LA-12893-C
Conference

Stable Isotope Applications in Biomolecular Structure and Mechanisms

Proceedings of the Conference
Santa Fe, New Mexico
March 27 – 31, 1994

Edited by
Jill Trewhella,
Timothy A. Cross,
and Clifford J. Unkefer

Edited by Jody H. Heiken, CIC-1

Cover illustration by Ward Zaelke, CIC-1

ABSTRACT

Knowledge of biomolecular structure is a prerequisite for understanding biomolecular function, and stable isotopes play an increasingly important role in structure determination of biological molecules. The first Conference on Stable Isotope Applications in Biomolecular Structure and Mechanisms was held in Santa Fe, New Mexico, March 27–31, 1994. More than 120 participants from 8 countries and 44 institutions reviewed significant developments, discussed the most promising applications for stable isotopes, and addressed future needs and challenges. Participants focused on applications of stable isotopes for studies of the structure and function of proteins, peptides, RNA, and DNA. Recent advances in NMR techniques, neutron scattering, EPR, and vibrational spectroscopy were highlighted in addition to the production and synthesis of labeled compounds. This volume includes invited speaker and poster presentations as well as a set of reports from discussion panels that focused on the needs of the scientific community and the potential roles of private industry, the National Stable Isotope Resource, and the National High Magnetic Field Laboratory in serving those needs.

DISCLAIMER

LA-12893-C
Conference

UC-908
Issued: December 1994

Stable Isotope Applications in Biomolecular Structure and Mechanisms

A meeting to bring together producers and users of stable-isotope-labeled compounds to assess current and future needs

Edited by
Jill Trewhella, Timothy A. Cross, and Clifford J. Unkefer

Los Alamos
NATIONAL LABORATORY
Los Alamos, New Mexico 87545

Conference Sponsors

Los Alamos National Laboratory, US Department of Energy

University of New Mexico Cancer Research and Treatment Center

Cell Biology, University of New Mexico, School of Medicine

The National High Magnetic Field Laboratory

MARTEK Biosciences Corporation

Cambridge Isotope Laboratories

ICON Services, Inc.

ISOTEC, Inc.

Program Committee

Nicholas A. Matwiyoff, Chairperson
University of New Mexico

Gerald T. Babcock
Michigan State University

Timothy A. Cross
Florida State University

Roger A. Jones
Rutgers University

John L. Markley
University of Wisconsin-Madison

Arthur Pardi
University of Colorado

Clifford J. Unkefer
Los Alamos National Laboratory

Organizing Committee

Jill Trewhella, Chairperson
Donald Ott
Louis A. "Pete" Silks
Mary Ann D. Martinez
Jean Stark

Los Alamos National Laboratory

Invited Speakers

Peter E. Wright, Keynote Speaker
The Scripps Research Institute

Gerald T. Babcock
Michigan State University

William W. Bachovchin
Tufts University

Adrian Bax
National Institutes of Health

Walter J. Chazin
The Scripps Research Institute

Timothy A. Cross
Florida State University

Gary Drobny
University of Washington

Heinz G. Floss
University of Washington

Angela M. Gronenborn
National Institutes of Health

Brian Hoffman
Northwestern University

Roger A. Jones
Rutgers University

Masatsune Kainosho
Tokyo Metropolitan University

Robert Kaptein
Utrecht University

Garry C. King
University of New South Wales

David M. LeMaster
Northwestern University

John L. Markley
University of Wisconsin

Arthur Pardi
University of Colorado

Jacob Schaefer
Washington University

Anthony S. Serianni
University of Notre Dame

Ignacio Tinoco
University of California, Berkeley

Jill Trewhella
Los Alamos National Laboratory

Clifford J. Unkefer
Los Alamos National Laboratory

William H. Woodruff
Los Alamos National Laboratory

CONTENTS

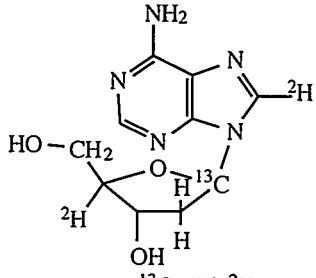
Discussion Group Reports

Abstracts for Posters

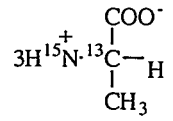
Nomenclature and Acronyms

Isotopic Labeling

The IUPAC nomenclature for isotopically modified organic compounds [*Pure and Appl. Chem.* (1979) *51*, 353-380] was used to name labeled compounds. Examples are illustrated to the right.



2'-deoxy-[1'-^{13}C, 4',8-^{2}H$_2$]adenosine



L-[α-^{13}C, ^{15}N]Alanine

Atom Numbers

Atoms are identified with their atomic symbol, followed by a number indicating the position. Substituents are identified by their atomic symbol, which is preceded by a number indicating the carbon to which they are attached. For example, the alanine illustrated above right is ^{13}C-labeled at C2 (or Cα) and ^{15}N-labeled at 2N (or αN); the adenine ring is labeled with deuterium at 8H. By convention, substituents in sugars have the same number as their attached carbons. Therefore, when naming atoms of carbohydrates, the number of the substituent follows its atomic symbol. For example, the 2'-deoxy-ribose in the 2'-deoxy-adenosine shown is ^{13}C-labeled at C1' and deuterated at H4'.

Sequence Numbering

CGCGATTGCG GlyAlaCysHisPro

Sequence numbers appear as a superscript that follows the one-letter code for nucleotides and the three-letter code for amino acids. For example, in the sequences above the adenine is would be identified as A^5 and cysteine as Cys^3. As is the case for isotope numbering, subscripts indicate number of copies.

Acronyms

2D = two dimensional

3D = three dimensional

γ = magnetogyric ratio

COSY = two-dimensional correlated spectroscopy

DEPT = distortionless enhancement by polarization transfer

DESERT = deuterium substitution effects on ^{1}H relaxation times

E. COSY = exclusive correlation spectroscopy

HCACO = three-dimensional intraresidue αH-$^{13}C\alpha$-C1 (carbonyl) correlation spectroscopy

HCA(CO)N = three-dimensional αH-$C\alpha$ of one residue to ^{15}NH of the next residue (via carbonyl) correlation spectroscopy

H(CA)NH = three-dimensional αH (via $C\alpha$) to nitrogen to amide proton correlation spectroscopy

HCCH-COSY = three-dimensional ^{1}H-^{13}C-^{13}C-^{1}H correlation spectroscopy via $^{1}J_{CC}$ carbon couplings

HCCH-TOCSY = three-dimensional ^{1}H-^{13}C-^{13}C-^{1}H total correlation spectroscopy with isotropic mixing of ^{13}C magnetization

HE-NOSEY = Hartman-echo NOSEY

HNCA = three-dimensional $^{13}C\alpha$-$^{15}N\alpha$-NH correlation spectroscopy

HNCO = three-dimensional $^{13}C1$-$^{15}N\alpha$-^{1}HN spectrum correlating amide ^{1}H and $^{15}N\alpha$ with the C1 shift of the preceding residue

HMQC = heteronuclear multiple quantum coherence

HMQCJ = variant of HMQC-COS; differs from the normal HMQC experiment by the addition of an ^{1}H 90° pulse immediately before acquisition

$^{1}H\{^{15}N\}MBC$ = heteronuclear multiple-bond correlation spectroscopy

HSQC = heteronuclear single-quantum correlation

HOHAHA = two-dimensional homonuclear Hartman-Hahn spectroscopy

INEPT = insensitive nuclei enhanced by polarization transfer

MeA = methylalanine

NMR = nuclear magnetic resonance

NOE = nuclear Overhauser effect

NOESY = two-dimensional nuclear Overhauser enhancement spectroscopy

P.E. COSY = primitive exclusive two-dimensional correlated spectroscopy

ppm/b = parts per million/billion

REDOR = rotational-echo, double resonance

ROESY = rotating-frame Overhauser spectroscopy

rms = root mean square

r.m.s.d. = root mean square deviation

S_N2 = substitution nucleophilic bimolecular

T_1 = longitudinal relaxation rate

T_2 = transverse relaxation rate

TOCSY = total correlation spectroscopy

ZQ = double quantum

PREFACE

An understanding of biological function at the molecular level is the foundation for many applications in health and biotechnology. For example, we have entered the era of "molecular medicine," in which we seek to treat disease using rationally designed drugs or gene therapy; in another arena we seek to reduce the trillion-dollar environmental cleanup bill we face as a nation by using biological structures in remediation and waste treatment; in energy and manufacturing, we are employing biological molecules, or molecules that mimic them, to develop useful new materials. Knowledge of biomolecular structure is a prerequisite for understanding biomolecular function, and stable isotopes are playing an increasingly important role in structure determination of biological molecules. In the past decade, we have seen uniform ^{13}C and ^{15}N labeling push the limits of structure determination by NMR spectroscopy to larger and larger proteins; ^{2}H labeling for neutron scattering studies of biomolecular complexes has become easier and more economical, and specific labeling has proven to be a powerful tool for vibrational spectroscopy and magnetic resonance techniques. Each of these technologies has played a critical role in our effort to understand biological function at the molecular level. In the future, optimal progress will continue through the successful combination of isotope-labeling technology with the respective physical and engineering tools we use.

Observing the rapid developments in this field, we believed it was important to bring together the users and producers of stable-isotope-labeled compounds for biomolecular research to examine the future and determine the community's needs and priorities. As scientists, we are each aware of shrinking resources and the need for careful planning in order to best serve the community. The first Conference on Stable Isotope Applications in Biomolecular Structure and Mechanisms was held at the Eldorado Hotel in picturesque and historic Santa Fe, March 27–31, 1994. An early spring snow storm greeted over 120 participants who came from 8 countries and 44 institutions (including 13 companies) not only to review significant developments and the most promising applications for stable isotopes, but also to address future needs and challenges. During the

four-day meeting, participants discussed applications of stable isotopes to the study of the structure and function of proteins, peptides, RNA, and DNA. Recent advances in NMR techniques, neutron scattering, EPR, and vibrational spectroscopy were highlighted, as well as the production and synthesis of labeled compounds.

This volume includes the invited speaker and poster presentations as well as a set of reports from discussion panels that met during the meeting to focus on the needs of the community and the potential roles of private industry, the National Stable Isotope Resource, and the National High Magnetic Field Laboratory in serving those needs.

The invited-speaker presentations provide powerful testimony about the importance of stable isotope technology as it leads us to a greater understanding of biomolecular structure and mechanisms. Peter Wright and Angela Gronenborn show how uniform isotope enrichment is helping to extend the limits of NMR applications for larger protein structures, including protein/protein interactions; David LeMaster and Cliff Unkefer describe specific isotope-labeling strategies that could push these limits even further. John Markley discusses the ways uniform isotope enrichment facilitates NMR studies of difficult paramagnetic systems such as respiratory ferrodoxins. Gerry Babcock, using vibrational spectroscopy, and Bill Bachovchin, using NMR, capitalize on specific isotope labeling to study the active sites of respiratory proteins (cytochrome oxidase, MMO, cytochrome P450) and enzymes (serine proteinases). Woody Woodruff describes the power of isotope labeling for time-resolved studies of both changes in respiratory protein structure during function and the fast reactions of protein folding.

Jacob Schaefer describes modern solid-state NMR methods for structural analysis of stable isotope labeled proteins. Tim Cross's elegant example of specific labeling and solid-state NMR spectroscopy reveals the complete three-dimensional structure of the membrane-embedded polypeptide gramicidin. Jill Trewhella's descriptions of neutron scattering experiments with uniform deuterium labeling reveal information on the conformations and interactions of proteins responsible for Ca^{2+}-dependent biochemical regulation. Walter Chazin, Roger Jones, and Masatsune Kainosho describe NMR experiments that use specific isotope labeling to study DNA structures and DNA/drug interactions. Garry King shows how ^{13}C NMR relaxation measurements provide dynamics information on the HIV-1 TAR RNA

element. Art Pardi and Ignacio Tinoco apply uniform labeling and multidimensional NMR to obtain structural data for important RNA molecules, including RNA enzymes and a retrovirus RNA element. Tony Serianni's eloquent description of labeling strategies for carbohydrates and nucleosides includes examples of their application in NMR studies of oligosaccharide and oligonucleotide conformations.

From the research described in this volume, one dominant theme emerges—advanced stable-isotope labeling strategies are critically important in the application of more sophisticated spectroscopic techniques to increasingly complex problems in biomolecular structure and mechanisms.

Our thanks go to Professor Nicholas Matwiyoff, who first suggested that we organize this conference. We especially thank the organizers, program committee, sponsors, meeting participants, and contributors to this volume for making the conference technically exciting and visionary and for ensuring that this Proceedings is a valuable reference and planning document. We are grateful to MaryAnn Martinez for her outstanding management of the conference logistics and correspondance and to Jody Heiken, whose assistance in composing this volume has been invaluable. We also thank Griselda Hernandez for assistance in proofreading.

We hope to see you all in Santa Fe for our next meeting!

Jill Trewhella
Timothy A. Cross
Clifford J. Unkefer

KEYNOTE ADDRESS

Isotope Labeling for NMR Studies of Macromolecular Structure and Interactions

Peter E. Wright

Department of Molecular Biology
The Scripps Research Institute
La Jolla, CA 92037

Implementation of biosynthetic methods for uniform or specific isotope labeling of proteins, coupled with the recent development of powerful heteronuclear multidimensional NMR methods, has led to a dramatic increase in the size and complexity of macromolecular systems that are now amenable to NMR structural analysis. In recent years, a new technology has emerged that combines uniform ^{13}C, ^{15}N labeling with heteronuclear multidimensional NMR methods to allow NMR structural studies of systems approaching 25 to 30 kDa in molecular weight. In addition, with the introduction of specific ^{13}C and ^{15}N labels into ligands, meaningful NMR studies of complexes of even higher molecular weight have become feasible. These advances usher in a new era in which the earlier, rather stringent molecular weight limitations have been greatly surpassed and NMR can begin to address many central biological problems that involve macromolecular structure, dynamics, and interactions.

This article primarily reviews recent work from this laboratory that exemplifies the power and promise of isotope labeling for NMR structural studies of large and complex macromolecular systems. Applications of uniform ^{13}C, ^{15}N labeling to determine the three-dimensional structure and dynamics of the *B. subtilis* glucose permease IIA domain (IIA^{glc}), containing 162 amino acids, are described. In the phosphotransferase system, a phosphoryl group is transferred to IIA^{glc} from the smaller phosphocarrier protein HPr (88 residues). Binding interfaces of IIA^{glc} and HPr have been identified and mapped by ^{15}N-edited and ^{15}N-filtered NMR studies of the complex. Applications of isotope-edited NMR methods to much larger

complexes—in this case, an ^{15}N-labeled peptide antigen bound to the 50 kDa Fab' fragment of a cognate antipeptide antibody—will also be described.

NMR Structural Studies of *B. subtilis* Glucose Permease IIA Domain

Bacterial phosphoenolpyruvate:sugar phosphotransferase systems (PTS) mediate the concomitant transmembrane transport and phosphorylation of a number of simple carbohydrates. The PTS consists of two nonspecific energy-coupling proteins, enzyme I and HPr, and a sugar-specific permease complex known as enzyme II (for reviews see Saier and Reizer, 1992; Meadow *et al.*, 1990; Reizer *et al.*, 1988; Saier, 1989). In PTS-mediated glucose transport, a phosphoryl moiety is transferred sequentially from phosphoenolpyruvate (PEP) to enzyme I, HPr, glucose-specific enzyme IIA (IIA^{glc}), membrane-bound IIB (IIB^{glc}), and finally to the sugar.

B. subtilis possesses a glucose permease in which a cytosolic C-terminal IIA domain is covalently linked to the membrane-bound $IICB^{glc}$ domains that are directly involved in sugar transport (Sutrina *et al.*, 1990). Obviously, the intact membrane-bound protein is not accessible to NMR studies; however, the 162-residue IIA^{glc} domain can be readily expressed in *E. coli* as a soluble protein suitable for NMR analysis (Sutrina *et al.*, 1990; Reizer *et al.*, 1992). For a protein of this size, extensive resonance assignments are not possible by homonuclear two-dimensional NMR methods, and it is necessary to resort to uniform isotope labeling and heteronuclear multidimensional NMR techniques. Fortunately, IIA^{glc} can be readily and economically labeled with ^{13}C and ^{15}N (to > 95%) by growing the cells in minimal medium that contains ^{15}N-labeled ammonium sulfate and ^{13}C-labeled glucose as the sole nitrogen and carbon sources (Fairbrother *et al.*, 1992a).

Resonance Assignments

Assignment of the ^{1}H and ^{15}N backbone resonances and determination of the secondary structure of IIA^{glc} was achieved by using primarily three-dimensional heteronuclear ^{1}H-^{15}N NMR spectroscopy (Fairbrother *et al.*, 1991). When using these techniques, it was not possible to assign the aliphatic ^{1}H resonances of many residues because (a) extensive chemical-shift overlap occurs in the aliphatic region of the spectra and (b) coherence transfer in the three-dimensional ^{1}H-^{15}N TOCSY-HMQC experiment is inefficient for the longer sidechains. Such problems, typical for proteins the size of the IIA^{glc} domain, have led to development of heteronuclear NMR techniques that rely on uniformly large one-bond ^{1}H-^{13}C, ^{13}C-^{13}C, ^{13}C-^{15}N, and ^{1}H-^{15}N scalar J coupling constants to establish through-bond connectivities (Kay *et al.*, 1990a,b; Ikura *et al.*, 1990; Fesik *et al.*, 1990; Bax *et al.*, 1990a,b; Clore *et al.*, 1990). Beginning with the previously assigned amide ^{15}N and $\alpha^{1}H$ resonances, Fairbrother *et al.*, (1992a)

confirmed the sequential assignments and obtained the $^{13}C\alpha$ chemical shifts from a three-dimensional HCA(CO)N triple-resonance experiment (Kay *et al.*, 1990b; Ikura *et al.*, 1990; Powers *et al.*, 1991; Palmer *et al.*, 1992). Assignment of the sidechain spin systems was completed by using three-dimensional HCCH-COSY (Bax *et al.*, 1990a; Clore *et al.*, 1990) and HCCH-TOCSY (Bax *et al.*, 1990b; Clore *et al.*, 1990) experiments. In this way, virtually complete assignments could be obtained; these assignments were the basis for subsequent investigations of three-dimensional solution structure, dynamics, and macromolecular interactions.

Structure Determination

The solution structure of *B. subtilis* IIA^{glc} was determined by combining metric matrix distance geometry methods with subsequent restrained molecular dynamics refinement (Fairbrother *et al.*, 1992b; Chen *et al.*, submitted). The availability of uniformly labeled protein and the use of three- and four-dimensional ^{13}C- and ^{15}N-edited NOESY spectra made it possible to derive a large number (2137) of distance constraints from the NMR data. These distance constraints were supplemented by 95 backbone and sidechain torsion-angle constraints. The resulting structures are well defined (Fig. 1) and are very similar to the independently determined x-ray structure of the *B. subtilis* IIA^{glc} domain (Liao *et al.*, 1991). The major difference is at the N-terminus, where the residues of the linker to the IICB integral membrane domain are disordered in solution but adopt a well-defined conformation in the crystal.

Dynamics of IIA^{glc}

In recent years, a number of groups have begun to exploit the potential of NMR relaxation measurements for probing the internal molecular dynamics of isotope-labeled proteins. The measurement of ^{15}N or ^{13}C relaxation rates provides valuable information about the internal dynamics of proteins on time scales shorter than the rotational correlation time. Taking advantage of the availability of uniformly ^{15}N-labeled IIA^{glc}, it was possible to characterize the backbone dynamics by using ^{15}N relaxation measurements (Stone *et al.*, 1992). The T_1 and T_2 relaxation time constants and steady-state $\{^1H\}^{15}N$ NOEs for 137 of the 151 protonated backbone nitrogens were measured. These data were analyzed by using the model-free approach of Lipari and Szabo (Lipari and Szabo, 1982a,b). In this approach, two parameters are used to describe internal motion—a generalized order parameter S^2 and an effective internal correlation time τ_e. The value of S^2 can range from 0 (indicating completely isotropic internal motions) to 1 (for internal motions that are completely restricted relative to a fixed molecular frame of reference). Most of the residues in *B. subtilis* IIA^{glc} exhibit relatively restricted internal motions with order parameters in the range of 0.75 to 0.90 (Fig. 2). These values are typical for

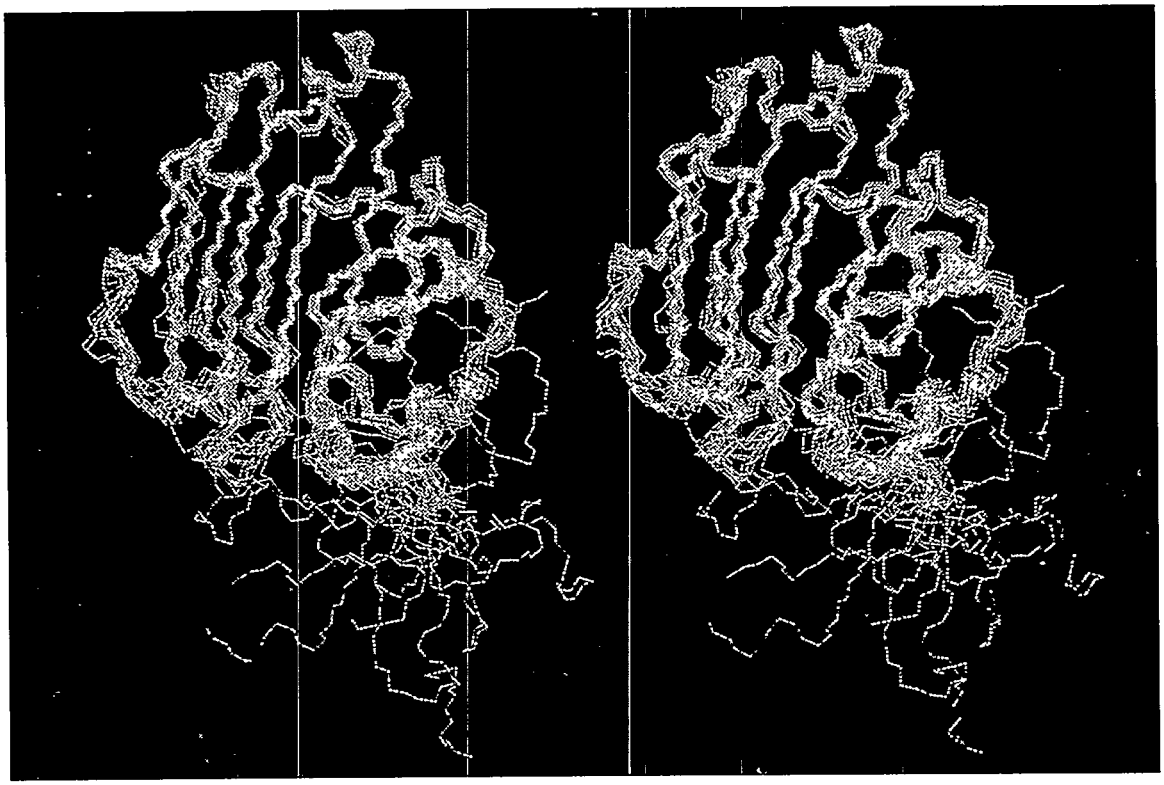


Fig. 1. Stereoview of backbone superposition of MD-refined structures of B. subtilis *IIA^{glc} domain (Chen* et al., *submitted).*

globular proteins. Two areas of the molecule, however, are significantly more flexible, with smaller S^2 values (<0.5) and a τ_e of approximately 100 ps. The segment of greatest flexibility is the N-terminal linker region, which is conformationally disordered in the NMR structures. A further region of significant flexibility is at the center (around residue 32) of the Ω-loop adjacent to the active site. In addition, residues 146 to 149, located in another loop region, are more flexible than average. The flexibility of the N-terminal region was expected because the first 13 residues are part of the Q-linker that joins the IIA^{glc} domain to the membrane-bound $IICB^{glc}$ domains. The high mobility of the Ω-loop, with its proximity to the active site, has led to the suggestion that it may form part of the binding surface for HPr and/or IIB^{glc} (Stone *et al.*, 1992).

Mapping the Binding Interfaces of IIA^{glc} and HPr

The availability of isotopically labeled protein offers unique advantages for spectral editing, which enhances characterization of intermolecular complexes. In the present case, a complex was formed between uniformly ^{15}N-labeled IIA^{glc} and unlabeled HPr. The exchange is fast on the NMR time scale, and chemical shift changes on complexation were used to identify the surfaces on each protein that form the binding interface (Chen *et al.*, 1993). Resonances of IIA^{glc} can be observed selectively in the presence of unlabeled HPr by using inverse-detected ^{1}H-^{15}N correlated NMR experiments. Conversely, by applying ^{15}N-filtered NMR experiments (Otting and Wüthrich, 1990), resonances from the unlàbeled HPr can be observed selectively. The ^{1}H-^{15}N HSQC spectrum of the IIA^{glc}-HPr complex is shown in Fig. 3; each of the peaks arises solely from the ^{15}N-labeled IIA^{glc}, and HPr resonances are completely absent from the spectrum. Chen *et al.* (1993) observed significant chemical shift changes for several residues of IIA^{glc} (indicated by boxes in Fig. 3). These include residues 33–40 of the Ω-loop, whose apex is found close to the active site; several residues from the eight-stranded β-sheet (61–64, 70–72, 79–81, and 132–134); and residues 87 and 89, which are also in a loop near the active site. The backbone amide ^{15}N and ^{1}H chemical shifts of the active-site histidine

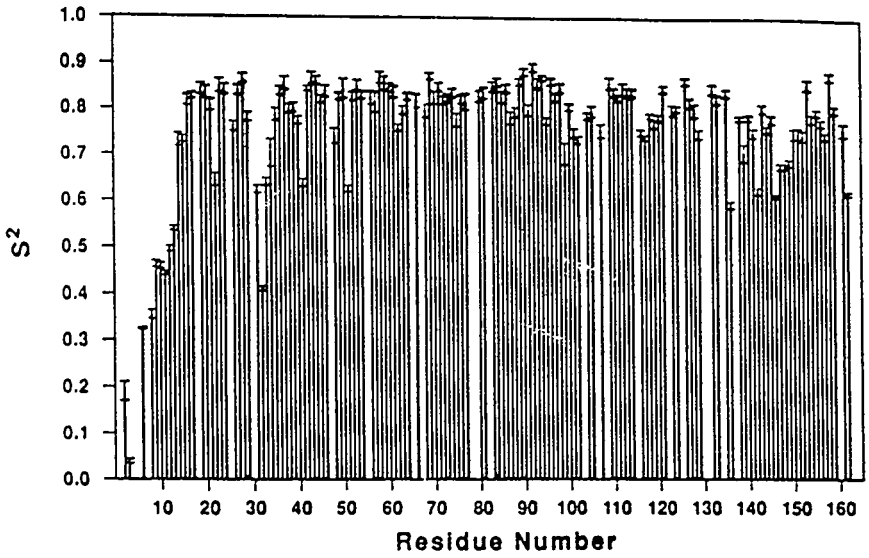


Fig. 2. Order parameters (S^2) for* B. subtilis *IIA^{glc} domain obtained from ^{15}N relaxation measurements.

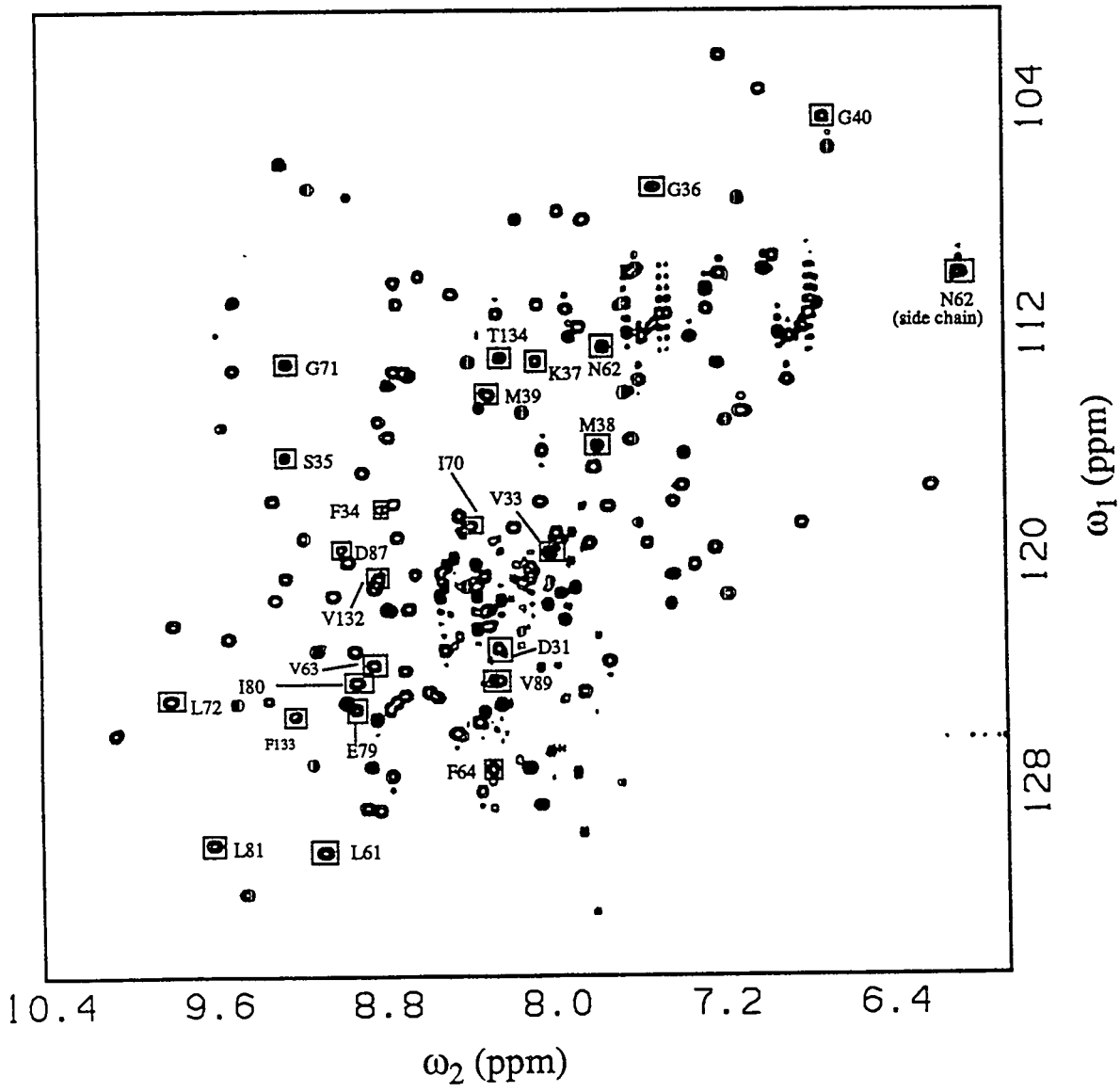


Fig. 3. ^{1}H-^{15}N HSQC spectrum of the IIA^{glc}-HPr complex (adapted from Chen et al., *1993). The boxes indicate residues for which amide ^{1}H or ^{15}N chemical shifts change significantly on complex formation. Amino acid residues are designated by their one-letter code followed by their position in the sequence. A = Ala, D = Asp, E = Glu, F = Phe, G = Gly, H = His, I = Ile, K = Lys, L = Leu, M = Met, N = Asn, S = Ser, T = Thr, V = Val. The spectrum contains only resonances of the ^{15}N-labeled IIA^{glc}.*

residues 83 and 68 and the remainder of the protein are essentially unchanged. All of the residues for which chemical shift changes were observed upon binding of HPr are located in the vicinity of the active site and, for the most part, form a continuous surface.

Chemical shift changes in HPr upon complex formation with IIA^{glc} were identified using ω_2-half-filtered COSY and TOCSY spectra (Chen *et al.*, 1993). A region of the half-filtered TOCSY spectrum of the IIA^{glc}-HPr complex is shown in Fig. 4. The spectrum only contains crosspeaks from HPr; the signals from the ^{15}N amide protons of the IIA^{glc} are eliminated by the ^{15}N filtering. Resonance assignments were made on the basis of published data for free HPr (Wittekind *et al.*, 1990). Significant backbone chemical shift perturbations in HPr were observed for residues 13–23, which include the active-site His^{15} and a short segment of α-helix. Shift changes were also observed for residues 51–56, which form part of a loop between an α-helix and a strand of the β-sheet, for residues 46 and 47, and for a few residues in the β-sheet (residues 33 and 41–44). These residues form a continuous surface, including the active site and an area immediately adjacent to the active site.

It is notable that residues that exhibit chemical shift changes in both proteins are in regions that are predominantly hydrophobic in nature. Because the areas of observed chemical shift changes in both HPr and IIA^{glc} are in the vicinity of the known active sites and appear to be complementary to each other, it is highly probable that they indicate residues that are involved in the binding interface. Interaction between HPr and IIA^{glc} is a necessary first step for the transfer of a phosphoryl group between the two proteins.

Conformation of an Fab'-Bound Peptide

NMR studies of high-molecular-weight systems are greatly facilitated by isotope-edited NMR techniques; for example, NMR studies (Tsang *et al.*, 1992) of the conformation and dynamics of the peptide antigen MetHisLysAspPheLeuGluLysIleGlyGlyLeu bound to the Fab' fragment (molecular weight ~50 kDa) of a monoclonal antibody, as described in this article. The antibody, termed B13A2, was raised against a 19-residue peptide immunogen (Feiser *et al.*, 1987) derived from the C-helix of myohemerythrin; the antigen corresponds to a truncated version of the original peptide used to elicit the antibody response. A series of synthetic peptides labeled with ^{15}N (98%) or ^{13}C (99%) at the backbone of individual amino acid residues was produced (Tsang *et al.*, 1992). Well-resolved amide ^{1}H and ^{15}N resonances were observed in isotope-edited spectra of the peptide-antibody complex, despite the high molecular weight. Significant variations in chemical shift and resonance linewidth were observed at different sites along the peptide backbone. These linewidths reflect the conformational flexibility of the bound peptide; the broadest

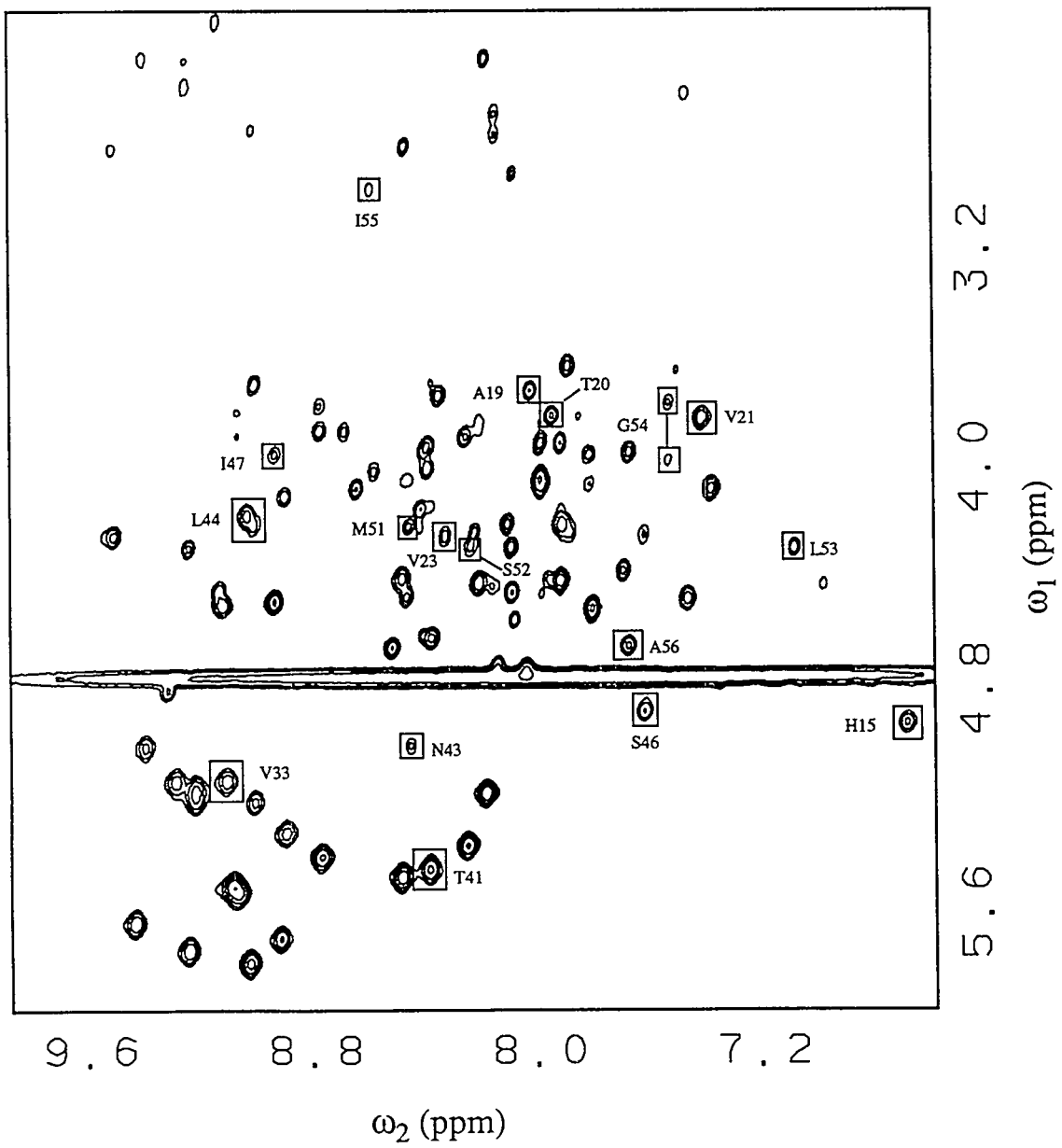


Fig. 4. Fingerprint region of ω_2-half-filtered TOCSY spectrum of the IIA^{glc}-HPr complex (adapted from Chen et al.*, 1993). Boxes indicate residues for which αH or NH resonances shift significantly on complex formation. Amino acid designations are those described in the Fig. 3 caption. The spectrum contains only HPr resonances; all resonances of the ^{15}N-labeled IIA^{glc} are eliminated by the filter.*

resonances were from amino acids within the immunologically defined epitope. The backbone motions of the peptide appear to be significantly restricted within the antibody combining site, whereas residues outside this region exhibit considerable flexibility.

To obtain information on the conformation of the antibody-bound peptide, Tsang *et al.* (1992) performed 1D and 2D isotope-edited NOE experiments. An example of a 1D ^{15}N-edited NOE spectrum of the complex is shown in Fig. 5, enhanced by specific deuteration of the peptide to eliminate undesirable magnetization leakage away from the nuclei of interest (Tsang *et al.*, 1990). Strong sequential amide-amide NOE connectivities observed throughout most of the epitope correspond to inter-proton distances shorter than 3 Å (Tsang *et al.*, 1992). These short NH-NH distances provide strong evidence that the peptide adopts helical backbone conformations in the combining site of the Fab' fragment of monoclonal antibody B13A2—unlike the free peptide, which exhibits a small population of nascent helical structures in the epitope region in aqueous solution (Dyson *et al.*, 1988). Unfortunately, it is difficult to measure longer range NOEs between protons separated by more than 3Å (which would be extremely valuable for structure determination) because of the high molecular weight of the antibody complex and the broadness of the resonances. This situation could probably be alleviated by selective deuteration of the peptide (coupled, of course, with ^{13}C and/or ^{15}N labeling) and perdeuteration of the protein.

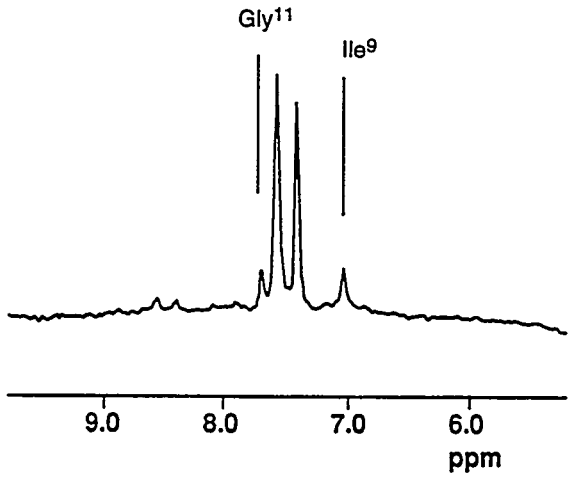


Fig. 5. The ^{15}N-edited NOE spectrum of Fab'-peptide complex (adapted from Tsang et al., 1992). The peptide MetHisLysAspPheLeuGluLysIleGlyGlyLeu is labeled with ^{15}N at Gly^{10}, α-deuterated at both Gly^{10} and Gly^{11}, and perdeuterated at Ile^{9}. Strong Ile^{9}-Gly^{10} and Gly^{10}-Gly^{11} amide-amide NOEs characteristic of helix are indicated. The resonance of Gly^{10} appears as a doublet as a result of ^{15}N coupling.

Acknowledgements

The author is indebted to many colleagues who have contributed to this work and, in particular, thanks the numerous members or ex-members of his laboratory who are directly responsible for the NMR studies reviewed above: Wayne Fairbrother, Art Palmer, Yuan Chen, Martin Stone, Garry Gippert, Pearl Tsang, Mark Rance, and Jane Dyson. This work was supported by grants GM36643 and CA27498 from the National Institutes of Health.

References

Bax, A., Clore, G.M., Driscoll, P.C., Gronenborn, A.M., Ikura, M., and Kay, L.E. (1990a) *J. Magn. Reson. 87,* 620-627.

Bax, A., Clore, G.M., and Gronenborn, A.M. (1990b) *J. Magn. Reson. 88,* 425-431.

Chen, Y, Reizer, J., Saier, M.H., Jr., Fairbrother, W.J., and Wright, P.E. (1993) *Biochemistry 32,* 32-37.

Clore, G.M., Bax, A., Driscoll, P.C., Wingfield, P.T., and Gronenborn, A.M., (1990) *Biochemistry 29,* 8172-8184.

Dyson, H.J., Rance, M., Houghten, R.A., Wright, P.E., and Lerner, R.A. (1988) *J. Mol. Biol. 201,* 201-217.

Fairbrother, W.J., Cavanagh, J., Dyson, H.J., Palmer, A.G., III, Sutrina, S.L., Reizer, J., Saier, M.H., Jr., and Wright, P.E. (1991) *Biochemistry 30,* 6896-6907.

Fairbrother, Palmer, A.G., III, Rance, M., Reizer, J., Saier, M.H., Jr., and Wright, P.E. (1992a) *Biochemistry 31,* 4413-4425.

Fairbrother, W.J., Gippert, G.P., Reizer, J., Saier, M.H., Jr., and Wright, P.E. (1992b) *FEBS Lett. 296,* 148-152.

Fesik, S.W., Eaton, H.L., Olejniczak, E.T., Zuiderweg, E.R.P., McIntosh, L.P., and Dahlquist, F.W. (1990) *J. Am. Chem. Soc. 112,* 886-887.

Fieser, T.M., Tainer, J.A., Geysen, H.M., Houghten, R.A., and Lerner, R.A. (1987) *Proc. Natl. Acad. Sci. USA 84,* 8568-8572.

Ikura, M., Kay, L.E., and Bax, A. (1990) *Biochemistry 29,* 4659-4667.

Kay, L.E., Ikura, M., and Bax, A. (1990a) *J. Am. Chem. Soc. 112,* 888-889.

Kay, L.E., Ikura, M., Tschudin, R., and Bax, A. (1990b) *J. Magn. Reson. 89,* 496-514.

Liao, D.-I., Kapadia, G., Reddy, P., Saier, M.H., Jr., Reizer, J., and Herzberg, O. (1991) *Biochemistry 30,* 9583-9594.

Lipari, G. and Szabo, A. (1982a) *J. Am. Chem. Soc. 104,* 4546-4559.

Lipari, G. and Szabo, A. (1982b) *J. Am. Chem. Soc. 104,* 4559-4570.

Meadow, N.D., Fox, D.K., and Roseman, S. (1990) *Ann. Rev. Biochem. 59,* 497-542.

Otting, G. and Wüthrich, K. (1990) *Q. Rev. Biophys. 23,* 39-96.

Palmer, A.G., III, Fairbrother, W.J., Cavanagh, J., Wright, P.E., and Rance, M. (1992) *J. Biomol. NMR 2,* 103-108.

Powers, R., Gronenborn, A.M., Clore, G.M., and Bax, A. (1991) *J. Magn. Reson. 94,* 209-213.

Reizer, J., Saier, M.H., Jr., Deutscher, J., Grenier, F., Thompson, J., and Hengstenberg, W. (1988) *CRC Critical Rev. Microbiol. 15,* 297-338.

Reizer, J., Sutrina, S.L., Wu, L.-F., Deutscher, J., Reddy, P., and Saier, M.H., Jr. (1992) *J. Biol. Chem. 267,* 9158-9169.

Saier, M.H., Jr. and Reizer, J. (1992) *J. Bacteriol. 174,* 1433-1438.

Saier, M.H., Jr. (1989) *Microbiological Reviews 53*, 109-120.

Stone, M.J., Fairbrother, W.J., Palmer, A.G., III, Reizer, J., Saier, M.H., Jr., and Wright, P.E. (1992) *Biochemistry 31*, 4394-4406.

Sutrina, S.L., Reddy, P., Saier, M.H., Jr., and Reizer, J. (1990) *J. Biol. Chem. 265*, 18581-18589.

Tsang, P., Rance, M., Fieser, T.M., Ostresh, J.M., Houghten, R.A., Lerner, R.A., and Wright, P.E. (1992) *Biochemistry 31*, 3862-3871.

Tsang, P., Wright, P.E., and Rance, M. (1990) *J. Am. Chem. Soc. 112*, 8183-8185.

Wittekind, M.G., Reizer, J., and Klevit, R.E. (1990) *Biochemistry 29*, 7191-7200.

Invited Papers

Oxygen Chemistry in Biology: Vibrational Spectroscopy, Stable Isotopes, and Future Applications

Gerald T. Babcock

Department of Chemistry
Michigan State University
East Lansing, MI 48824

Dioxygen is an ideally suited substrate for enzymatic manipulation in oxidation-reduction chemistry and in substrate transformation. It is a powerful oxidant with a midpoint potential of 0.815 V at neutral pH; at the same time, however, it exists in a triplet state in its most stable electronic configuration (George, 1965). This latter property confers kinetic inertness as a result of spin-conservation restrictions on reaction chemistry. If these restrictions can be overcome and controlled, dioxygen's high redox potential can be used to maximize efficiency in free-energy conversion processes and to effect activation of relatively inert substrates.

Evolutionary pressure has clearly selected a means by which to take advantage of the unique chemistry of O_2, and a number of enzymes have evolved that are able to activate dioxygen kinetically and control its redox chemistry. In mitochondrial respiration, for example, the use of O_2 as the terminal oxidant substantially increases the free-energy release made available for conservation in ATP formation. In this process, electrons enter the respiratory chain as the reducing substrate NADH ($E_m^{o'} = -0.33$ V) and exit at O_2. This releases >1 V of free energy, a major fraction of which is captured as a pH and potential gradient to be ultimately used in driving endergonic cellular processes (Babcock and Wikström, 1992). The use of a less highly oxidizing terminal electron acceptor (for example, elemental sulfur) would decrease the free-energy span between input and output stages of the respiratory chain and reduce the overall free-energy release significantly. The same reaction principles are apparent in enzymatic systems that use O_2 as a substrate in catalyzing difficult bond activation chemistry. Both cytochrome P450 and methane monooxygenase, for instance,

use O_2 as a reactant in activating C-H bonds for oxygen insertion to form alcohol products (Guengerich, 1992; Rosenzweig *et al.*, 1993). Ribonucleotide reductase, galactose oxidase, and prostaglandin synthase all bind O_2 in the initial step in their catalytic cycles and use this substrate to oxidize a tyrosine phenol sidechain ($E_m^{o'} \approx 1.0$ V) to produce a neutral tyrosyl radical in later phases of the reaction (Reichard and Ehrenberg, 1983; Babcock *et al.*, 1992; Picot *et al.*, 1994). The amino-acid radical intermediate is activated to carry out the subsequent chemistry that generates the ultimate reaction products of the catalysis.

Considerable effort has been expended to understand the mechanisms by which dioxygen is mobilized to carry out these enzymatic processes. Approaches have ranged from employing model organic and inorganic chemistry to using time-resolved spectroscopies on the reacting enzyme system. In our own work on the cytochrome oxidase system, we have taken advantage of the unique properties of the heme a_3/Cu_B active site, using time-resolved resonance Raman to characterize the transient chemistry that leads to O_2 reduction to water (Babcock and Wikström, 1992; Babcock and Varotsis, 1993). Because the vibrational modes that we observe in the reaction time course can be related in a direct and straightforward way to chemical structure, we have been able to use stable oxygen isotope substitution strategies to identify specific intermediates at various times during the reaction. This has allowed us to write a detailed mechanism for O_2 reduction by cytochrome oxidase and to expand these insights to broader considerations of the control of oxygen chemistry in enzyme active sites.

Oxygen Chemistry and its Control in Heme-Enzyme Active Sites

Several themes emerge in the studies of biological control of dioxygen chemistry. First, in enzymes that activate O_2, redox-active metal ions with unpaired d-orbital electrons in at least one of the valence states that they occupy during catalysis are often used. The most commonly encountered metals in these processes are iron and copper in mononuclear or binuclear centers. Second, for the iron-based enzymes, both nonheme and heme iron active sites occur and, in fact, an emerging principle appears to be that enzymes with parallel functions occur in these two classes of iron-containing proteins. The nonheme sites appear to be more prevalent in procaryotic systems, and heme iron-containing centers are used more extensively in eucaryotic organisms.

An important third theme appears to be that Nature has elaborated significant control functions in developing the means by which to exploit dioxygen chemistry. O_2 is a multi-electron oxidant, and its reduction to water is a four-electron, four-proton process:

$$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O \quad . \qquad [1]$$

Because of this complexity, we necessarily expect intermediate species during its

metabolism and can write the following general scheme for its activation:

$$O_2 \xrightarrow[e^-]{} \text{"intermediates"} \xrightarrow[H^+]{} H_2O \quad . \quad [2]$$

Electron addition to O_2 generates intermediate species that are subsequently protonated to achieve O–O bond cleavage, reactive species generation, and product formation. A wide variety of intermediates are possible in this process, most of which are potentially toxic to the cellular host. These species include superoxide, hydroxyl radicals, peroxy radicals, and peroxides. Given the complexity of the reaction chemistry of dioxygen and the potential for cellular damage, one expects that control mechanisms exist; in fact, two have emerged and can be discussed within the context of the skeletal scheme in Reaction [2] above (Babcock and Varotsis, 1993).

The first control mechanism involves control at the electron-injection step and is epitomized by cytochrome P450 reaction chemistry. Figure 1 shows a representative model for the action of this class of enzyme. The oxidized enzyme binds the substrate and is reduced externally to form the ferrous enzyme, which then binds O_2 to form a stable ferrous oxy species. During the slow step in the overall process, the ternary complex is reduced externally; a series of intermediates (shown schematically in the illustration as peroxy and ferryl species) are produced and subsequently generate the hydroxylated product. Rate constants estimated for the various steps in this reaction sequence can be used to generate concentration/time profiles for each of the intermediates in Fig. 1. Results of this analysis are shown in Fig. 2. The important point gained from these profiles is that, because the slow step in the reaction occurs at the second electron-injection step, the concentrations of the chemically and catalytically active species do not build to significant levels. This reaction is an example of electron-transfer control; that is, the active site is "preloaded" with substrate (and also protons), and addition of an electron triggers a series of subsequent fast O–O bond cleavage and oxygen-atom insertion steps. These reactive intermediates are maintained at a low concentration, and potential cellular toxicity is minimized. Unfortunately, the kinetics of the system are such that our ability to probe the chemistry that occurs between disappearance of the oxy and appearance of the product is severely limited; indeed, intermediates in P450 catalysis beyond the oxy enzyme have not been trapped or even observed transiently in time-resolved measurements.

A second control mechanism that can be envisioned within the context of Reaction [2] involves control at the proton-transfer step. In such a mechanism, the electron transfer occurs to an "unloaded" active site, and only as the reaction develops are conditions (including proton uptake) established for O–O bond activation and cleavage. As a consequence, the rate limitation in the process occurs late in the reaction sequence and intermediate

species build to detectable levels. At first glance, such a control scheme seems to be counterproductive because of the potential toxicity of the intermediates that accumulate. Nonetheless, the initial flow-flash experiments on reduction of dioxygen by the terminal mitochondrial respiratory enzyme cytochrome oxidase, carried out by Gibson and Greenwood, suggested that intermediates in this process could be detected if time resolution were achieved on the microsecond scale (Gibson and Greenwood, 1963; Greenwood and Gibson, 1967). This observation, made with room-temperature optical techniques, was extended to the oxidase reaction by Chance and his co-workers (1975) and subsequently by both Malmström (Clore *et al.*, 1980) and Chan (Blair *et al.*, 1985) in their triple-trapping approach. Their technique allowed reactive intermediates in the

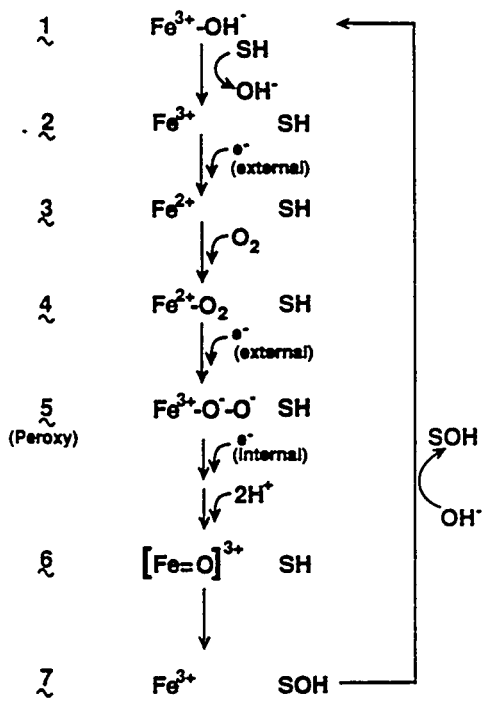


Fig. 1. Postulated P450 reaction cycle.

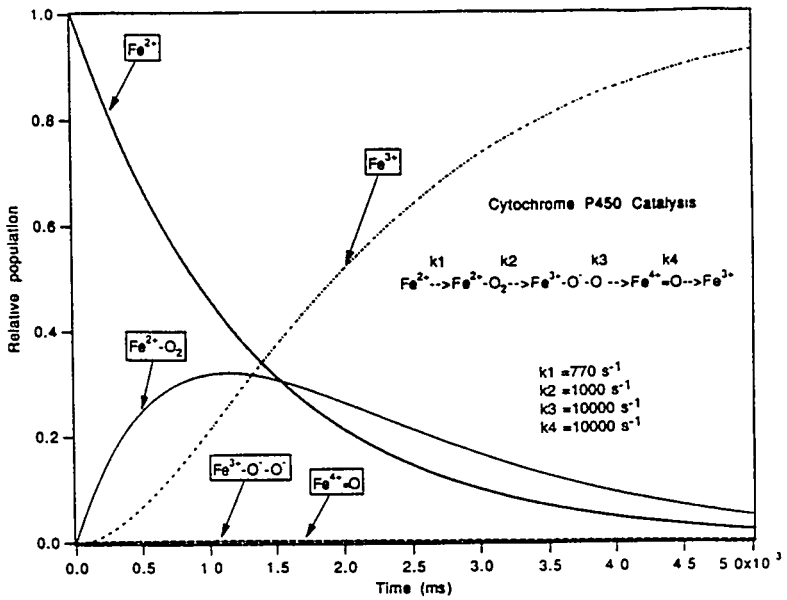


Fig. 2. Simulated concentration/ time profiles for P450; mechanism and rates are indicated.

reaction to be trapped at cryogenic temperatures. They used EPR and optical spectroscopies for characterization and postulated a number of molecular structures, including an initial oxy-adduct of the heme a_3 Fe at the active site.

We were able to bring vibrational techniques to bear on the oxidase reaction by adapting the flow-flash method so that time-resolved resonance Raman methods could be applied (Babcock *et al.*, 1984, 1985). We subsequently established the molecular properties of the oxy-adduct by detecting it directly (Varotsis *et al*, 1989a,b) and by comparison with heme *a*--O_2 model compounds (Oertling *et al.*, 1990). Both the Rousseau and Kitagawa groups have developed variations of the resonance Raman/ flow-flash methods. We and they have been active in extending the Raman approach, and, at present, four different oxygen-isotope-sensitive intermediates in the reaction sequence have been detected (Varotsis and Babcock, 1990; Varotsis *et al.*, 1990, 1993; Han *et al.*, 1990a,b,c; Ogura *et al.*, 1990, 1991, 1993).

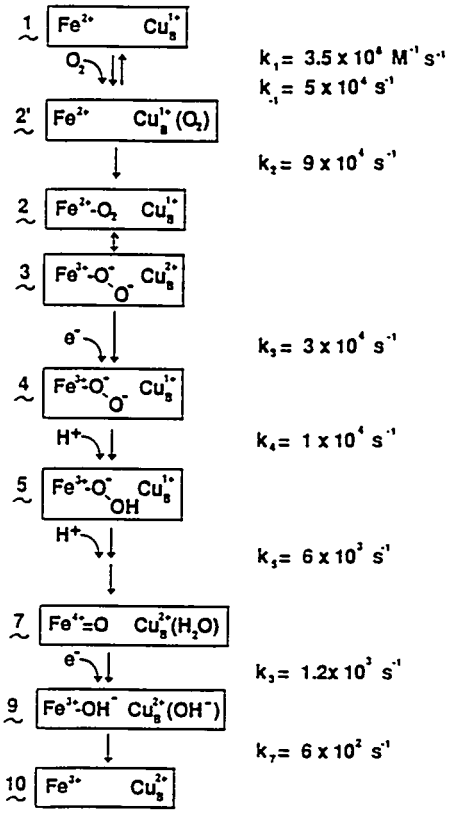


Fig. 3. Postulated reaction sequence for cytochrome oxidase (from Varotsis* et al., *1993).

Recently, we have combined these results with optical flow-flash results, particularly those obtained by Malmström, Nilsson, and their coworkers (Oliveberg *et al.*, 1989, 1991; Hallen and Nilsson, 1992), to write a complete reaction sequence for the reduction of O_2 by cytochrome oxidase (Babcock and Wikström, 1992; Varotsis *et al.*, 1993). This scheme is summarized in Fig. 3. Briefly, the O_2 reduction chemistry occurs at the binuclear heme a_3/Cu_B center. Following its reduction to the ferrous/cuprous pair, O_2 binding occurs, initially at the copper and then at the a_3-Fe^{2+}. Rapid electron transfer from the low-spin heme *a* center generates the three-electron reduced, O_2-bound binuclear center. Slower proton transfers generate the peroxy, which cleaves to produce the ferryl intermediate. The final electron transfer generates a transient hydroxy species, which decays to produce

the oxidized form of the enzyme. Within this model, the four FeO stretching vibrations that have been observed by Raman are assigned to oxy (Fe–O_2), protonated peroxy (Fe–OOH), ferryl (Fe=O), and hydroxy (Fe–OH) species. With the rate constants shown in Fig. 3, the reaction sequence can be simulated to generate concentration-time profiles, analogous to Fig. 2, for the various species involved (Varotsis *et al.*, 1993). These are shown in Fig. 4 and are qualitatively different than those obtained for the P450 reaction. In the oxidase reaction, reaction intermediates build to substantial concentrations throughout the time course; the ferryl, for example, accumulates to 60% of the total enzyme concentration. This behavior reflects the proton-control aspects of the oxidase reaction and the fact that the proton-uptake reactions in Fig. 3, not the electron-transfer reactions, rate limit the

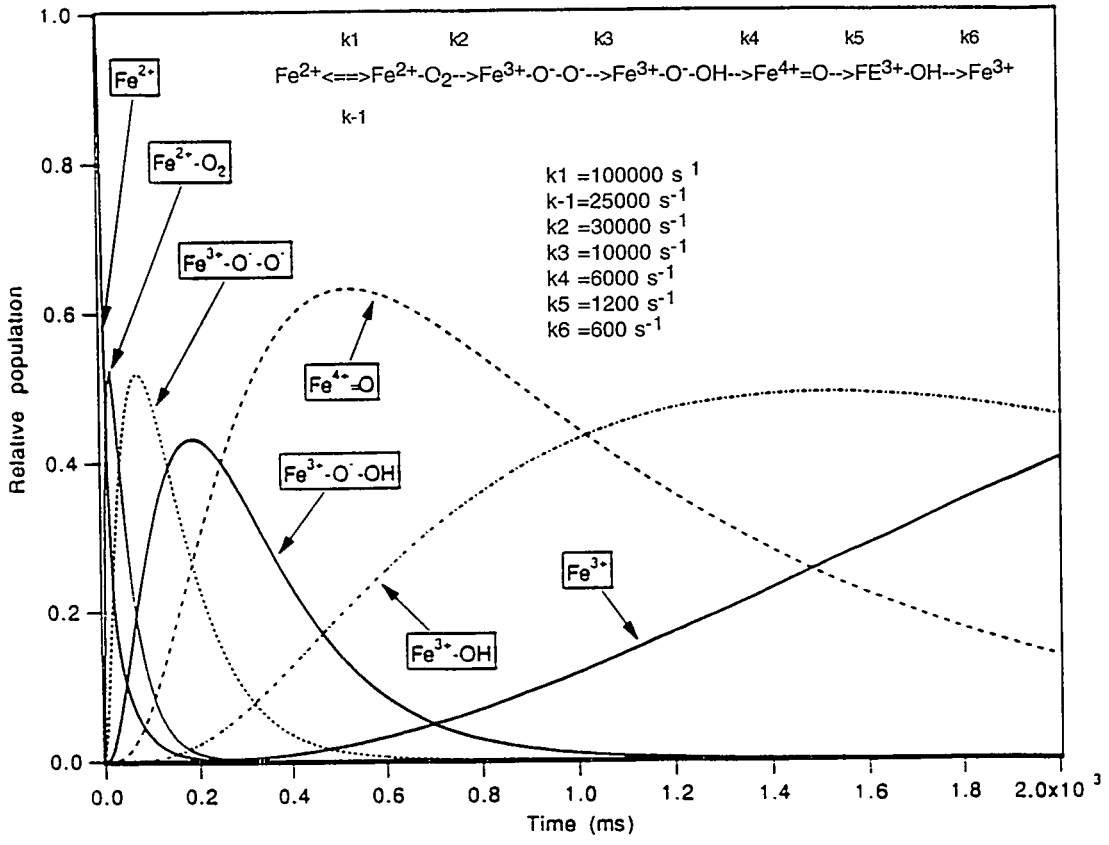


Fig. 4. Simulation of concentration-time profiles for several intermediates in the reaction mechanism of Fig. 3.

reaction chemistry. This point has been stressed by us as well as by the Göteborg group; the occurrence of high intermediate concentrations provides the potential for detailed characterization of O_2 activation and reduction chemistry in the enzyme active site.

Although it was not apparent at the time of the original work by Gibson and Greenwood (1993; Greenwood and Gibson, 1967), the physiological basis for proton control in oxidase can be rationalized on the basis of the fact that electron-transfer and O_2-reduction chemistry in cytochrome oxidase is coupled to proton pumping against the membrane potential in the mitochondrion. The free-energy drop through the protein from its reductant cytochrome *c* to its oxidant O_2 is on the order of 550 mV, and the enzyme has evolved so that a significant fraction of this free energy is conserved in the chemiosmotic gradient; that is, cytochrome oxidase acts as a redox-linked proton pump and contributes directly to the buildup of the membrane potential that ultimately drives the synthesis of ATP. Wikström has been the leader in demonstrating this function (Wikström, 1977, 1989). In his initial work, he showed that the proton-pumping stoichiometry was one proton translocated per electron transferred through the enzyme. Thus, in the four-electron reduction of O_2, the overall reaction can be written as

$$4\ \text{cyt}\ c^{2+} + O_2 + 8H^+_{in} \rightarrow$$

$$4\ \text{cyt}\ c^{3+} + 2H_2O + 4H^+_{out}\ , \qquad [3]$$

where H^+_{in} and H^+_{out} refer to protons on the matrix and cytosolic sides of the mitochondrial membrane, respectively. In his subsequent work, Wikström showed that, although the overall H^+/e^- stoichiometry is 1:1, the actual pump is coupled asynchronously to the electron-transfer reactions. Only the late intermediates in the reduction reaction, the peroxy and the ferryl forms of the enzyme in Fig. 3, are coupled to H^+ translocation; the free energy released in the one-electron reduction of each of these is used to drive the translocation of $2H^+$. Within this context, the basis for proton control in oxidase becomes clear: limiting the O_2-reduction chemistry at the level of protons provides a means by which to ensure that the H^+ pumping sites are loaded before the redox chemistry that drives the pump is triggered.

Wikström demonstrated the tight coupling that exists between proton translocation and O_2 chemistry by using an imposed pH and electrostatic potential to reverse electron flow through the enzyme (Wikström, 1981; Wikström and Morgan, 1992). Beginning with the fully oxidized enzyme, he has been able to produce forms of the enzyme equivalent (in electron stoichiometry) to the ferryl and peroxy species in Fig. 3. These correspond, respectively, to one- and two-electron reversal of the O_2 reduction chemistry. He refers to these species as F and P and has characterized them optically; in a collaborative effort with his group, we have obtained an initial set of Raman data that directly links

the forward reaction sequence in Fig. 3 and his reversed electron-transfer results (Varotsis, Wikström, and Babcock, in preparation).

F and P intermediates have also been generated by adding hydrogen peroxide to the oxidized form of the enzyme (Watmough *et al.*, 1994; Lauraeus *et al.*, 1993; Vygodina *et al.*, 1993). These species have distinctive optical characteristics that resemble closely those observed by Wikström in his reversed electron-transfer work. As implied above, however, the detailed molecular structures of these species and those in Fig. 3 are on tenuous ground at present. We have an overall kinetic framework within which to discuss oxidase O_2 chemistry, we have a good idea of electron stoichiometries, and the tools necessary to extend this are clearly available. Testing the proposed structures for various oxygen intermediates and establishing the molecular bases for their coupling to proton translocation remain important goals in oxidase work. Stable oxygen isotopes have played key roles in the progress that has been made in understanding dioxygen chemistry in heme proteins. Their continued availability will be crucial in expanding the insights that have been achieved to date.

Recent Developments in O_2 Chemistry in Non-Heme Iron Proteins

In the past 3 to 4 years, several groups, including those of Stubbe, Sjöberg, Lipscomb, Que, Münck, Lippard, and Sanders-Loehr, have shown that the nonheme iron class of enzymes, which includes ribonucleotide reductase (RNR), methane monooxygenase (MMO), and fatty acid Δ^9-desaturase, activates dioxygen in a fashion that is amenable to detailed characterization by spectroscopic methods (Bollinger *et al.*, 1991b; Ling *et al.*, 1994; Lee *et al.*, 1993a,b; Dong *et al.*, 1993; Liu *et al.*, 1994; Sahlin *et al.*, 1990). The kinetic constants for the O_2 chemistry are such that the reaction slows progressively as it occurs, and concentration/time profiles like those in Fig. 4 result; moreover, the rate constants are 2 to 3 orders of magnitude slower than those that we have detected in the cytochrome oxidase/O_2 reaction. As a consequence, fast-trapping methods are useful in these systems. Thus far, both RNR, which uses O_2 to generate a tyrosyl radical at Tyr^{122} in its R2 subunit during the iron-binding and catalytic activation process, and MMO, which catalyzes a P450-like oxygen atom insertion reaction into a methane C-H bond, have been studied in some detail. Although this work is in its early stages and definitive molecular structures have not yet been assigned to the species detected, it appears that the oxygen chemistry that occurs in these systems and in heme oxidases and peroxidases will have common themes. In MMO, for example, at least three intermediate species have been detected (Lee *et al.*, 1993a,b; Liu *et al.*, 1994), including a 420-nm absorber that has been designated intermediate Q and is thought to be the active hydroxylating species in the cycle. The iron valence and oxygen structure

in Q are, at present, being debated, as is the issue of whether oxygen insertion chemistry is concerted or stepwise. Nonetheless, observation of an activated oxygen species bound at a binuclear metal center in MMO is of considerable relevance to the putative peroxy species that we assigned in the binuclear oxidase center during the reaction sequence in Fig. 3.

Reactive intermediates have also been trapped in the RNR reaction, including one in which substantial coupling to an oxygen atom has been observed by using stable isotope substitution methods (Bollinger *et al.*, 1993a,b). The relationship between the intermediates observed in RNR and MMO is unclear now, but is likely to be established fairly quickly. Also unclear in the nonheme iron systems is the mechanistic basis for the kinetic properties of the O_2 reaction; that is, the basis for a kinetic scheme in which reaction rates diminish as the reaction proceeds. Whether these systems are subject to proton control, as described above for oxidase chemistry, is not known, although this scenario would explain the kinetic behavior. In any event, why these enzymes would follow such an apparently risky reaction path, with reactive oxygen species accumulating to significant extents during the sequence, is an important mechanistic question to address and is of immediate relevance to O_2 activation chemistry in cytochrome oxidase, in P450s, and in peroxidases.

Model compound approaches for studying O_2 activation have provided considerable insight and guidance in efforts to understand the chemistry that occurs in enzyme-active sites (for reviews, see Watanabe and Groves, 1992; Que and True, 1990). We, for example, have drawn heavily on both Chang's and Balch's work (Young and Chang, 1985; LaMar *et al.*, 1983) on oxy and ferryl oxo complexes of model heme compounds in characterizing iron-oxygen stretching frequencies in these species (Oertling *et al.*, 1990). The heme iron side-bound peroxy model species synthesized and characterized by Valentine and her coworkers (McCandlish *et al.*, 1980), the oxo-bridged Fe/Cu adducts of Karlin and of Holm (Lee and Holm, 1993; Karlin *et al.*, 1994), and the binuclear copper and iron dioxygen adducts that have been prepared recently (Dong *et al.*, 1993; Kitajima *et al.*, 1992) are all of potential importance to oxidase chemistry. In our own work, we continue our collaboration with Professor Chang and are currently characterizing hydrogen-bonding effects to metalloporphyrin oxo species. Determining whether the vibrations detected in the 800-cm^{-1} region during the oxidase/O_2 reaction arise from differentially hydrogen-bonded species is an important goal in our continuing work, and model compound work is essential.

Conclusions

Recent advances in understanding dioxygen activation in biology have been substantial. In metal-based enzymes,

a variety of intermediates in the process have been detected and identified. This work has relied on both magnetic-resonance and vibrational spectroscopic methods and has benefited significantly from the ready availability of stable oxygen isotopes. In fact, much of the progress described above would not have been possible in their absence. From this work, detailed mechanisms for oxygen-reaction chemistry can be determined. This development is essential for understanding biological chemistry at a fundamental, molecular level. Moreover, oxygen-derived free radicals that are generated in the biological milieu have increasingly been implicated in triggering a variety of pathological states. The fact that we are now beginning to develop a detailed view of oxygen metabolism promises that we will soon have methods by which to understand and eventually control the chemistry so that deleterious generation of radical species can be minimized.

Acknowledgements

The research in the author's laboratory that is described in this article is being supported by NIH GM25480.

References

Babcock, G.T., Jean, J. M., Johnston, L.N., Palmer, G., and Woodruff, W.H. (1984) *J. Am. Chem. Soc. 106*, 8305-8306.

Babcock, G.T., Jean, J. M., Johnston, L.N., Palmer, G., and Woodruff, W.H. (1985) *J. Inorg. Biochem. 23*, 243-251.

Babcock, G.T. and Wikström, M. (1992) *Nature 356*, 301-309.

Babcock, G.T., El-Deeb, M.K., Sandusky, P.O., Whittaker, M.M., and Whittaker, J.W. (1992) *J. Am. Chem. Soc. 114*, 3727-3734.

Babcock,G.T. and Varotsis, C. (1993) *SPIE Vol. 1890, Biomolecular Spectroscopy (III)*, pp. 104-113.

Blair, D.F., Witt, S.N., and Chan, S.I. (1985) *J. Am. Chem. Soc. 107*, 7389-7399.

Bollinger, J.M., Jr., Stubbe, J., Huynh, B.H., and Edmondson, D. E. (1991a) *J. Am. Chem. Soc. 113*, 6289-6291.

Bollinger, J.M., Jr., Edmondson, D.E., Huynh, B.H., Filley, J., Norton, J.R., and Stubbe, J. (1991b) *Science 253*, 292-298.

Chance, B., Saronio, C., and Leigh, J.S., Jr. (1975) *J. Biol. Chem. 250*, 9226-9237.

Clore, G.M., Andreasson, L.E., Karisson, B., Aasa, R., and Malmström, B. G. (1980) *Biochem. J. 185*, 139-154.

Dong, Y., Menage, S., Brennan, B.A., Elgren, T.E., Jang, H.G., Pearce, L.L., and Que, L., Jr. (1993) *J. Am. Chem. Soc. 115*, 1851-1859.

George, P. (1965) in *Oxidases and Related Redox Systems*, Vol. 1., T.E. King, H.S. Mason, and M. Morrison, eds., Wiley, New York, pp. 3-33.

Gibson, Q. and Greenwood, C. (1963) *Biochem. J. 86*, 541-554.

Greenwood, C. and Gibson, Q.H. (1967) *J. Biol. Chem. 242*, 1782-1787.

Guengerich, F. P. (1992) *FASEB J. 6*, 667-668.

Hallen, S. and Nilsson, T. (1992) *Biochemistry 31*, 11853-11859.

Han, S., Ching, Y.-C., and Rousseau, D.L. (1990a) *Proc. Natl. Acad. Sci. USA 87*, 2491-2495.

Han, S., Ching, Y.-C., and Rousseau, D.L. (1990b) *Proc. Natl. Acad. Sci. USA 87*, 8408-8412.

Han, S., Ching, Y.-C., and Rousseau, D.L. (1990c) *Nature 348*, 89-90.

Karlin, K.D., Nanthakumar, A., Fox, S., Murthy, N.N., Ravi, N., Huynh, B.H., Orosz, R.D., and Day, E.P. (1994) *J. Am. Chem. Soc. 116*, 4753-4763.

Kitajima, N., Fujisawa, K., Fujimoto, C., Moro-oke, Y., Hashimoto, S., Kitagawa, T., Toriumi, K., Tatsumi, K., and Nakamura, A. (1992) *J. Am. Chem. Soc. 114*, 1277-1291.

LaMar, G.N., deRopp, J.S., Latos-Grazynski, L., Balch, A.L., Johnson, R.S., Smith, K.B., Parish, D.W., and Cheng, R.J. (1983) *J. Am. Chem. Soc. 105*, 782-787.

Lauraeus, M., Morgan, J.E., and Wikström, M. (1993) *Biochemistry 32*, 2664-2670.

Lee, S.C. and Holm, R.H. (1993) *J. Am. Chem. Soc. 115*, 5833-5834.

Lee, S.-K., Fox, B.G., Nesheim, J.C., Froland, W.A., Münck, E., and Lipscomb, J.D. (1993a) *J. Inorg. Biochem. 43*, 300 (1993).

Lee, S.-K., Fox, B.G., Froland, W.F, Lipscomb, J.D., and Münck, E. (1993b) *J. Am. Chem. Soc. 115*, 6450-6451.

Lee, S.-K., Nesheim, J.C., and Lipscomb, J.D. (1993c) *J. Biol. Chem. 268*, 21569-21577.

Ling, J., Sahlin, M., Sjöberg, B.-M., Loehr, T.M., and Sanders-Loehr, J. (1994) *J. Biol. Chem. 269*, 5596-5601.

Lippard, S.J. and Nordlund, P. (1993) *Nature 366*, 537-543.

Liu, K. E., Wang, D., Huynh, B. H., Edmondson, D. E., Salifoglou, A., and Lippard, S. J. (1994) *J. Am. Chem. Soc. 16*, 7465-7466.

McCandlish, E., Miksztal, A.R., Nappa, M., Sprenger, A.Q., Valentine, J.S., Strong, J.D., and Spiro, T.G. (1980) *J. Am. Chem. Soc. 102*, 4268-4271.

Oertling, W.A., Kean, R.T., Wever, R., and Babcock, G.T. (1990) *Inorg. Chem. 29*, 2633-2645.

Ogura, T., Takahashi, S., Shinzawa-Itoh, K., Yoshikawa, S., and Kitagawa, T. (1990) *J. Am. Chem. Soc. 112*, 5630-5631.

Ogura, T., Takahashi, S., Shinzawa-Itoh, K., Yoshikawa, S., and Kitagawa, T. (1991) *Bull. Chem. Soc. Jpn. 64*, 2901-2907.

Ogura, T., Takahashi, S., Hirota, S., Shinzawa-Itoh, K., Yoshikawa, S., Appelman, E.H., and Kitagawa, T. (1993) *J. Am. Chem. Soc. 115*, 8527-8536.

Oliveberg, M., Brzezinski, P., and Malmström, B.G. (1989) *Biochim. Biophys. Acta 977*, 322-328.

Oliveberg, M., Hallen, S., and Nilsson, T. (1991) *Biochemistry 30*, 436-440.

Picot, D., Loll, P.J., and Garavito, R.M. (1994) *Nature 367*, 243-249.

Que, L., Jr. and True, A.E. (1990) in *Progress in Inorganic Chemistry: Bioinorganic Chemistry*, Vol. 38, S.J. Lippard, ed., John Wiley & Sons, New York, pp. 97-200.

Reichard, P. and Ehrenberg, A. (1983) *Science 221*, 514-519.

Sahlin, M., Sjöberg, B.-M., Backes, G., Loehr, T., and Sanders-Loehr, J. (1990) *Biochem. Biophys. Res. Commun. 167*, 813-818.

Varotsis, C., Woodruff, W.H., and Babcock, G.T. (1989a) *J. Am. Chem. Soc. 111*, 6439-6440.

Varotsis, C., Woodruff, W.H., and Babcock, G.T. (1989b) *J. Am. Chem. Soc. 112*, 1297.

Varotsis, C. and Babcock, G.T. (1990) *Biochemistry 29*, 7357-7362.

Varotsis, C., Woodruff, W.H., and Babcock, G.T., (1990) *J. Biol. Chem. 265*, 11131-11136.

Varotsis, C., Zhang, Y., Appelman, E.H., and Babcock, G.T. (1993) *Proc. Natl. Acad. Sci. USA 90*, 237-241.

Vygodina, T.V., Schmidmaier, K., and Konstantinov, A.A. (1993) *Biol. Mem. 6*, 883-906.

Watanabe, Y. and Groves, J.T. (1992) in *The Enzymes*, S. Sigman, ed., Academic, San Diego, pp. 405-452.

Watmough, N.J., Cheesman, M.R., Greenwood, C., and Thomson, A.J. (1994) *Biochem. J. 300*, 469-475.

Wikström, M. (1977) *Nature 266*, 271-273 .

Wikström, M. (1981) *Proc. Natl. Acad. Sci. USA 78*, 4051-4054.

Wikström, M. (1989) *Nature 338*, 776-778.

Wikström, M. and Morgan, J.E. (1992) *J. Biol. Chem. 267*, 10266-10273 .

Young, R. and Chang, C.K. (1985) *J. Am. Chem. Soc. 107*, 898-909.

Use of Specifically ^{15}N-Labeled Histidine to Study Structures and Mechanisms Within the Active Sites of Serine Proteinases

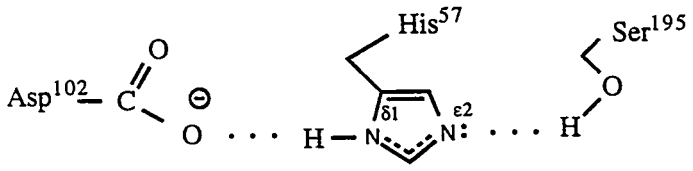
William W. Bachovchin

Department of Biochemistry
Tufts University School of Medicine
Boston, MA 02111

The current emphasis in biological NMR work is on determining structures of biological macromolecules in solution. This emphasis is appropriate because NMR is the only technique capable of providing high-resolution structures that are comparable to those of x-ray crystallography for molecules in solution. This structural knowledge is immensely valuable and is needed in many areas of investigation. However, as valuable as such structural knowledge is, it never provides all the answers; a structure often reveals more questions than answers.

Production of a three-dimensional structure, however, is not the limit of NMR's potential for contributing to our understanding of biological macromolecules. NMR spectroscopy is an incredibly versatile investigatory tool. A host of NMR experiments can provide information to complement and enhance the picture provided by structural methodologies. The objective of this paper is to briefly summarize one such NMR experiment and review how it has contributed to—and continues to enhance—our knowledge of a class of enzymes that have been very well characterized by structural methodologies. This NMR experiment involves extracting the information contained in the behavior of ^{15}N signals of histidine residues. The structurally well characterized enzymes to which we have applied this method are the serine proteinases.

Serine proteinases all contain a catalytic triad of Asp-His-Ser residues at their active sites, as illustrated on the next page. Enzymes belonging to this family are extraordinarily numerous. Although they all catalyze the same basic reaction



Catalytic triad

(that is, the hydrolysis of peptide bonds), the serine proteinases are very diverse with respect to the biological functions nature has recruited them to perform. For example, thrombin, factors IX, X, and XI function in blood clotting. Dipeptidyl amino peptidase IV, also known as CD26, is found on CD4+ T cells, where it apparently functions as a co-stimulatory molecule and contributes to regulation of the immune response (Tanaka *et al.*, 1993; Flenke *et al.*, 1991; Kubota *et al.*, 1992). Serine proteinase functions are not always helpful to humans. For instance, IgA proteinases, produced by pathogenic strains of *Neisseria gonorrhoeae* and *Hemophilus influenzae*, are thought to contribute to the virulence of these human pathogens by cleaving human IgA in the hinge region (Bachovchin *et al.*, 1990). This cleavage separates the antigen binding function from the effector function, thereby disarming these antibodies, which are the first line of defense against invading microbes.

Despite the fact that hundreds of serine proteinases have already been identified, we continue to discover new ones. More often than not, these new enzymes have key biological functions that make them potential targets for drug-design efforts. The DP IV and the IgA proteinases mentioned above fall into this category. Although one may question why nature has chosen to employ serine proteinases so often and to adapt them to so many different functions, their importance to biological systems is clear. Understanding proteinase structure, specificity, and mechanisms is therefore highly desirable, because such knowledge should contribute to the development of new therapeutic agents for many different diseases.

The structure and mechanism of serine proteinases have been studied intensively for decades. Chymotrypsin, a pancreatic digestive serine proteinase, was only the third protein structure solved by x-ray crystallography (Blow *et al.*, 1969). Since then, many high-resolution structures of serine proteinases and their complexes with various inhibitors have been solved, and new structures continually appear—testifying to the importance of this class of enzymes. Despite the wealth of structural information we have accumulated for serine proteinases, many aspects of these enzymes are not well understood. For example: How do these enzymes convert substrate binding energy into catalytic power? This phenomenon, which they all exhibit and which may hold the key to understanding specificity, is still essentially a mystery. Nor is there a satisfactory explanation for why

substituting Cys for the active site Ser destroys the catalytic activity of most members of this family, yet nature seems to have carried out this same experiment in the picornavirus family of proteinases without ill effect (Allaire *et al.*, 1994).

The ^{15}N NMR approach to extracting information from histidyl residues is complimentary to x-ray and NMR structural analysis methodologies in several ways. Unlike NMR structural analysis methodologies, which essentially rely on information provided by NOE and coupling constant measurements, the ^{15}N NMR experiment relies to a large extent on chemical shifts, although linewidths and coupling constants to the attached protons also contribute. Specific information supplied by this approach is in large measure complementary to that supplied by structural methodologies. For example, the location of protons, the existence and relative strengths of H-bonds involving the histidyl residue, the tautomeric structure of the imidazole ring, and the interaction of the histidyl residue with other atoms such as metals or boron are revealed by this methodology. And they are revealed with striking clarity (Bachovchin and Roberts, 1978; Bachovchin, 1986; Bachovchin *et al.*, 1988; Farr *et al.*, 1989; Farr-Jones *et al.*, 1993; Blomberg *et al.*, 1977). This is true even when the interactions with His are too short-lived (because of rapid exchange) to be detected by x-ray diffraction or NMR structural analysis methodologies. Several other factors make this experiment useful and complementary to NMR structural methodologies.

- The efficiency with which the crucial information can be obtained—often a few or even a single one-dimensional NMR experiment is sufficient to determine the structural features of interest
- The experiment can be applied to molecules with molecular weights far greater than those now feasible for NMR structural analysis methodologies.

Framework for Interpreting ^{15}N Chemical Shifts

Figure 1 summarizes a framework that we have developed over the years for interpreting chemical shifts of imidazole ring nitrogen atoms (Bachovchin, 1986; Bachovchin *et al.*, 1988; Farr-Jones *et al.*, 1993). It is important to stress that this framework applies only to histidines in aqueous environments (Farr-Jones *et al.*, 1993; Shuster and Roberts, 1979). Briefly, the nitrogen atoms in an imidazole ring can adopt three canonical forms: pyrrole-like (>N-H), pyridine-like (⩾N:), and pyrrole-like in a protonated imidazole ring (+⩾N-H). The chemical shifts of pyrrole- and pyridine-like nitrogen differ by ~80 ppm. A pyrrole-like nitrogen in a protonated imidazole ring resonates ~10 ppm downfield from its position in a neutral imidazole ring and thus differs from a pyridine-like nitrogen by ~70 ppm. The ^{15}N shifts, therefore, are never ambiguous about whether or not a nitrogen atom has an attached proton. The ^{15}N shifts also respond to H-bonding: the H-donating >N-H- and +⩾N-H-type

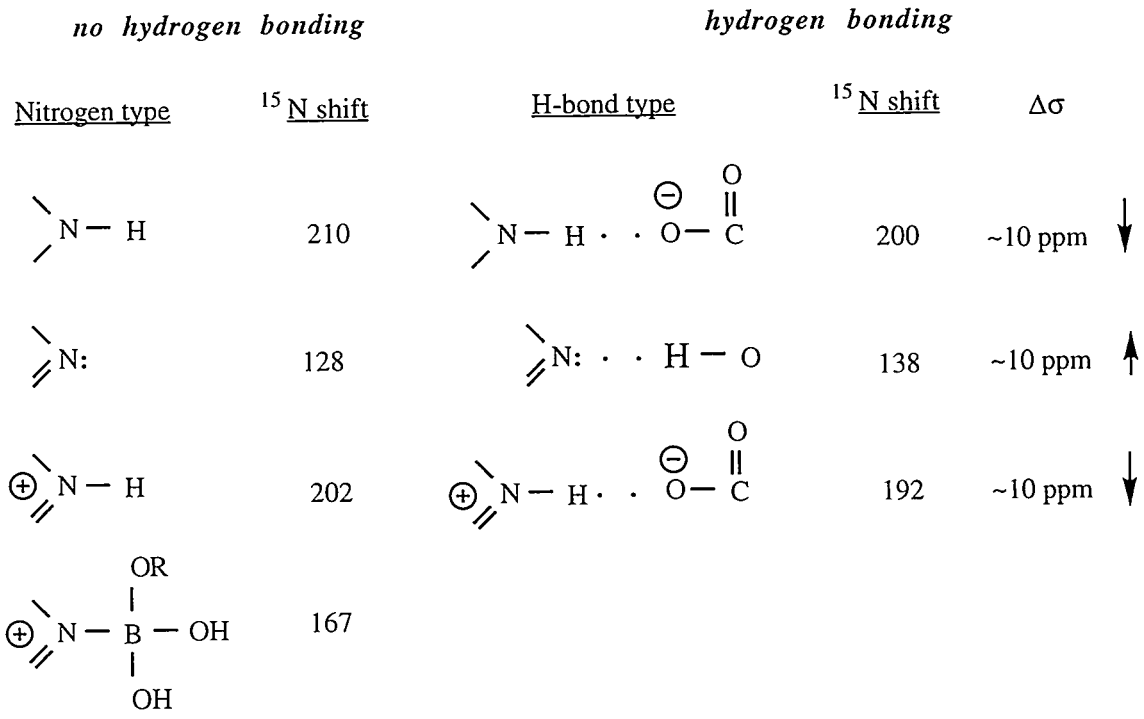


Fig. 1. Typical ^{15}N chemical shifts of the canonical types of imidazole ring nitrogen atoms and the effect of hydrogen bonding on these chemical shifts.

nitrogens are moved downfield as much as 10 ppm; the H-accepting ⩾N:-type nitrogens are moved upfield by the same amount.

Nitrogen atoms in 4-substituted imidazoles such as histidine are magnetically non-equivalent. Each gives rise to a signal that can—and often does—represent a weighted average of the shifts of all three canonical nitrogen types (depending on pH) because of the fast rate at which equilibrium is established between the protonated and the two neutral tautomeric species, as is illustrated in Fig. 2.

Shifts for the "pure" +⩾N-H-type nitrogens are easy to obtain because the protonated species dominates at low pH. The shifts for the pure >N-H- and ⩾N:-type nitrogens are more difficult to determine because the neutral tautomers are usually both present and rapidly equilibrating at room temperature in a pH > 6.0 solution. Some effort has gone into estimating these chemical shifts from model system studies (Blomberg *et al.*, 1977; Shuster and Roberts, 1979; Roberts *et al.*, 1982; Alei *et al.*, 1980). However, values given in Fig. 2 are from direct

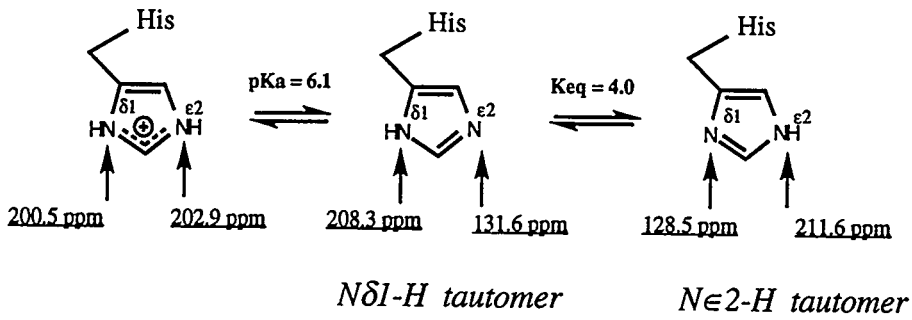


Fig. 2. Acid-base and tautomeric behavior of the imidazole ring of histidine and corresponding ^{15}N chemical shifts in aqueous solution.

observations of the individual tautomers of histidine itself at -60°C, where exchange is slow on the NMR time scale (Farr-Jones *et al.*, 1993).

Resting Enzyme

Figure 3 compares the ^{15}N shifts of the active-site histidyl residue in the catalytic triad of α–lytic protease at the low-pH limb of its titration curve to the shifts of the protonated form of histidine (Bachovchin and Roberts, 1978).

The shifts show that both nitrogen atoms of His^{57} have directly bonded protons and, thus, that the imidazole ring is protonated. Both nitrogen atoms are therefore +⩾N-H-type nitrogens. The position of the Nδ1 resonance at 191.6 ppm—more than 10 ppm downfield from the canonical value for a +⩾N-H-type nitrogen (Fig. 1) and ~9 ppm downfield from the shift of this nitrogen of histidine itself when protonated—reveals the presence of the strong H-bond to Asp^{102} (Fig. 3).

Figure 4 compares the ^{15}N shifts of His^{57} at the high-pH limb of its titration curve to the shifts of the Nδ1-H tautomer of histidine. The high-field position of Nδ1 relative to Nε2 makes it immediately obvious that His^{57} exists predominately as the Nδ1-H tautomer. These shifts, however, contain more information: notice that Nδ1 is ~10 ppm downfield of a canonical >N-H type, whereas Nε2 is 10 ppm upfield of a canonical ⩾N:-type nitrogen. There are two possible explanations for these shifts.

- Tautomeric exchange: ~0.15 mole fraction of the Nε2-H tautomer averages with a predominate Nδ1-H tautomer.

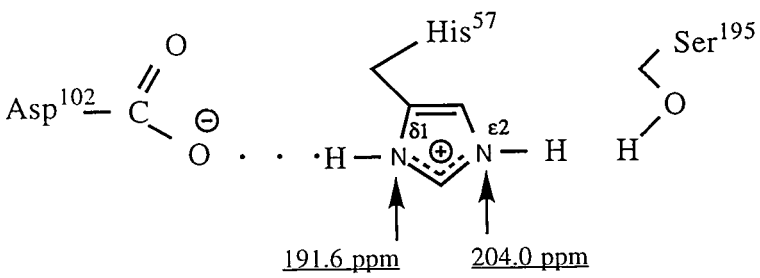
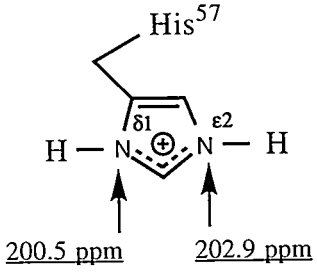


Fig. 3. Comparison between the ^{15}N shifts of His^{57} of α-lytic proteinase at low pH and those of histidine free in solution at the same pH.

α-lytic proteinase at pH 4.0



histidine at pH 4.0

- Both nitrogens are strongly H-bonded. A strong H-bond to Asp^{102} would be expected to move Nδ1 ~10 ppm *downfield*, whereas a strong H-bond from Ser^{195} to Nε2 would be expected to move it ~10 ppm *upfield.*

Inactivation of α–lytic protease with PMSF, or with DIFP or other such reagents that form covalent adducts with Ser^{195}, resolves the ambiguity. PMSF inactivation induces the Nε2 resonance to move downfield ~10 ppm to a position characteristic of that of canonical ⩾N:-type nitrogen; Nδ1 moves very little (Fig. 4). This experiment directly reveals the existence of the His-Ser H-bond and shows that it is responsible for inducing ~10-ppm upfield chemical shift change on Nε2 of His^{57} (Bachovchin, 1986) while establishing that the second explanation above is correct. With a shift of ~128 ppm in PMSF inhibited enzyme, Nε2 accounts for one full pyridine type (⩾N:) nitrogen. Nδ1 must therefore be fully a >N-H type because a neutral imidazole ring must contain, on average, one pyridine- and one pyrrole-type nitrogen. The position

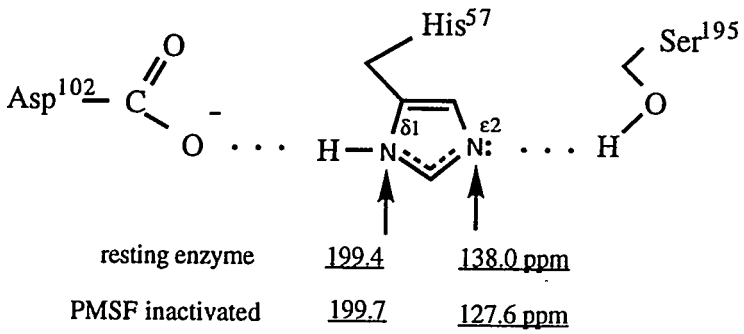
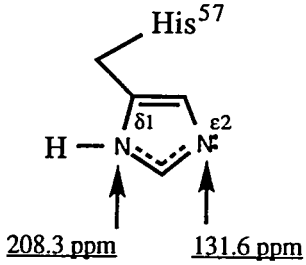


Fig. 4. Comparison of the ^{15}N shifts of His^{57} at high pH and those of the Nδ1-tautomer of free histidine.

α-lytic proteinase at pH 9.0



Nδ1-H tautomer of histidine

of Nδ1 at 199 ppm, ~10 ppm downfield from that of a canonical >N-H-type nitrogen at ~210 ppm can be ascribed to the strong H-bond to Asp^{102}.

Thus, the behavior of the ^{15}N shifts of His^{57} reveal the presence of strong H bonds between it and Asp^{102} on one side and Ser^{195} on the other. The ^{15}N signals also show that His^{57} exists exclusively as the Nδ1-H tautomer when neutral. The location of the proton in the triad has aroused substantial interest and debate because of its mechanistic significance. The ^{15}N shifts of His^{57} provided the first unambiguous experimental answer to this question (Bachovchin and Roberts, 1978), and they remain the strongest and clearest evidence that the proton is indeed located on Nδ1 of His^{57}.

Direct observation of the ^{15}N signals also provides information about the exchange dynamics of the protons directly bonded to imidazole ring nitrogen atoms. Although the shifts indicate that His^{57} exists exclusively as the Nδ1-H tautomer, the ^{15}N-H coupling is not observed at room temperature, demonstrating that the Nδ1 proton is in fast exchange with

solvent with respect to the ~90-Hz coupling constant. The spin coupling, however, is resolved at 5°C (Bachovchin, 1985, 1986; Bachovchin *et al.*, 1988).

Boronic Acid and Peptide Boronic Acid Inhibitor Complexes

Peptide boronic acids are extraordinarily potent inhibitors of serine protease. The K_i values of these inhibitors are typically in the nanomolar range and often reach into the picomolar range. This high affinity is widely attributed to the boronyl group's ability to mimic the transition state of the enzyme-catalyzed reaction. Peptide boronic acid complexes with serine proteinases are therefore of considerable interest.

The behavior of ^{15}N NMR signals from the active-site histidine has revealed that boronic acid inhibitors form at least two distinct types of complexes with serine proteinases—referred to here as serine and histidine adducts (Bachovchin *et al.*, 1988).

In a typical serine adduct (illustrated in Fig. 5), both ring nitrogens are bonded to protons, as demonstrated by the high-field position of the ^{15}N signals and by the observation of one-bond ^{1}H-^{15}N coupling. This coupling is well resolved, even at room temperature, demonstrating that the ^{1}H exchange is markedly slower in this complex than in the resting enzyme. Both ring nitrogen atoms resonate ~10 ppm downfield from the position of a canonical pyrrole-type nitrogen in a protonated imidazole ring—very similar to the position of the Nδ1-H proton resonance in resting enzyme at low pH (Fig. 3). The position of these nitrogen signals reveals that both N-H groups of His^{57} are engaged in strong H-bonds as proton donors in this complex. This conclusion predicts there should be

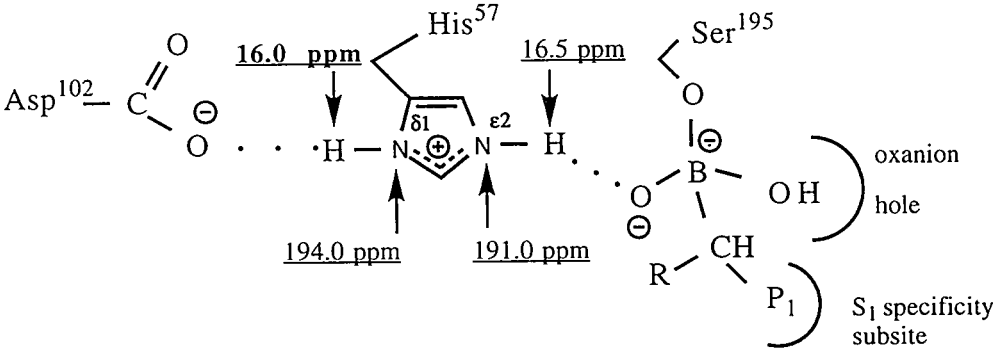


Fig. 5. ^{15}N and ^{1}H chemical shifts and structure of a typical serine adduct complex.

two low-field ^{1}H resonances instead of the usual one, and the prediction is borne out. One ^{1}H signal is observed at ~16.5 ppm, the other at 16.1 ppm, from the protons on Nε2 and Nδ1, respectively (Fig. 5) (Bachovchin *et al.*, 1988). The observation of the second low-field proton signal confirms the interpretation of the ^{15}N chemical shifts.

Interestingly, both the ^{15}N and ^{1}H signals on the serine side of His^{57} (Nε2) resonate even further downfield than their counterparts on the Asp^{102} side (Nδ1). This suggests that the hydrogen bond is even stronger than the H-bond to Asp^{102} in these complexes (Fig. 5). This is remarkable because the Asp-His hydrogen bond is widely considered an exceptionally strong hydrogen bond.

Both the ^{15}N and the ^{1}H signals in serine adduct complexes are pH-independent over the range of 4.0 to 10.0. With some especially high affinity inhibitors, the pH-independence extends to pH 11.0 (Bachovchin *et al.*, 1988), which demonstrates that the active-site histidine does not titrate and remains protonated and positively charged as the imidazolium ion for as long the boronic acid inhibitor remains bound. The electrostatic interaction set up between imidazolium ion of His^{57} and the negatively charged boronyl group is likely to be an important component of the high affinity of these inhibitors. Boronic acids that form serine adducts with α-lytic protease include Boc-Ala-Pro-boroVal, MeOSuc-Ala-Ala-Pro-boroVal, and MeOSuc-Ala-Ala-Pro-boroAla (see Table 1).

In a typical histidine adduct complex (summarized in Fig. 6), Nδ1 carries a proton, as indicated by its high-field chemical shift and the observation of one-bond ^{1}H coupling. The ^{15}N resonance of Nε2, however, is found at ~167 ppm,

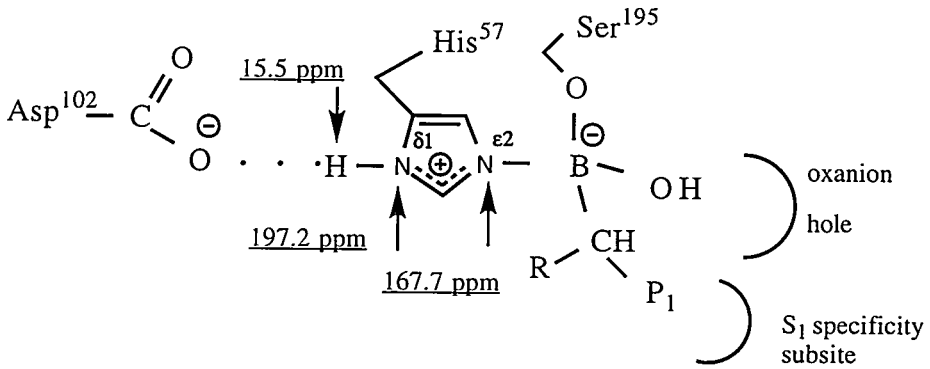


Fig. 6. ^{15}N and ^{1}H chemical shifts and structure of a typical histidine adduct complex.

which is about halfway between that of canonical >N-H-type and ⩾N:-type nitrogens. The Nε2 resonance is pH-independent—ruling out acid/base exchange and shift averaging as the explanation for its intermediate chemical shift. Model system studies showed that such an intermediate ^{15}N chemical shift value is characteristic of imidazole nitrogens bonded to the boron atom of a boronyl group (Bachovchin *et al.*, 1988). X-ray crystallographic studies have confirmed the existence of an N-B bond in the histidine adduct complex between α-lytic protease and MeOSuc-Ala-Ala-Pro-boroPhe (Bone *et al.*, 1989). In addition, a histidine adduct for elastase with CBZ-Ala-Ile has been observed by x-ray crystallography (Takahashi *et al.*, 1989).

Histidine adduct complexes are easily distinguished from serine adduct complexes by their low-field proton spectra (Bachovchin *et al.*, 1988; Tsilikounas *et al.*, 1992). Unlike serine adducts, which exhibit two resonances, histidine adducts exhibit only a single low-field ^{1}H resonance. In addition, the position of this resonance is approximately halfway (~15.5 ppm) between the high- and low-pH positions of the low-field proton in the resting enzyme (13.8 and 17.2 ppm, respectively, for α-lytic protease). The two resonances exhibited by serine adduct complexes are found further downfield.

Boronic acid inhibitors that form histidine adduct complexes with α-lytic protease include MeOSuc-Ala-Ala-Pro-boroPhe, Boc-Ala-Pro-D-boroVal, benzene boronic acid, and boric acid itself. Table 1 contains a summary of the types of complex α-lytic protease forms with various boronic acid inhibitors.

The results obtained thus far indicate that for a boronic acid to form a transition-state-like serine adduct with α-lytic protease requires that it be an analog of a "good" substrate. Or, said another way: for a boronic acid to form a serine adduct, it must be able to favorably occupy the specificity subsites while simultaneously presenting the boronyl group at the "transition-state-site."

Is this phenomenon general? Do other serine proteinases also require occupancy of specific subsites before permitting boronic acid inhibitors to form serine adducts? Or is it peculiar to α-lytic protease? To address this question, we examined boronic-acid-inhibited complexes of chymotrypsin, trypsin, trypsinogen, and elastase (our results are summarized in Table 1) and found that the phenomenon is general—at least among the trypsin family proteinases. The details differ from one enzyme to another, which likely reflects differences in specificities. The data summarized in Table 1 were obtained largely using the low-field proton spectra diagnostically for serine and histidine adducts.

Especially provocative is the discovery that Ac-boroAla and Ac-boroVal form histidine adducts with α-lytic protease (Table 1). The preferred cleavage sites

of α-lytic protease are after Ala and Val residues. These inhibitors should therefore be able to favorably occupy the S_1 subsite while the boronyl group simultaneously forms a transition-state-like adduct with Ser^{195}. Thus engaged, essentially nothing is left of these inhibitors that could interact unfavorably with the enzyme to counteract the presumably highly favorable interactions in S_1 and in the "transition-state binding site." Why, then, do these inhibitors not form transition-state-like serine adducts? The addition of one more amino acid to form the dipeptide boronic acid Ac-Pro-boroVal improves the inhibitor's affinity for α-lytic protease ~500-fold (Bachovchin *et al.*, 1988) and switches the complex from a

Table 1. Structures of Boronic Acid Complexes with Serine Proteases Determined by One-Dimensional NMR Spectroscopy

Serine Adducts	Histidine Adducts
α-Lytic Protease	
Boc-Ala-Pro-boroVal	MeOSuc-Ala-Ala-Pro-boroPhe
MeOSuc-Ala-Ala-Pro-boroVal	Boc-Ala-Pro-D-boroVal
MeOSuc-Ala-Ala-boroAla	benzene boronic acid
Ac-Pro-boroVal	boric acid
	Ac-boroAla
	Ac-boroVal
	H-boroAla
	H-boroVal
Chymotrypsin	
MeOSuc-Ala-Ala-Pro-boroPhe	MeOSuc-Ala-Ala-Pro-D-boroPhe
Ac-boroPhe	methane boronic acid
	boric acid
Trypsin	
H-D-Val-Leu-boroArg	boric acid
Ac-Gly-boroArg	methane boronic acid
	butane boronic acid
	triethanolamine boronic acid
Trypsinogen	
H-D-Val-Leu-boroArg	H-D-Val-Leu-boroArg
Ac-Gly-boroArg	Ac-Gly-boroArg
	boric acid
	methane boronic acid
	butane boronic acid
	triethanolamine boronic acid
Elastase	
MeOSuc-Ala-Ala-boroAla	MeOSuc-Ala-Ala-Pro-boroPhe
	benzene boronic acid

histidine to a serine adduct (Farr-Jones, Tsilokounas, and Bachovchin, manuscript in preparation). The addition of a third amino acid to form the tripeptide boronic acid inhibitor Ac-Ala-Pro-boroVal produces an additional ~10,000-fold increase in binding affinity (Bachovchin *et al.*, 1988) (Table 2).

This phenomenon is very reminiscent of the widely observed phenomenon in serine protease substrates as the peptide chain is elongated. For example, Ac-Ala-NH_2 is a very poor substrate for α-lytic protease (Bauer *et al.*, 1981). The addition of Pro in position 2 (P^2), then Ala in P^3, produces successively dramatically better substrates (Fig. 3). The improvement is mostly in k_{cat} rather than in K_m, which leads to the conclusion that serine proteinases convert binding energy from the nonreactive portion of the substrate molecule (that is, P^2 and P^3) into transition-state-stabilization energy.

There are two general mechanisms through which the substrate binding energy might be converted into catalytic power: strain and induced fit. At present, the strain argument seems to be favored. This argument states that the binding interaction in the extended subsites is used to "pay for" the unfavorable interaction between the planar substrate bond to be cleaved and a binding site on the enzyme designed to bind and stabilize a tetrahedral transition state. The net effect is to strain the planar peptide bond toward the tetrahedral transition-state structure.

The improvements in binding affinity, as well as the switch from histidine to serine adducts as the peptide chain of the boronic acid inhibitor is elongated, suggest the operation of the same type of mechanisms; binding energy from interactions in extended specificity

Table 2. Comparison of the Kinetic Parameters of a Homologous Series of Substrates and Peptide Boronic Acid Inhibitors

Substrate	$k_{cat}s^{-1}M^{-1}$	Δk_{cat}	K_m mM	ΔK_m
Ac-Ala-NH_2	0.0004		300	
Ac-Pro-Ala-NH_2	0.07	200-fold	110	3-fold
Ac-Ala-Pro-Ala-NH_2	0.7	10-fold	23	4-fold

Inhibitor	K_i M	ΔK_1	Adduct Type
Ac-boroVal	1.4 x 10-3		His
Ac-Pro-boroVal	3.3 x 10-6	~500-fold	Ser
Ac-Ala-Pro-boroVal	3.5 x 10-10	~10,000-fold	Ser

The upper data are from Bauer *et al.* (1981); the lower set is from Kettner et al.(1988).

subsites is being converted into transition-state-stabilization energy. In this case, transition-state stabilization refers to the the ability of the enzyme to bind to the boronyl group as reflected in the K_i and the type of complex formed. With the boronic acid inhibitors, however, the phenomenon cannot reflect the operation of a strain-like mechanism because the tetrahedral transition-state-like state for a boronyl group is not a high-energy one and thus would not need to be strained to assume this structure. The results are better explained by a mechanism involving substrate-induced changes in the enzyme that make it better able to interact with the transition-state-like structure of the boronyl group. Such an induced-fit mechanism can explain both the inhibitor and substrate results, whereas a strain mechanism can account for only the substrate results. We therefore believe that an induced-fit-type mechanism is likely to be operating in these enzymes. More work is needed, however, to establish this theory and to better define the mechanism. We are currently focusing our efforts in this direction.

Subtilisin BPN'

The work described thus far has been performed on serine proteinases that belong to the trypsin family of mammalian pancreatic serine proteinases. The subtilisins represent another, unrelated family of serine proteinases. They have no sequence or tertiary structural homology with the trypsin family, except that they contain the active site Asp-His-Ser triad. The existence of this family raises several interesting questions. For example, how closely do the H-bond interactions across the triad in the subtilisin family duplicate those in the trypsin family as measured by the ^{15}N chemical shift behavior and the tautomeric structure of the active site His? Do the rules outlined above regarding the binding of boronic acid to the trypsin family of proteinases also apply to the subtilisin family? These questions have important mechanistic implications because the catalytic mechanisms of these otherwise unrelated families is widely believed to be the same. We began comparative studies of subtilisin to address these questions. An interesting and potentially useful feature of the subtilisin family of enzymes for studies of transition-state-binding mechanisms is that the oxanion hole is constructed of amino acid sidechains rather than the backbone NHs found in the trypsin family. This feature permits mutagenesis experiments that are not possible with the trypsin family.

Subtilisin BPN' contains six histidine residues: three are buried, two are located on the surface, and one is at the active site. Figure 7 shows a representative one-dimensional ^{15}N spectra of ^{15}N-histidine-labeled subtilisin (M. Vincent, R. Day, and W. Bachovchin, manuscript in preparation).

All 12 histidine signals are present and resolvable. Interestingly, four of the histidine residues do not titrate over the range of 5.0 to 11. They remain neutral

and in single tautomeric structures—one in the Nδ1-H state and three in the Nε2-H state, as demonstrated by spectra of singly ^{15}Nδ1-His-labeled enzyme. Notice that one signal is found at 118 ppm; this is from an Nδ1 that is a ⩾N:-type nitrogen and is in one of the nontitrating histidines. The downfield position of this resonance indicates it is in a hydrophobic environment that resembles an organic solvent such a benzene (Farr-Jones *et al.*, 1993; Shuster and Roberts, 1979). This result suggests that the ^{15}N shifts of histidine residues may prove useful as probes of the hydrophobicity of protein interiors. Interestingly, one of the four nontitrating histidines must be a surface histidine because there are only three buried histidines.

The signals from the active-site histidine His^{64} (shown in Fig. 7A) were obtained under conditions in which the imidazole ring is protonated. The behavior of the ^{15}N signals of subtilisin His^{64}, summarized in Fig. 8, is remarkably similar to that described for α-lytic protease. The positions of the signals reveal the presence of a strong H-bond to Asp^{23} at both high and low pH as well as the presence of a strong H-bond to Ser^{221} at high pH. They also show that the neutral form of His^{64} exists exclusively as the Nδ1-H tautomer, just as His^{57} of α-lytic protease.

Addition of MeOSucAla-Ala-Pro-boroPhe, a good substrate analog inhibitor of subtilisin, induces the ^{15}N signals of His^{64} to behave much like those of His^{57} of α-lytic protease in serine adduct complexes. One

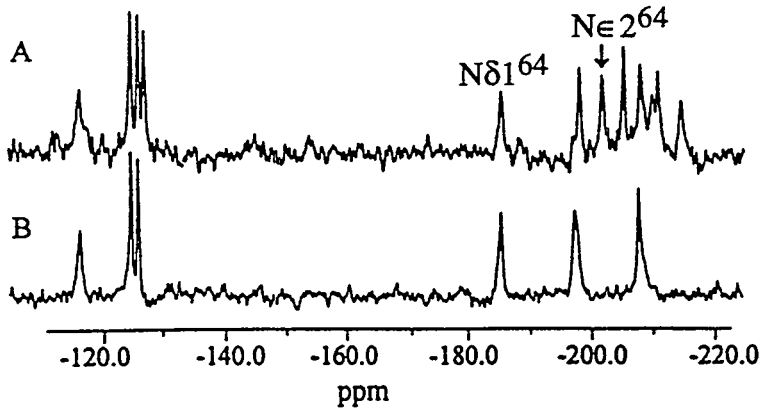


Fig. 7. ^{15}N NMR spectra of ^{15}N-His-labeled subtilisin BPN'. A: doubly ^{15}N ring nitrogen labeled; B: Singly ^{15}Nδ1 labeled.

similarity is that both imidazole ring nitrogens of His^{64} carry protons, and the imidazole ring remains protonated over the pH range of 5.5 to 9.5. Other similarities include the fact that exchange of His NH protons is slowed sufficiently to allow direct observation of the N-H coupling and that both N-H protons produce observable resonances in the low-field region of the ^1H NMR spectrum. This is remarkable in itself because resting subtilisin does not exhibit an observable low-field proton resonance.

All other boronic acid inhibitors examined so far also form serine adducts with subtilisin. These include Ac-boroPhe, benzene boronic acid, and even MeOSucAla-Ala-Pro-D-boroPhe and boronic acid. The fact that the latter two inhibitors form serine adducts is especially surprising on the basis of our results with trypsin-forming enzymes.

The amide sidechain of Asn^{155} is believed to contribute directly to transition-state-stabilization as part of the oxyanion hole. Replacing Asn^{155} with an alanine residue reduces the k_{cat} of subtilisin 150-fold (Bryan *et al.*, 1986; Braxton and Wells, 1991). We have incorporated ^{15}N-labeled histidine into an Asn155Ala mutant of subtilisin BPN' and examined the behavior of the His^{64} ^{15}N signals in several boronic acid complexes (R. Day and W. Bachovchin, manuscript in preparation). Our objective was to determine if weakening the transition-state-stabilizing ability of the enzyme would cause some boronic acid inhibitors to form histidine

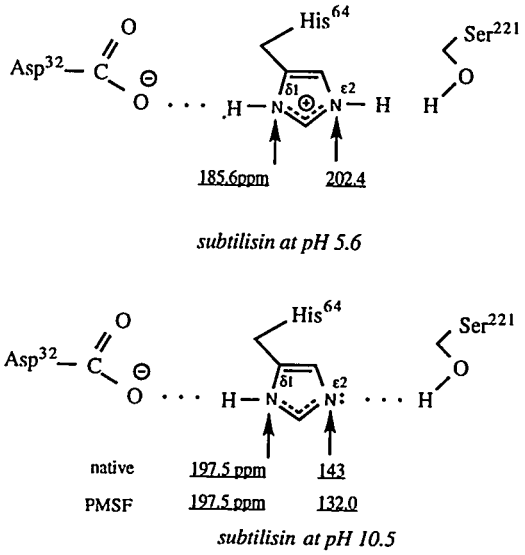


Fig. 8. ^{15}N shifts of His^{64} of resting subtilisin at the high- and low-pH ends of its titration and on PMSF inactivation.

adduct complexes. Surprisingly, however, thus far we have only observed serine adducts. This suggests that the subtilisins behave entirely differently than the trypsin family with respect to binding of boronic acids. However, another more likely possibility is that subtilisin BPN' is a remarkably nonspecific protease and that a subtilisin with more restrictive specificity, such as furin, would exhibit histidine adducts.

References

Alei, M.J., Morgan, L.O., Wageman, W.E., and Whaley, T.W. (1980) *J. Am. Chem. Soc. 102* , 2881-2887.

Allaire, M., Chernaia, M.M., Malcolm, B.A., and James, M.N. (1994) *Nature 369,* 72-76.

Bachovchin, W. W. (1985) *Proc. Natl. Acad. Sci. US* A*82,* 7948-7951.

Bachovchin, W.W. (1986) *Biochemistry 25,* 7751-7759.

Bachovchin, W.W. and Roberts, J.D. (1978) *J. Am. Chem. Soc. 100,* 8041-8047.

Bachovchin, W.W., Plaut, A.G., Flentke, G.R., Lynch, M., and Kettner, C.A. (1990) *J. Biol. Chem. 265,* 3738-37439.

Bachovchin, W.W., Wong, W.Y., Farr, J.S., Shenvi, A.B., and Kettner, C.A. (1988) *Biochemistry 27,* 7689-7697.

Bauer, C.-A., Brayer, G.D., Sielecki, A.R., and James, M.N.G. (1981) *Eur. J. Biochem. 120,* 289-294.

Blomberg, F., Maurer, W., and Ruterjans, H. (1977) *J. Am. Chem. Soc. 99,* 8149-8159.

Blow, D.M., Birktoft, J.J., and Hartley, B.S. (1969) *Nature 221,* 337-340.

Bone, R., Frank, D., Kettner, C.A., and Agard, D.A. (1989) *Biochemistry 28,* 7600-7609.

Braxton, S. and Wells, J.A. (1991) *J. Biol. Chem. 266,* 11797-11800.

Farr, J.S., Smith, S.O., Kettner, C.A., Griffin, R.G., and Bachovchin, W.W. (1989) *Proc. Natl. Acad. Sci.USA 86,* 6922-6924.

Bryan, P., Pantoliano, M.W., Quill, S.G., Hsiao, H.Y., and Poulos, T. (1986) *Proc. Natl. Acad. Sci.USA 83,* 3743-5.

Farr-Jones, S., Wong, W.Y.L., Gutheil, W.G., and Bachovchin, W.W. (1993) *J. Am. Chem. Soc . 115,* 6813-6819.

Flentke, G.R., Munoz, E., Huber, B.T., Plaut, A.G., Kettner, C.A., and Bachovchin, W.W. (1991) *Proc. Natl. Acad. Sci.USA 88,* 1556-1559.

Kettner, C.A., Bone, R., Agard, D.A., and Bachovchin, W.W. (1988) *Biochemistry 27,* 7682-8.

Kubota, T., Flentke, G.R., Bachovchin, W.W., and Stollar, B.D. (1992) *Clinical Exper. Immunol. 89,* 192-197.

Roberts, J.D., Chun, Y., Flanagan, C., and Birdseye, T.R. (1982) *J. Am. Chem. Soc. 104,* 3945-3949.

Shuster, I.I. and Roberts, J.D. (1979) *J. Org. Chem. 44,* 3864-3867.

Takahashi, L.H., Radhakrishnan, R., Rosenfield, R.J., and Meyer, E.J. (1989) *Biochemistry 28,* 7610-7617.

Tanaka, T., Kameoka, J., Yaron, A., Schlossman, S.F., and Morimoto5, C. (1993) *Proc. Natl. Acad. Sci.USA 90,* 4586-4590.

Tsilikounas, E., Kettner, C.A., and Bachovchin, W.W. (1992) *Biochemistry 31,* 12839-12846.

Complex DNA Structures and Structures of DNA Complexes

Walter J. Chazin, Göran Carlström, Shiow-Meei Chen, Siobhan Miick, Luigi Gomez-Paloma, Jarrod Smith, and Jan Rydzewski

Department of Molecular Biology
The Scripps Research Institute
10666 North Torrey Pines Road
La Jolla, CA 92037

Complex DNA structures (for example, triplexes, quadruplexes, junctions) and DNA-ligand complexes are more difficult to study by NMR than standard DNA duplexes are because they have higher molecular weights, show nonstandard or distorted local conformations, and exhibit large resonance linewidths and severe ^{1}H spectral overlap. These systems also tend to have limited solubility and may require specialized solution conditions to maintain favorable spectral characteristics, which adds to the spectroscopic difficulties. Furthermore, with more atoms in the system, both assignment and structure calculation become more challenging. In this article, we focus on demonstrating the current status of NMR studies of such systems and the limitations to further progress; we also indicate in what ways isotopic enrichment can be useful.

In contrast to the situation for proteins and RNA, application of new and powerful stable isotope-based strategies for determining the three-dimensional structures of DNA systems using NMR has not been possible; uniform isotopic enrichment by either chemical or biochemical methods remains a major obstacle. Current applications of stable isotopes to DNA systems are therefore restricted to chemical or semibiosynthetic methods that produce selective ^{15}N- or ^{13}C-labeled precursors for chemical DNA synthesis or isotopic enrichment of a DNA ligand. However, the future looks

promising because several different strategies for uniform isotope enrichment of DNA are under development. The current state of progress in NMR analysis and the application of stable isotopes to study complex DNA structures and DNA-ligand complexes will be discussed using the specific examples of DNA crossover structures and drug-DNA complexes.

Methods

The preparation and purification of oligonucleotides and SN-6999 has been described in detail (Leupin *et al.*, 1986, 1990; Chen *et al.*, 1991, 1992, 1993). The enediynes and their complexes with DNA have all been prepared in the laboratory of Professor K.C. Nicolaou. Typically, complementary oligonucleotides were mixed in equimolar amounts and then lyophilized and redissolved in 420 μℓ stock buffer (50 mM sodium phosphate buffer at pH 7.0, 100 mM NaCl, 0.1% NaN_3, 5% D_2O). For experiments to examine only nonlabile protons, the solutions were repeatedly lyophilized from 99.6% D_2O, followed by dissolution in 99.996% D_2O (MSD Isotopes, Montréal, Canada). Titrations were monitored by following the imino, aromatic, and proton resonances in the ^1H NMR spectrum. For experiments that examined labile protons, solutions were lyophilized and dissolved in 95% H_2O/5% D_2O.

NMR experiments were performed using Bruker AM and AMX spectrometers operating at 500 and 600 MHz for ^1H. The 2Q spectra were acquired using the standard pulse sequence (Braunschweiler *et al.*, 1983). The transmitter was placed to the low-field side of the H1′ region and base proton correlations were folded in the ω_1 dimension. Two-dimensional nuclear Overhauser effect spectroscopy (NOESY) spectra were acquired using the standard pulse sequence (Macura and Ernst, 1980). For spectra in H_2O, the last pulse was replaced by a composite jump-return sequence (Plateau and Guéron, 1982). For spectra in D_2O, a short Hahn-echo sequence was inserted just before acquisition to improve the quality of the baseline (Rance and Byrd, 1983; Davis, 1989). The ^{13}C-directed (Rance *et al.*, 1987) and ^{13}C-ω_1-half-filtered (Otting *et al.*, 1986) NOESY spectra for the SN-6999 complexes were acquired at 301 K with 64 or 96 scans/t_1 value, a 200-ms mixing time and t_{1max} = 18 ms.

All data were processed using SUN workstations or a CONVEX C240 computer with FTNMR and FELIX software (Hare Research, Inc., Woodinville, Washington). Typical acquisition and processing parameters can be found in Chen *et al.* (1991, 1992, 1993).

Holliday junction (HJ) models were created from regular B-form DNA by an in-house graphics program (T. Macke, D.A. Case, unpublished) that generates atomic coordinates of nontraditional nucleic acid structures, including HJs. Each of these structures were energy-minimized *in-vacuo* in the AMBER force-field.

Results and Discussion

Holliday Junctions

The Holliday junction (HJ) is a four-arm, branched DNA structure that is a common intermediate to most currently accepted models for genetic recombination (Holliday, 1964). A significant body of evidence has accumulated indicating that the structure at the junction has a central role in determining the outcome of the recombination event (Lilley and Clegg, 1993). HJs or HJ-like structures also appear to play significant roles in other cellular processes such as replication and telomere resolution. Thus, an understanding of the molecular basis of these cellular processes will require a detailed knowledge of HJ structure and the nature of interactions with HJ-resolving enzymes. The principal goals of our research are to determine the three-dimensional structures of Holliday junctions, to understand how the sequence at the junction influences structure, and to determine why the cellular recombination machinery is influenced by these differences.

Figure 1 shows the basic scheme of the 32 base-pair Holliday junction models we study, as well as the numbering systems used to designate each nucleotide and the four arms. J2 is a permutation of J1 with one base pair exchanged between arms III and IV, as is indicated by the arrows in Fig. 1. J2P1 is a one-step cyclic permutation of J2. The four base pairs at the center of these four-arm DNA structures compose the "junction," and the base pair at the open end of each duplex is termed the arm terminus.

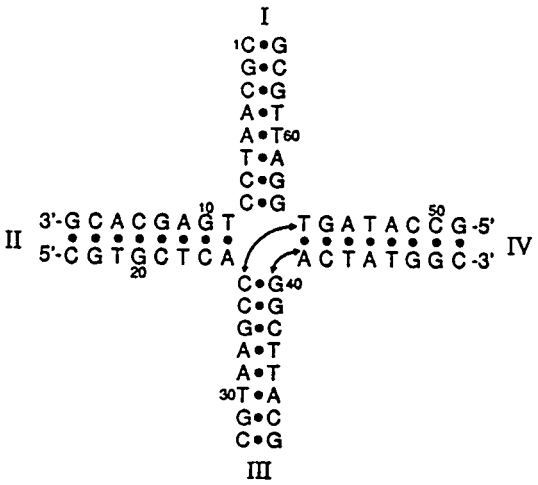


Fig. 1. The sequence and numbering systems of the 32 base-pair models of the Holliday junction. The sequence of J1 is shown; the changes required to produce J2 are indicated by the arrows, and J2P1 corresponds to a single-step cyclic permutation of the J2 sequence. The oligonucleotide sequence runs consecutively from the 5' to the 3' end of strand 1 (residues 1–16), then on in the same manner to strands 2 (17–32), 3 (33–48), and 4 (49–64). The four arms of the junction are labeled with roman numerals.

Resonance Assignments

The standard methods for ^{1}H NMR analysis of duplex DNA have been extensively reviewed (*e.g.*, Wüthrich, 1986); the 32 base-pair HJs we study are clearly at the limit of current homonuclear NMR methodology. Their large

size causes two specific problems; that is, large linewidths of the NMR signals and severe overlap of resonances resulting from the large number of ^{1}H signals. The linewidth problem is an inherent property of the system—the result of the relatively short rotational correlation time for this relatively large (from a ^{1}H NMR perspective) molecule. Our success with J1, J2, and J2P1 has shown that although the linewidths must be carefully considered, this problem does not preclude analysis by available methods. Our approach for obtaining the sequential resonance assignments is to identify as many of the observed NOE connectivities as possible and, correspondingly, to obtain assignments by as many sequential pathways as possible (Chen *et al.*, 1991). This step was necessary to overcome the complexity of the assignment problem and serves to increase the reliability of the final result. A second important aspect of our strategy is to simultaneously analyze all connectivities in all of the spectra rather than to analyze a specific region of a single spectrum.

The extremely large number of cross peaks in the characteristic regions of the spectrum precludes straightforward analysis of any one region for obtaining assignments. Invariably, there is overlap, and it becomes difficult to resolve all of the expected intra- and interresidue cross peaks. A typical example is presented in Fig. 2, which shows the base 6H/8H to the ribose H2′/H2″ region of a 600-MHz NOESY spectrum of J2P1. The strategy that we have found successful for analysis relies on the fact that, except across the junction itself, the HJs are composed of regular B-form helices. A search is made for all connectivities that correspond to inter-proton distances <6 Å for all pairs of oligonucleotides in B-form DNA. The connectivities are analyzed in parallel to establish a fully self-consistent set of resonance assignments. In addition to the standard base (6H/8H) to Me, H1′, H2′, and H2″ connectivities, a number of other connectivities have proven useful (Chen *et al.*, 1991). Using this strategy, complete sequence-specific assignments have been obtained for the labile and nonlabile aromatic protons, and the C1′, C2′, and C3′ sugar protons of J1, J2, and J2P1 (Chen *et al.*, 1993; Chen and Chazin, 1994; Carlström and Chazin, in preparation).

To further confirm our results for the key residues at the junction, we recently began using ^{15}N-labeled thymidine phosphoramidites synthesized for us by the National Stable Isotope Resource (at Los Alamos National Laboratory) to prepare HJs with single residues at the junction labeled with ^{15}N. This labeling provides a vast degree of spectral simplification in the ^{1}H spectra, as well as ^{15}N signals that can be probed for structural and dynamic parameters. Figure 3 shows the ^{15}N-^{1}H correlation signal from a two-dimensional HMQC spectrum of selectively ^{15}N labeled J2P1. In the long term, when methods become available for uniform ^{13}C/^{15}N labeling of DNA, multidimensional multinuclear NMR strategies will be applied to the HJ system.

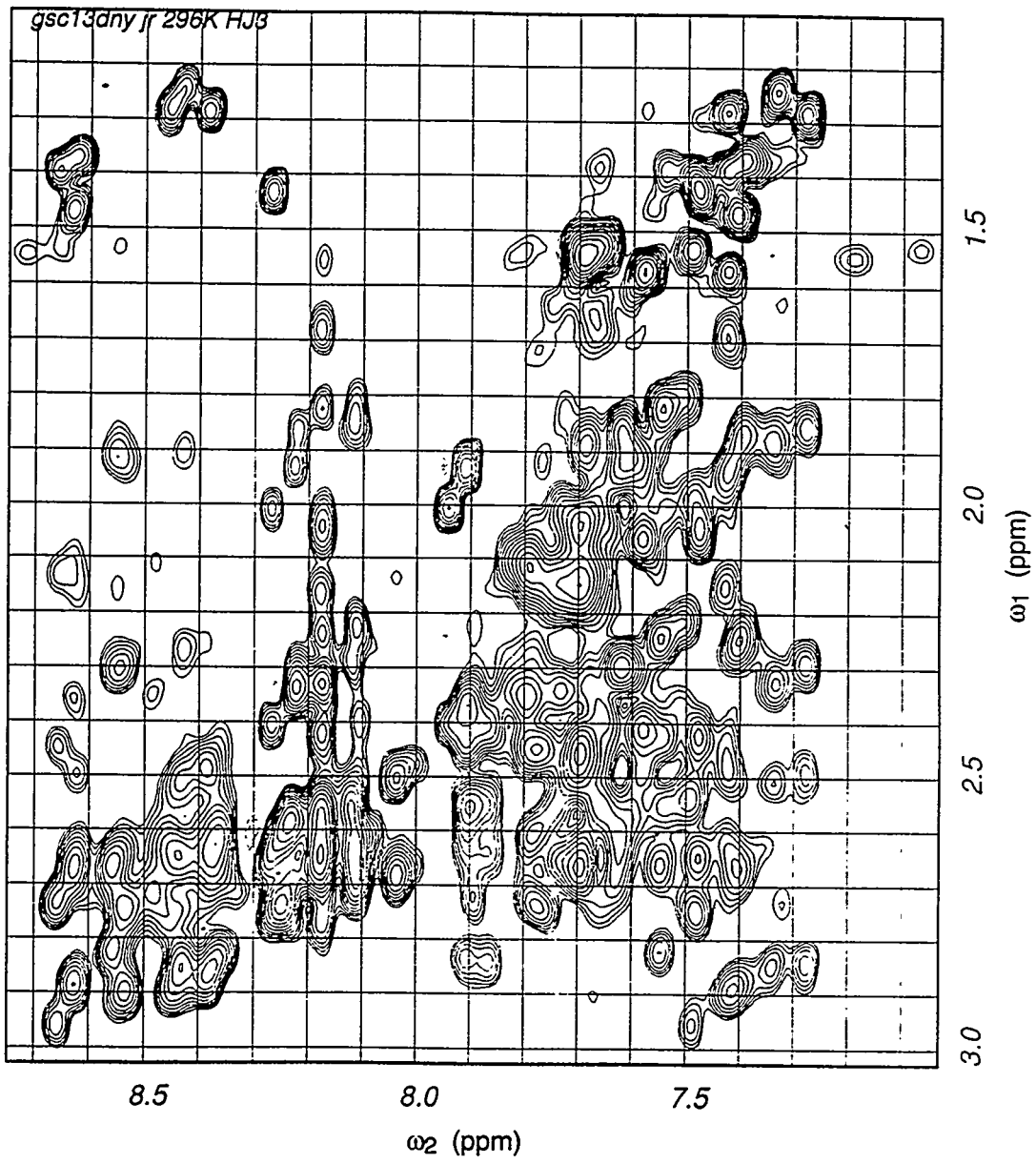


Fig. 2. Crosspeak overlap in NOESY spectra of 32 base-pair model HJs. Expanded region of a 600-MHz homonuclear ^1H JR-NOESY spectrum showing the highly overlapped base (6H/8H) to 5-CH_3, H2', and H2" connectivities. The spectrum was acquired at 296 K on a Bruker AMX 600 from a 1.0 mM solution of J2P1.

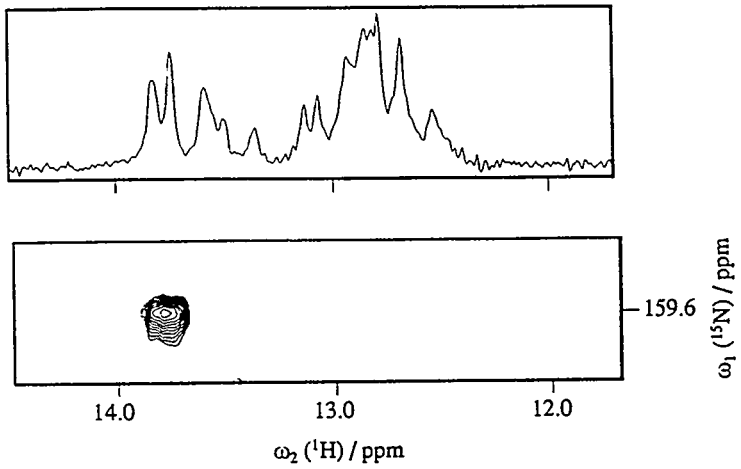


Fig. 3. Spectral simplification achievable with selective isotopic enrichment. One-dimensional ^1H (upper panel) and two-dimensional ^{15}N-^1H zero quantum correlation (lower panel) spectra of 0.8 mM J2P1 labeled selectively with ^{15}N at the N3 (and N1) position of T^9, directly at the junction. This spectrum was acquired at 300 K on a Bruker AMX500 spectrometer.

Information on the Structures of J1 and J2

Once sequence-specific resonance assignments were made, the interpretation of the NMR data in terms of its implications about junction structure could be initiated. The analysis of the labile protons of J1 and J2 provided sufficient information to define the base pairing scheme and to infer the duplex stacking arrangements of these two HJs (Chen *et al.*, 1993). It is of importance to note that not all of the expected NOEs are observed between the stacked arms across the junction. This occurs for two reasons. First, it is logical to assume some degree of structural perturbation at the junction, and in fact, sequential and intra-residue NOEs between protons of the base and sugar rings indicate some deviations from idealized B-form geometry. Second, the resonance linewidths of the residues at the junction tend to be broad, which makes detection of the associated cross peaks more difficult.

A number of NOEs between nonlabile protons have been identified that are indicative of the nature of the geometry at the junction; J1 has arm I stacked over arm II (Chen and Chazin, 1994). An explanation of how the observed NOEs distinguish one arm stacking arrangement from the other is given in Macke *et al.*, (1992). The determination of stacking geometry was more difficult for J2 than for J1 because the detection of NOEs for the residues at and near the junction was more difficult. The primary stacking geometry implicated in the data for J2 is arm I with arm II and arm III with arm IV, as determined for J1.

However, not all of the observed NOEs for J2 can be explained solely by this arrangement; a set of unambiguous but uniformly weaker NOEs were identified that are clearly inconsistent with the I/II stacking geometry. For example, NOEs from both C^8 and A^{40} to T^9 were

observed. In models built with standard B-form geometry, these residues can never all be close to each other. Even with substantial distortion from standard B-form geometry in the I over II stacking isomer, it is not possible to bring T^9 and A^{40} close enough to explain the observed NOEs. However, in the J2 structure with the alternate I over IV stacking arrangement, T^9 and A^{40} are stacked directly adjacent to each other, well within the range to give rise to NOEs between them. We have concluded that under the conditions of the NMR experiments, J2 exists in solution as an equilibrium distribution of the two possible stacking geometries in fast exchange on the NMR time-scale, with a preference for the (I/II,III/IV) *vs* the (I/IV,II/III) isomers of about 5:1. Further support for this interpretation came from a temperature-dependence study using the J2P1 sample labeled selectively at the N3 position of residue T^9. As the temperature is lowered below 300 K, a second resonance becomes visible in the ^{15}N filtered ^1H spectrum. As shown in Fig. 4, at 285 K a minor signal is readily apparent in approximately the same 5:1 ratio as estimated from the analysis of the NOESY spectra.

The inability to crystallize HJs for x-ray diffraction studies has prompted this laboratory to use NMR spectroscopy

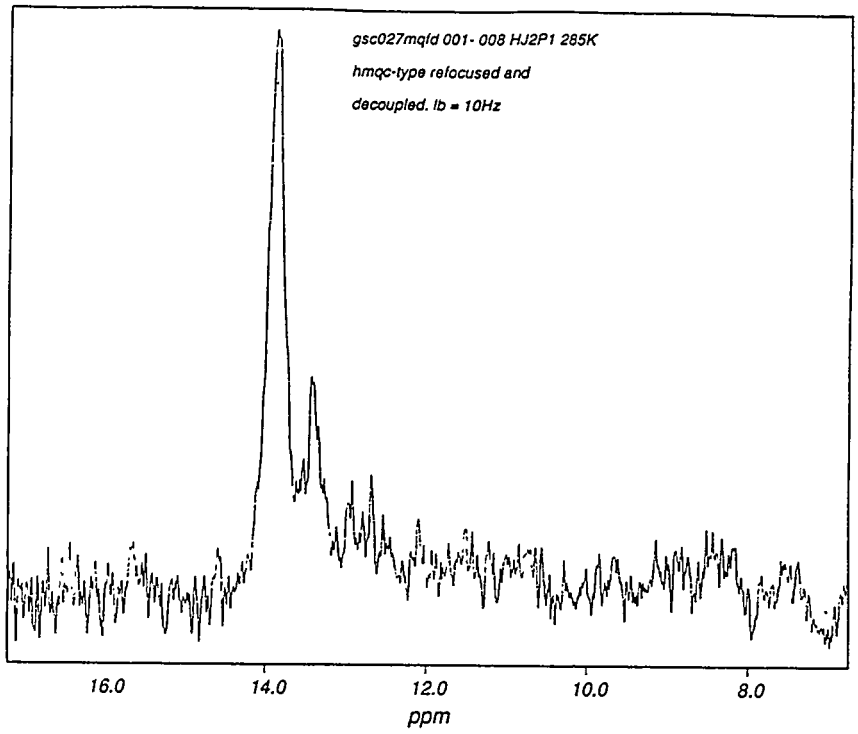


Fig. 4. The model HJ J2P1 exists as an equilibrium mixture of two conformers. This 1D ^{15}N-^1H HMQC spectrum of ^{15}N-labeled J2P1 at 285 K was acquired from the same sample and on the same instrument as the spectrum shown in Fig. 3.

to begin characterization of the three-dimensional structure and dynamics of 32 base-pair HJs in solution. Sequence-specific ^{1}H NMR assignments have been made for three model HJs. However, the calculation of the complete three-dimensional structure of these model HJs is greatly hindered by the combination of severe resonance overlap and large linewidths for these ~20-kDa molecules. Uniform isotopic enrichment of DNA will provide a vast improvement in the resolution of cross peaks and also will allow the determination of three-dimensional solution structures at high resolution.

Drug-DNA Complexes

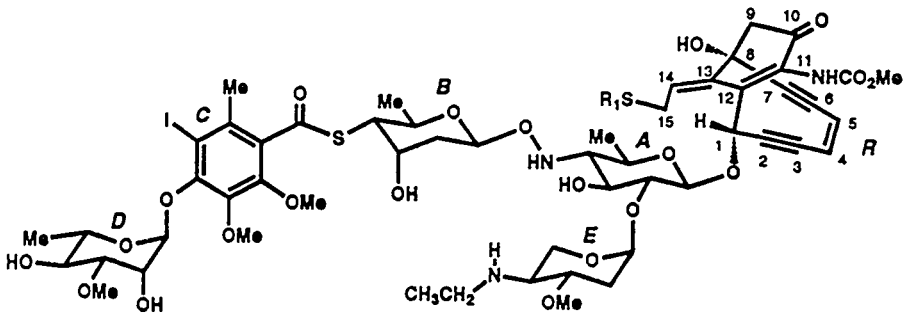
DNA binding agents play a significant role in current clinical strategies for cancer therapy. The activity of a large majority of antitumor drugs in present clinical use or under preclinical study is thought to occur by interference with DNA metabolism (Zimmer and Wähnert, 1986). The chemical structures of these drugs vary considerably, matched by a great diversity in modes of interaction ranging from groove binding and intercalation to direct chemical reactivity. The primary goal of structural research in this field is to understand the interactions of anti-tumor drugs with DNA at the molecular level. The objective of our research is to identify the molecular basis for the interaction of specific classes of compounds with the minor groove of duplex DNA using NMR and computational methods as the primary tools.

Two examples of drug-DNA complexes will be discussed here: an enediyne DNA cleaving agent (calicheamicin) and a noncovalent binding bisquaternary ammonium heterocycle (SN-6999). The significance of this type of research is that if a better understanding of the relationship between molecular structure and antitumor activity can be obtained, this will provide a rational basis for design of new, more effective therapeutic agents.

Calicheamicin

Enediynes are a new class of extremely potent antitumor agents (Nicolaou and Dai, 1991). These molecules exert their biological activity by cleaving duplex DNA through formation of a highly reactive diradical intermediate that abstracts hydrogen atoms from the DNA backbone. The site-specific recognition of the DNA minor groove is governed by the complex carbohydrate moiety of the molecule. NMR studies of the complex formed between d(GCATCCTAGC)• d(CGTAGGATCG) and the calicheamicin γ_1^I analog calicheamicin θ (Fig. 5) have revealed an important aspect of the DNA cleaving properties of this family of agents (Gomez-Paloma *et al.*, 1994).

After titration of the drug to a 1:1 ratio, spectra were acquired at several temperatures to make the assignments of the drug and DNA resonances in the complex. The perturbations of DNA chemical shifts induced by drug binding clearly indicate binding in the previously identified high affinity d(TCCT)•d(AGGA) binding site.



Calicheamicin θ $R_1=CH_3CO$
Calicheamicin γ_1^I $R_1=CH_3SS$

Fig. 5. Molecular structures of calicheamicin γ_1^I and calicheamicin θ.

An in-depth analysis of the NOESY spectra has revealed intermolecular contacts for docking the drug into the minor groove of the duplex. A structural model for the complex was generated by refining manually docked initial structures using molecular dynamics calculations with NMR-derived distance restraints. Refined structures, starting from standard A-form and B-form DNA geometries, converged to an RMSD of 1.12 Å for the well-defined regions of the complex. This structural model is stabilized by several specific interactions including H-bonds and favorable Van der Waals interactions. Temperature-dependence studies of the ^1H NMR spectrum revealed that calicheamicin γ_1^I binds in two different modes within the d-(TCCT)•d-(AGGA) recognition site; two unequally populated species in slow exchange on the NMR time scale are observed at low temperature. The three-dimensional structure of the complex has been determined for the major conformer.

To explain the DNA cleavage properties of calicheamicin, it is necessary to invoke some dynamic process to alternately reposition the aglycone for abstraction of the hydrogen atoms. Molecular modeling reveals that it is highly unlikely that both hydrogen atom abstractions required for DNA cleavage can occur with the activated diradical intermediate in a fixed position midway between the two sites. Based on the observation of two low-energy conformers in the NMR spectra, we propose that a dynamic process occurs involving a slight movement of the aglycone in the minor groove. This movement is sufficient to allow the carbon radicals of the benzenoid ring to reach the required positions for the abstraction of the two H atoms. The two conformers observed by NMR thus correspond to alternate binding modes for the aglycone, one closer to the C^5 abstraction site on the TCCT strand and the other closer to the G^{19} abstraction site on the AGGA strand.

SN-6999

The bis-quaternary ammonium heterocycles (Braithwaite and Bagueley, 1980) are well-studied agents that have served as a model system for determining the molecular basis for recognizing the DNA minor groove. The binding of a specific agent from this family, SN-6999 (Fig. 6), to different DNA duplexes has been studied in detail by NMR spectroscopy. Upon titration of a DNA solution with one molar equivalent of drug, the two well-separated drug resonances near 8.6 ppm in the ^{1}H NMR spectrum show changes in chemical shift and increased linewidth that are characteristic of the binding of the drug to the DNA duplex. Another readily discernable indication of drug binding is the substantial change in the chemical shift of the imino proton resonances. Given that there is only one AT-rich binding site, the absence of multiple sets of NMR signals after binding the asymmetric drug implies either that the drug binds in only one orientation on the DNA or that a single, time-averaged NMR spectrum is observed due to fast exchange (on the NMR timescale) between two or more orientations of the drug.

One method of distinguishing between these two possibilities is to look for evidence of selective broadening of resonances in the NMR spectrum as a function of temperature. Some of the thymine methyl resonances in the high-field end of the spectrum are broader than other resolved signals in the one-dimensional spectrum, and a separation of each broadened methyl resonance into two discrete signals has been observed for the SN-6999:d-(GCATTAATGC)$_2$ complex (Leupin *et al.*, 1986). Thus, there is intermediate-to-fast exchange of the drug (on the NMR timescale) between at least two DNA binding sites.

Obtaining resonance assignments for the complex is much more difficult than for the free DNA because the ^{1}H-resonance linewidths of the complex are significantly larger. COSY-type experiments for identifying scalar correlations are not very informative, but 2Q and TOCSY experiments are more successful because the latter are less sensitive to cancellation problems associated with larger linewidths. As was the case in the HJ study, 2Q spectroscopy was the method of choice for identifying scalar correlations because the crosspeaks in TOCSY spectra tend to merge in spectral regions that are very crowded, whereas the 2Q peaks are

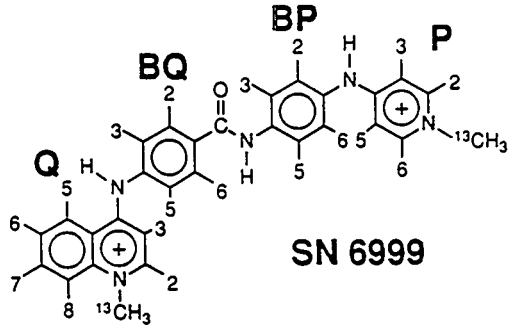


Fig. 6. Molecular structure of 4-[p-[p-(4-quinolylamino)benzamido]-anilino]pyridine (SN-6999). Each proton of the drug is identified by a letter specifying the aromatic ring (Q- quinolinium, BQ- benzamido, BP- anilino, P- pyridinium) and a number specifying its position in the ring (for example, P-2H).

easier to resolve because of their antiphase multiplet structure in the ω_2 dimension. With these scalar correlations, the sequence-specific assignment of resonances in the complex can be made from NOESY spectra in the usual manner. It is important to note that some of the characteristic intraresidue and sequential NOEs are observed only barely above the thermal noise as a result of substantial exchange broadening of resonances from residues that are drastically affected by binding of the drug.

With the availability of sequence-specific resonance assignments for both components of the complex, it becomes possible to identify the regions of intermolecular contact by examining the drug-binding-induced changes in ^1H chemical shifts and by noting the NOEs between drug and DNA protons. Profiles of the perturbations of chemical shifts indicate that the effects of drug binding are predominantly felt in the AT-rich segment of the duplex, as expected, but that these perturbations can extend one base (but not one base-pair) along each strand. The NOE data are fully consistent with the analysis of drug-induced chemical-shift changes, indicating that the drug is bound primarily in the AT-rich region of the duplex. Observation of intermolecular contacts to residues adjacent to the AT-rich region demonstrates that the stabilization of drug binding for the BQAH family of AT-specific, minor groove binding drugs is not exclusively based on interactions with AT basepairs (Chen *et al.*, 1992).

An inspection of the NOEs indicates that the asymmetric SN-6999 binds in the minor groove in two orientations relative to the helix axis and with no preference for one orientation over the other for most duplexes, as indicated by NOEs from specific drug resonances and DNA residues at opposite ends of the binding pocket. The intermolecular contacts identified by NOEs can be used to align the drug along the duplex in the two orientations. These results, in combination with the observations of only one set of NMR resonances and specific resonances that are exchange-broadened, leads to the conclusion that on the NMR timescale, the asymmetric drug exchanges rapidly between its two possible orientations in the binding pocket of the DNA duplex (Leupin *et al.*, 1986; Chen *et al.*, 1992).

Strong temperature-dependence of the linewidths of the C2 protons is also observed and two sets of signals are found in NOESY spectra at low temperatures, confirming the conclusion that the drug binds in both orientations in the DNA minor groove. The one- and two-dimensional NOESY spectra acquired at various temperatures have been carefully examined to determine coalescence temperatures and to estimate the exchange rate between the two binding sites, as well as the activation energy for the exchange processes. The similarity between the activation energies indicates that these exchange processes are due to the same physical phenomenon; that is, flipping of the SN-6999 molecule relative

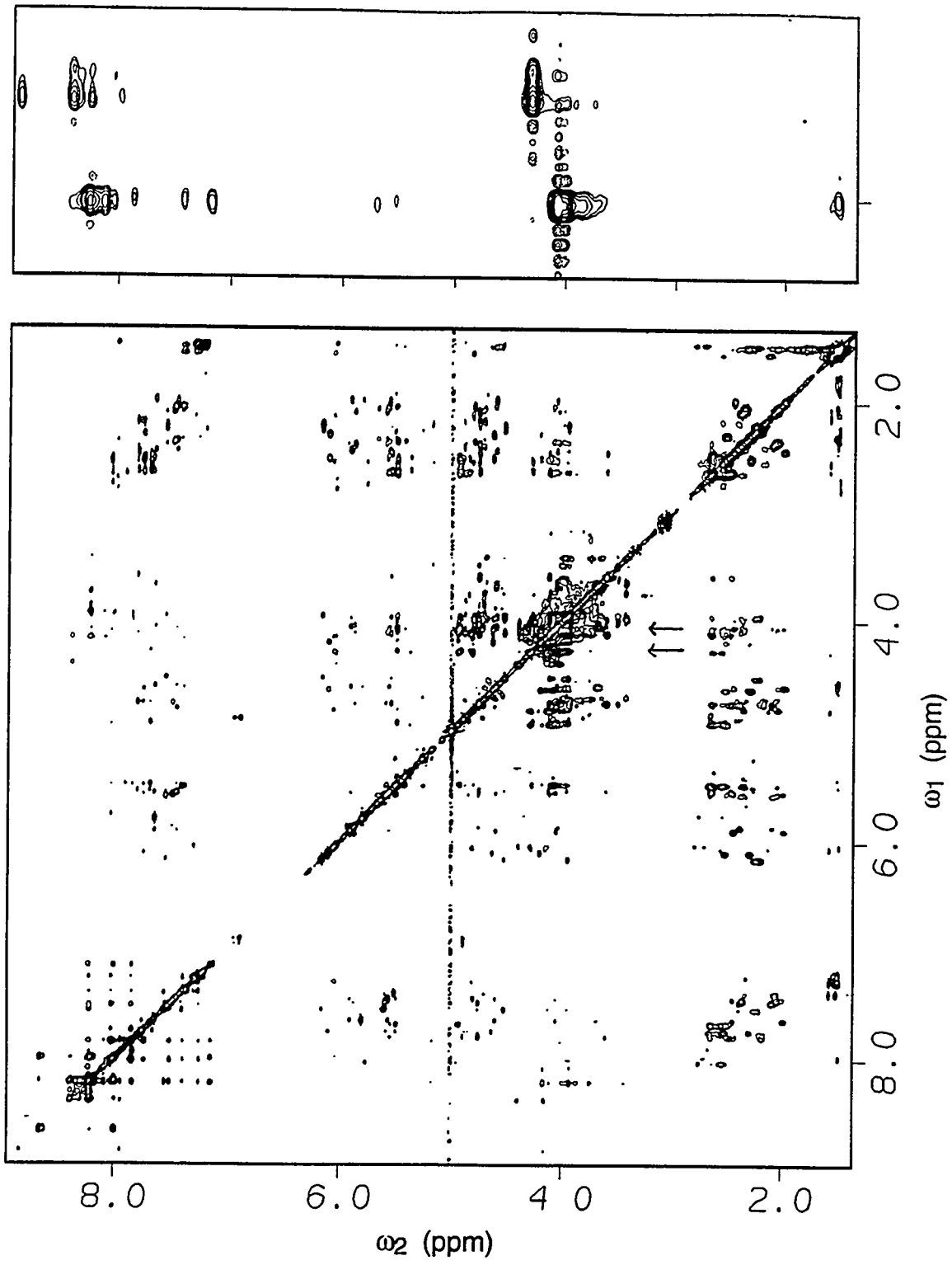
to the DNA duplex (Leupin *et al.*, 1986). Such a flip-flop mechanism has been postulated for the exchange of other minor-groove-binding drugs between two symmetrically related or similar DNA binding sites. Thus, the flip-flop process seems to represent a property inherent to complexes between minor-groove-binding ligands and duplex DNA.

In recent studies, the DNA duplex d(CCGAAAAAGCC)•d(GGCTTTTTCGG) was designed specifically to reduce the rate of the flip-flop exchange of the drug. The logic behind this design was that the narrowed minor groove of the dA_5 tract would provide an optimized binding site with greater complementarity between the Van der Waals surfaces of the drug and the minor groove of the DNA duplex. After titration of the drug, NMR spectra revealed that the drug binds with a strong preference for a single orientation in the dA-tract and with a longer lifetime than previously observed for the other SN-6999 complexes. The perturbations of chemical shifts of the DNA induced by drug binding clearly indicate the expected binding in the AT-rich central region of the duplex.

An in-depth analysis of the proton-proton contacts in the NOESY spectra revealed a significant number of the critical intermolecular contacts that allow a determination of the three-dimensional structure of the complex. Experiments using SN-6999 prepared with ^{13}C-labeled N-methyl groups provided a number of additional constraints that were not resolvable in the standard NOESY spectra. Figure 7 shows how the crowded homonuclear ^{1}H NOESY spectrum (lower panel) is vastly simplified by filtering out NOEs to the two ^{13}C-labeled N-methyl groups of the drug.

Although low-to-medium-resolution structures can now be determined with available homonuclear data, progress towards higher resolution structures is inhibited by an inability to identify and integrate numerous overlapped crosspeaks. Uniform ^{15}N, ^{13}C enrichment offers promise for considerable improvement in the number and quality of NMR-derived input constraints for calculating three-dimensional structures of drug-DNA complexes at high resolution.

Fig. 7. Spectral simplification of NOESY spectra from selective isotopic enrichment. NOESY spectra of the 1:1 complex of d(CCGAAAAAGCC)•d(GGCTTTTTCGG) and ^{13}C-labeled SN-6999. The experiments were acquired at 274 K from a 1 mM D_2O solution on a Bruker AMX2-500. The lower panel shows the full homonuclear ^{1}H HE-NOESY spectrum. The upper panel shows the ^{13}C-directed NOESY spectrum containing the NOEs to the two ^{13}C labeled N-methyl groups of the drug, which correspond to the two rows indicated by arrows in the lower panel.


8.0
6.0
4.0
2.0
ω2 (ppm)
2.0
4.0
6.0
8.0
ω1 (ppm)

Acknowledgments

Supported by operating grants from the American Cancer Society (CH-529) and the National Science Foundation (DMB 9019250) as well as fellowships to G. Carlström from the Swedish National Science Foundation and to W.J. Chazin from the American Cancer Society (JFRA-294, FRA-436).

We thank Beth Larson for her help with typing the manuscript.

References

Braithwaite, A.W. and Bagueley, B.C. (1980) *Biochemistry 19*, 1101-1106.

Braunschweiler, L., Bodenhausen, G., and Ernst, R. R. (1983) *Mol. Phys. 48*, 535-560.

Chen, S.-M. and Chazin, W.S. (1974) *Biochemistry 33*, 11453-11459.

Chen, S.-M., Heffron, F., Leupin, W., and Chazin, W.J. (1991) *Biochemistry 30*, 766-771.

Chen, S.-M., Leupin, W., Rance, M., and Chazin, W.J. (1992) *Biochemistry 31*, 4406-4413.

Chen, S.-M., Heffron, F., and Chazin, W.J. (1993) *Biochemistry 32*, 319-326.

Davis, D.G. (1989) *J. Magn. Reson. 81*, 603-607.

Gomez-Paloma, L., Smith, J., Chazin, W.J., and Nicolaou, K.C. (1994) *J. Am. Chem. Soc. 116*, 3697-3708.

Holliday, R. (1964) *Genet. Res. 5*, 282-304.

Leupin, W., Chazin, W.J., Hyberts, S., Denny, W.A., and Wüthrich, K. (1986) *Biochemistry 25*, 5902-5910.

Leupin, W., Otting, G., Amacker, H., and Wüthrich, K. (1990) *FEBS Lett. 263*, 313-316.

Lilley, D.M.J. and Clegg, R.M. (1993) *Quarterly Rev. Biophys. 26*, 131-175.

Macke, T., Chen, S.-M., and Chazin, W.J. (1992) *in Structure and Function, Vol. 1*, R.H. Sarma and M.H. Sarma, eds., Adenine Press, pp. 213-227.

Macura, S. and Ernst, R.R. (1980) *Mol. Phys. 41*, 95-117.

Nicolaou, K.C. and Dai, W.-M. (1991) *Angewandte Chemie 30,* 1387-1530.

Otting, G., Senn, H., Wagner, G., and Wüthrich, K. (1986) *J. Magn. Reson. 70,* 500-505.

Plateau, P. and Guéron, M.J. (1982) *J. Am. Chem. Soc. 104,* 7310-7311.

Rance, M. and Byrd, R.A. (1983) *J. Magn. Reson. 54,* 221-240.

Rance, M., Wright, P.E., Messerle, B.A., and Field, L.D. (1987) *J. Am. Chem. Soc. 109,* 1591-1593.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids,* Wiley, New York.

Zimmer, C. and Wähnert, U. (1986) *Prog. Biophys. Molec. Biol. 39,* 31-112.

High-Resolution Polypeptide Structure and Dynamics in Anisotropic Environments: the Gramicidin Channel

Timothy A. Cross, Kwun-Chi Lee, Randal R. Ketchem, Weidong Hu, Noel D. Lazo, and Shougin Huo

National High Magnetic Field Laboratory,
Institute of Molecular Biophysics, and Department of Chemistry
Florida State University
1800 E. Paul Dirac Drive
Tallahassee, FL 32306

To understand the details of macromolecular function, high-resolution structural and dynamic detail is essential. The polypeptide fold of the gramicidin channel has been effectively modeled for the past 20 years (Urry, 1971), yet the functional changes in conductance and channel lifetime associated with amino acid substitutions cannot be predicted. To accomplish this goal, high-resolution electrostatic modeling and the precise orientation of all dipoles are required.

Furthermore, an enhanced knowledge of the complex molecular environment of this membrane-bound peptide is needed. An aqueous environment is relatively uniform and achiral. The membrane environment is very heterogenous and chiral. A knowledge of the interactions, specific and nonspecific, between peptide and lipid will aid in developing a better understanding of this environment. To accomplish this goal, it is necessary to study the peptide in an extended lipid bilayer, rather than in a vesicular or micellar form. These latter environments are likely to possess increased dynamics, increased water penetration, and distorted interactions between the polypeptide and membrane surface. To perform NMR studies on bilayer bound peptides, solid state NMR methods are required, and for specific site information, isotopic labels are incorporated using solid phase peptide synthesis.

Orientational constraints derived from solid state NMR are sufficient for determining macromolecular three-dimensional

structure (Ketchem *et al.*, 1993). In anisotropic samples, the orientation-dependence of nuclear spin interaction tensors is not completely averaged to its isotropic value, and consequently the observable is dependent upon the tensor's orientation with respect to the magnetic field direction. If the sample is uniformly aligned with respect to this direction, then the observable is single-valued. It becomes a structural constraint when the orientation of the nuclear spin tensor with respect to the molecular frame is known.

Such constraints are being used in a variety of structural efforts. Several other groups are using these constraints not only for investigations of gramicidin (Cornell *et al.*, 1988; Smith *et al.*, 1989; Prosser *et al.*, 1991; Killian *et al.*, 1992; Koeppe *et al.*, 1994), but also for the filamentous viral coat proteins (Opella *et al.*, 1987; Shon *et al.*, 1991; McDonnell *et al.*, 1993), bacteriorhodopsin (Ulrich *et al.*, 1992; Ulrich and Watts, 1993) and magainin (Bechinger *et al.*, 1993). Similar to these bilayer studies are the experiments by Sanders' and Prestegard's groups in which disk-shaped liquid crystalline domains partially align various molecules of interest in a magnetic field (Aubin and Prestegard, 1993; Sanders and Schwonek, 1993).

Such constraints differ from those of NOE, REDOR (Garbow and McWherter, 1993; Hing and Schaefer, 1993), and R^2 (Farrar *et al.*, 1993), not only because they are an orientation rather than a distance, but also because they are a constraint fixed in the laboratory frame of reference (an absolute constraint) rather than constraining the position of one site in the molecule with respect to a second site in the same molecule (a relative constraint). From orientational constraints, the structure is determined relative to the laboratory frame of reference, and hence, the orientation of the molecule is determined with respect to the alignment axis, which for this study is the bilayer normal parallel to the magnetic field direction.

One of the prime challenges to be overcome before a high-resolution structure can be determined is the need to resolve dynamic and structural effects on the nuclear spin interaction tensors. The observed magnitudes of the interactions in aligned samples must be evaluated in light of the motionally averaged tensors. Such tensor characterizations can be approximated or characterized by detailed low-temperature (<200 K) studies of rapidly frozen samples (Lazo *et al.*, 1992, 1993; Evans *et al.*, 1993). In this way, it is possible to both avoid the conformational heterogeneity induced by large ice crystals and lipid phase transition and obtain the precise determination of the tensor elements.

Gramicidin is a 15-amino-acid polypeptide consisting of an alternating sequence of L and D amino acids: formyl – L-Val – Gly – L-Ala – D-Leu – L-Ala – D-Val – L-Val – D-Val – L-Trp – D-Leu – L-Trp – D-Leu – L-Trp – D-Leu – L-Trp – ethanolamine. Both ends of the peptide are blocked so that there are no formal charges.

Although this molecule forms a variety of well-defined structures in organic solvents, they are different from the channel state. In fact, there is an interesting question about membrane insertion: How do the double-helical structures in organic solvents convert to the folding motif of the channel state, which is a single-stranded N to N dimer of 6.3 residues per turn (Urry, 1971) with a right-handed helical sense (Nicholson and Cross, 1989). In nonhydrogen bonding organic solvents, like dioxane or tetrahydrofuran, the conformations are trapped—unable to interconvert between various double-helical states (Pascal and Cross, 1992). Similarly, in a lipid environment, certain nonminimum energy conformers can be trapped (Cross *et al.*, unpublished results). Recently, the backbone conformation and tryptophan orientations of the channel state in hydrated lipid bilayers was elucidated from orientational constraints (Ketchem *et al.*, 1993) and the rest of the sidechain conformations have now been determined (Lee *et al.*, unpublished results). This article will describe the high-resolution structural and dynamic details of this structure.

Methods and Materials

Isotopically labeled amino acids were obtained from ISOTEC and Cambridge Isotope Laboratories. The amino acids have been blocked for Fmoc chemistry and the synthesis conducted as previously described (Fields *et al.*, 1989). Purification of the peptides has been achieved with a semipreparative HPLC protocol when peptides are <98% by analytical HPLC analysis (Fields *et al.*, 1989). The gramicidin and dimyristoyl-phosphatidylcholine (DMPC) were cosolubilized in an organic solvent at a molar ratio of 1:8. After the organic solvent was removed, the samples were hydrated with ~50% by weight water for the unoriented samples. For the oriented samples, the organic solution was spread on many microscope cover slips; the solvent was removed and, after stacking these glass plates in a square glass tube (inner dimension 6 mm on a side), the samples were hydrated with ~50% by weight water. (Additional details for these protocols are given in Hu *et al.*, 1993.) Fast freezing of unoriented bilayer samples was achieved by plunging a thin film of the preparation into liquid propane and storing the sample over liquid nitrogen while the process was repeated numerous times to build up a large enough sample for solid state wideline spectroscopy (Lazo *et al.*, 1993).

Spectroscopy was performed on two instruments: a narrow-bore 200-MHz Bruker/IBM spectrometer that has been heavily modified for solid state NMR and a home-built spectrometer assembled around an Oxford 400/89 magnet and a Chemagnetics data acquisition system. ^{15}N spectra were obtained with cross polarization and high-level proton decoupling. The ^{2}H spectra were obtained with a quadrupolar echo pulse sequence. Spectral processing was performed on an SGI workstation with Felix software. Simulations were carried out on a Sun

Sparc station using in-house software. Insight II from Biosym was used to display the molecular models.

Results and Discussion

Dynamics

Protein dynamics are very complicated. To address the molecular dynamics from an experimental viewpoint, it is convenient to consider three dynamic modes. First, the global motions of the channel have the lowest frequencies; second, the librational motions have the highest frequencies; and third, large-amplitude local motions have intermediate frequency. To assess the magnitude of the librational motions, samples of [^{15}N]Ala^3-gramicidin have been fast-frozen, as described above, to avoid conformational heterogeneity. Below 200 K, the librational motions do not significantly average the tensor elements. Between 200 and 263 K, the changes in individual tensor elements can be monitored. The plot in Fig. 1 clearly shows that the asymmetry of the powder pattern is changing with temperature—an observation that dictates an asymmetric motion. Frequently, librational motions average the tensor elements uniformly so that there are no changes in the tensor's asymmetry. Here, the librational averaging is effectively modeled with librations about an axis collinear with the Cα-Cα vector. The librational amplitudes are relatively uniform along the channel axis and are typically ±20° (Lazo *et al.*, unpublished results).

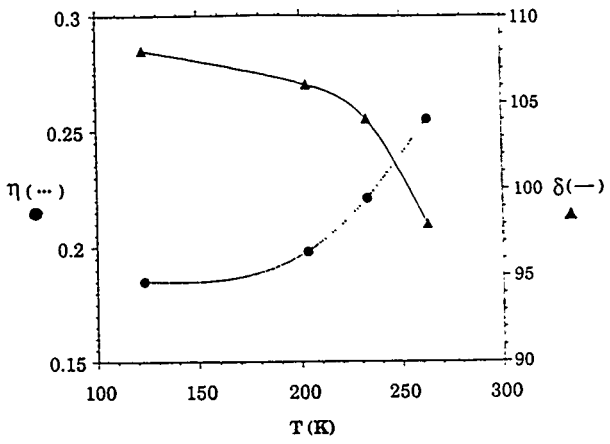


Fig. 1. The anisotropy (δ) and asymmetry (η) of the ^{15}N chemical shift tensors for the [^{15}N]Ala^3 site in the gramicidin channel. The change in asymmetry dictates anisotropic librational motions—motions that potentially rotate the carbonyl oxygens into the channel for solvating cations.

Global motion has been accurately assessed over a temperature range of 15 to 36°C—corresponding to correlation times from 4.3 ms to 29 μs, respectively (Lee *et al.*, 1993). This determination was achieved from the powder pattern spectra of the methyl deuterons of Ala^3 in hydrated lipid bilayer preparations of gramicidin. Figure 2 shows very similar data for the Ala^5 site in gramicidin. For this site, it is assumed that the only other motions experienced (such as the χ_1 motion for the methyl or librational motions about the Cα-Cβ axis) will be orders of magnitude more rapid than

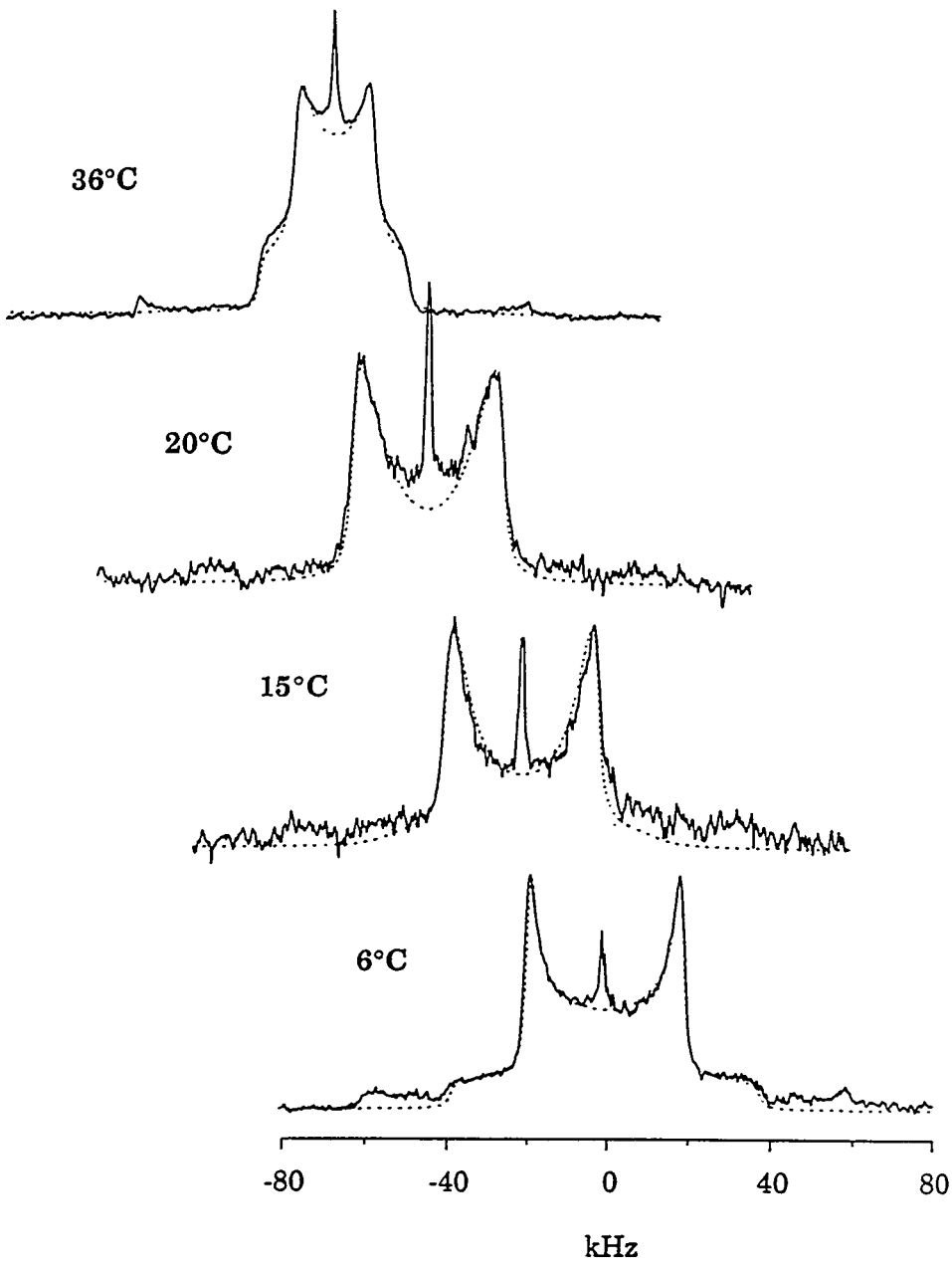


Fig. 2. ^2H powder pattern spectra of the [U-$^2H_4$]Ala^5-gramicidin in hydrated lipid bilayers. The signals are dominated by the methyl deuterons. The simulations model the cessation of global motion (Lee et al., *1993) with a constant value for librational amplitude. The conclusion is that librational amplitudes remain virtually constant over this temperature range.*

the global motions. Therefore, the observation of intermediate motional averaging is readily assigned to the global motions of the channel.

Averaging of the powder pattern spectra from 6°C (the global static limit: ≥1 s) to 52°C (the global fast exchange limit: ≤10 μs) can be completely accomplished with a single variable—global correlation time. The averaging due to this motion is dependent on the orientation of the Cα-Cβ axis, which was determined independently from the structural studies to be 102° (Ketchem *et al.*, 1993). There is no significant increase in the librational amplitudes after raising the temperature from 6 to 52°C. This result suggests that the majority of the librational amplitude is induced below 6°C, and furthermore, the librational amplitudes determined at -10°C have been effective in simulating these variable temperature spectra. If this result is valid, it further suggests that the lipid phase transition has relatively little effect on the amplitude of local motions. Consequently, we have continued to use spectra at temperatures below the phase transition, where the global motion is in the static limit and yet local motions can be effectively characterized.

There are only two residues that show large-amplitude local motions: Val^1 and Val^7. These sites undergo three-state jumps about their χ_1 axis. The populations of these states can be determined through simulation of low-temperature spectra, such as those shown in Fig. 3.

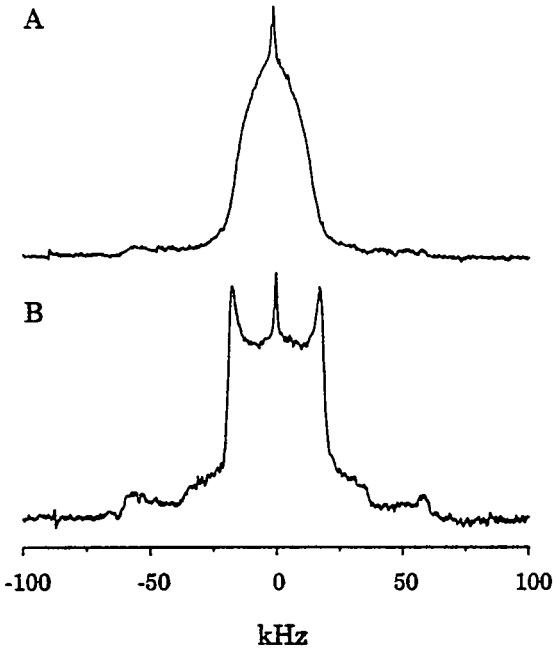


Fig. 3. ^2H powder pattern spectra of the [U-$^2H_8$]Val-labeled gramicidin in hydrated lipid bilayers at 5°C. As in Fig. 2, the signals are dominated by the methyl deuterons. (A) Val^1-labeled gramicidin shows a motionally averaged spectrum characteristic of large-amplitude motions occurring in the fast exchange limit. (B) The Val^8-labeled gramicidin shows no evidence of large-amplitude motions.

For Val^1, the trans conformation (χ_1 = 180°) dominates, and the gauss$^-$ and gauss$^+$ states have approximately equal populations. For these three states, the distributions are 75:10:15, respectively.

Using this approach, a detailed model of the channel dynamics can be achieved

within the assumptions made. These models describe the modes for the motion (that is, whether diffusional or jump motions), the axis about which the motion is occurring, and the amplitudes for the motions. Spectral densities can now be calculated and relaxation rates predicted, based on the experimentally defined motional model. A comparison with experimental relaxation rates, therefore, permits determination of the motional frequencies. Much of this work is still in progress for the gramicidin channel, but the initial results for the backbone librational frequencies generated two interesting conclusions. First, the motional frequencies obtained from ^{15}N relaxation measurements for various sites all suggest 10-ns timescale motions, a result that is 3 to 4 orders of magnitude slower than the rates predicted from molecular dynamics calculations (North and Cross, 1993). Second, the backbone librational amplitudes are much smaller than the amplitudes determined from powder pattern averaging results (North and Cross, unpublished results). This latter analysis is sensitive to all frequencies greater than 10^4 Hz, whereas relaxation analysis is sensitive only to the motions in the vicinity of the Larmor frequency. Consequently, it appears that the librational motions may be composed of several modes or frequency ranges, each with their own amplitude. Such experimental determinations of molecular dynamics will provide useful gauges against which the computational methods and their force fields can be refined.

Structure

After the dynamics have been estimated, the motionally averaged tensors can be predicted (if not observed directly), and then it is possible to determine the structure of the polypeptide from orientational constraints. Figure 4 shows an example of high-resolution orientational constraints from [2,3-$^2H_3$]Leu^{12}-labeled gramicidin. If the linewidth were completely due to the disorder in orientations of the channels with respect to the magnetic field direction, this disorder would be much less than 1°. Because this linewidth has significant relaxation contributions as well as contributions from coupling to nearby ^1Hs, the alignment of these

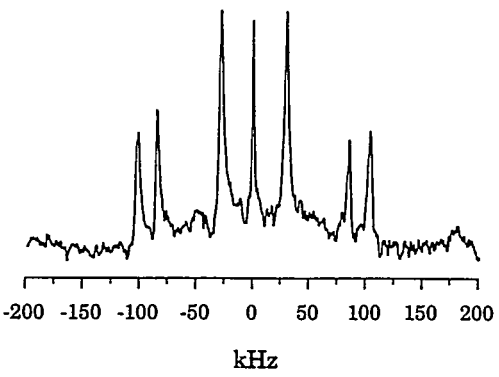


Fig. 4. ^2H NMR spectra of [2,3-$^2H_3$]Leu^{12}-gramicidin in uniformly aligned lipid bilayers so that the channel axis is parallel to the magnetic field. Based on the backbone structure and analysis of the [α,methyl-$^2H_7$]Leu^{12}-gramicidin spectra, it was possible to make the following assignments: α, +207 kHz; β1, +171 kHz; and β2, +49 kHz.

samples is remarkable. Although the bilayers are aligned between glass plates, the high degree of alignment is the result of magnetic alignment caused by the anisotropy of the diamagnetic susceptibility, which—for this sample and for α-helical membrane proteins—will tend to align the bilayer normal parallel with respect to the magnetic field.

The backbone structure of the channel has been determined by first obtaining the orientation of the ^{1}H-^{15}N and $^{13}C1$-^{15}N bonds with respect to the magnetic field through the respective dipolar interactions. This process required preparation of both single-site and doubly labeled gramicidin for each peptide plane. Assuming (only for the initial structure) that the peptide plane is indeed planar, the orientations of these two vectors thereby define the orientation of the peptide plane with respect to the magnetic field and the bilayer normal. By performing such an operation for adjacent peptide linkages and by taking advantage of the tetrahedral geometry of the shared α-carbon, it is possible to determine the relative orientation of the two planes and, hence, their torsion angles (Teng *et al.*, 1991). This analysis dictates the hydrogen bonding pattern and the folding motif for the channel, but orientational constraints such as the chemical shifts have not been optimally used to determine the initial structure. A refinement protocol based on a form of simulated annealing modifies the structure slightly, calculates the NMR observables and the hydrogen bonding lengths, and compares these with the observed NMR constraints and typical hydrogen bond lengths. The result is the development of a penalty function that can be minimized to yield a significantly improved structure; as a consequence, the perfect trans geometry for the peptide linkages has been relaxed, and the final result has the ω torsion angle varying between -175 and +171° (Ketchem *et al.*, 1993).

Further refinement from additional orientational constraints is still possible. A specific example is provided by having both the ^{15}N and $^{13}C1$ chemical shifts within a single peptide linkage. Through the process shown in Fig. 5, it is possible to analytically determine the ω torsion angle. This illustration shows the computed value for each chemical shift as a function of rotation about the C-N bond. For the ^{15}N chemical shift, the plane being rotated about this axis is defined by Cα-N-C1; for the ^{13}C chemical shift, the plane is defined by N-C1-Cα. The local maxima for these curves represent the parallel orientation of these planes with respect to the magnetic field. Because these planes are so close to parallel with respect to the magnetic field, the accuracy with which the ω torsion angle is determined is not high. However, the approach is demonstrated here; only from the highest resolution crystal structures has this information been previously obtained directly from experimental data (that is, without the aid of an energy-based computational method).

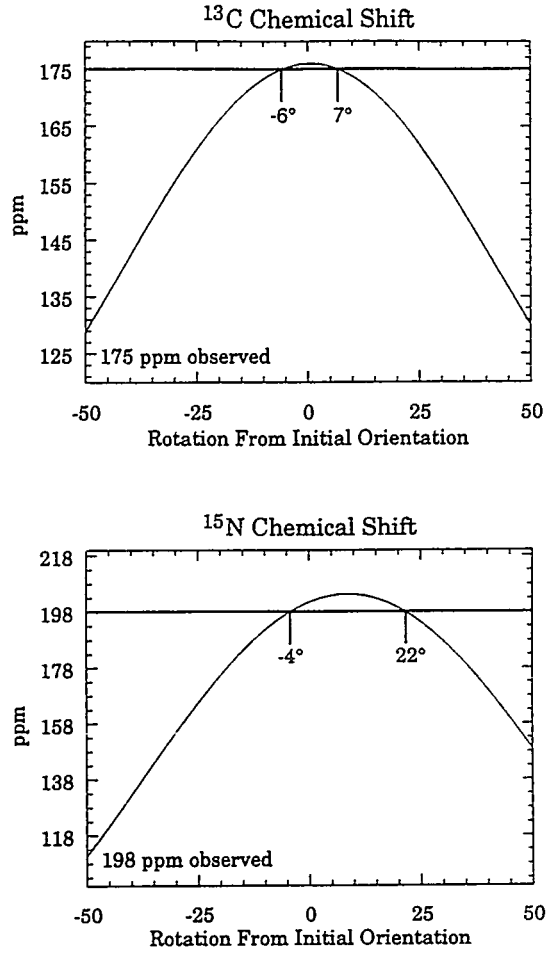


Fig. 5. Chemical shift as a function of rotation about the C1-N bond. The initial orientation is from the refinement protocol described in the text, and for this Gly^2-Ala^3 peptide plane the mean orientation is very close to parallel with the channel axis. The omega torsion angle is 171° from this refined structure. Here it is shown that in the absence of other constraints, the optimal orientation for the ^{13}C chemical shift tensor has the carbonyl group rotated another 6° into the pore and for the ^{15}N chemical shift the amide proton is rotated 4° away from the channel, resulting in an omega angle of 173°.

Unique sidechain structures can be achieved for all of the residues except tryptophan. For the four Trp sites, a unique orientation of the indole ring with respect to the channel axis and magnetic field direction can be achieved, but four χ_1 / χ_2 pairs are still possible, even with a defined Cα - Cβ axis orientation from the backbone structural analysis. Figure 6 shows the root-mean-square deviation between the observed bond orientations from quadrupolar splittings for the five indole carbon-bound deuterons as well as the predicted values for all χ_1 and χ_2 values. This plot clearly shows four equivalent minima. Mechanisms have been used to justify a most probable structure, but the necessary solid state NMR data have not yet been recorded to distinguish between these conformers.

With the unique orientation of the indole rings with respect to the channel axis and the bilayer normal, it is clear that the indole N-H bond is oriented toward the bilayer surface for hydrogen bond formation. Furthermore, the dipole moment associated with the indole ring is predominantly axial rather than perpendicular to the channel axis. Both facts about this high-resolution structure have considerable impact on the function of the channel. The hydrogen bond formation impacts on the dimerization constant, and it is thought that the migration of the indoles to the bilayer surface for hydrogen bonding may provide the energetic driving force for the conformational rearrangements associated with membrane insertion

(Zhang *et al.*, 1992). The dipole moments influence the potential energy surface for the transit of monovalent cations through the channel.

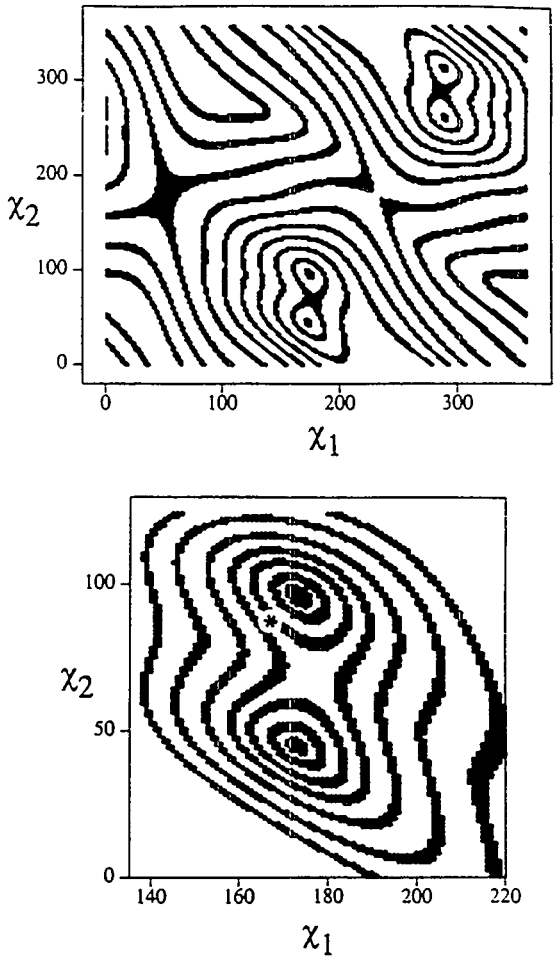


Fig. 6. Root-mean-square deviations between experimental and predicted bond orientations for the [indole-$^2H_5$]Trp^9 sidechain of gramicidin. The contour step for the full plot is 5° and for the expansion it is 2°. The 'star' identifies the χ_1 and χ_2 angles for this residue in the Arseniev structure determined in SDS micelles by solution NMR (Lomize et al., *1992).*

Figure 7 shows a side view and end view of the complete structure, with the exception of the ethanolamine group for which structural and dynamic information has yet to be resolved from the data. The sidechain conformations represent initial structures, whereas the backbone structure is refined. As with the backbone conformation, it is expected that the initial structure will be close to the refined conformation. There is little core-core overlap of the atoms, but there is some, and such incongruities must be resolved through a refinement protocol. As with the backbone, it is not anticipated that the rotameric state or even any large torsion angle changes will take place upon refinement.

Several very interesting features are apparent from this initial structure. First, the sidechains appear to be tightly packed, giving rise to an unusual shape, but this shape was predicted many years ago, albeit for a left-handed helix (Urry, 1971). Second, the surface of the channel is rough, with a very substantial surface for lipid interactions. Third, there appears to be considerable backbone exposed surface, especially near the monomer - monomer junction. This may permit sequestering of solvent water on the exterior of the channel, which could play an important role in the monomer - dimer equilibrium. Fourth, the indole nitrogen sites for the four tryptophans are within 3.5 Å on the Z axis of the channel. Consequently, all the indole N-H groups appear to be within hydrogen bonding and solvent exchange

distance with the hydrophilic-hydrophobic interface. Hydrogen bonding may be with the fatty acyl linkages, the phosphate groups, or waters of hydration. These hydrogen bonds provide the mechanism by which the channel orients itself with respect to the bilayer surface.

The local dynamics of the gramicidin chain provides surprising insights. With the exception of two valine residues, there are no large-amplitude local motions. On the other hand, there are significant librational amplitudes for the backbone and all of the sidechains. Compared to the high degree of gauss/trans isomerization in the fatty acyl chains, the polypeptide may appear to be quite rigid—as if there is a dynamic phase boundary between the peptide and lipid environments. Many computationalists have been forced to consider membrane proteins as rigid cylinders in a dynamic lipid matrix. Although this has generally been viewed as a rather gross assumption, it now appears that the peptide is indeed quite rigid. The considerable amplitude for the librational dynamics is also surprising. In the low dielectric environment of the bilayer center, α-helical proteins can be expected to be very rigid. The lack of exposure of the backbone to a hydrogen-bond-exchanging environment will result in a helix that is far more rigid than it would be in an aqueous environment. The gramicidin backbone, however, is exposed to a considerable aqueous surface from the pore and, possibly, even from the lipid side of the backbone, where waters of hydration may be sequestered.

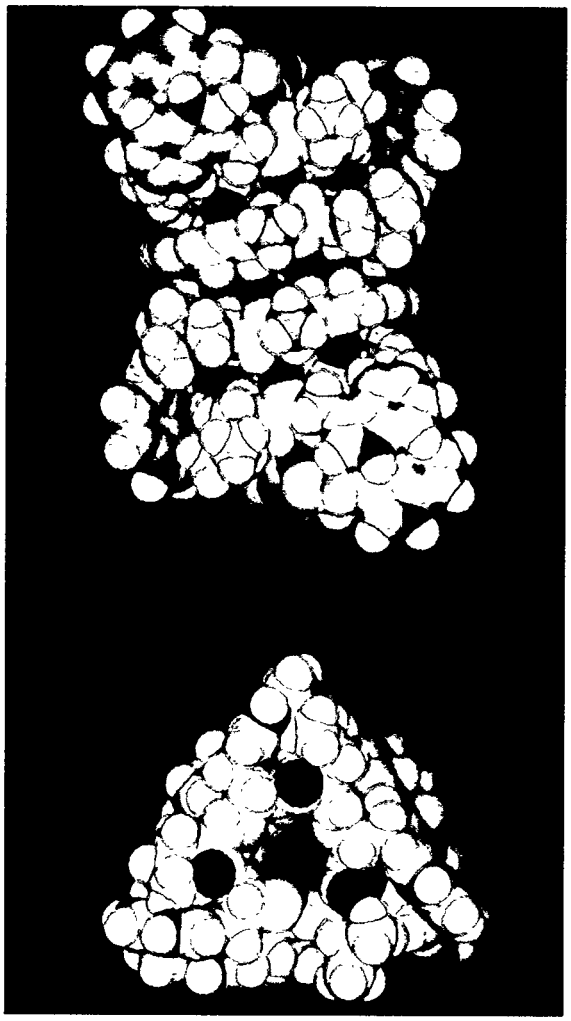


Fig. 7. The gramicidin channel conformation with the backbone structure refined and the sidechain structure in its initial form. These views show the indole N-H orientations oriented toward the bilayer surface and that significant segments of the backbone are exposed on the exterior toward the lipid environment. The endview shows the efficient packing of the sidechains and a narrow pore that can accept only a single file of water molecules.

Therefore, although there are few large-amplitude local motions, the structure has surprisingly large librational amplitudes.

Through the high-resolution structure and dynamics of this lipid bilayer bound polypeptide, we are gaining new insights into the membrane environment for proteins as well as how this peptide performs its channel activity.

Acknowledgements

We thank the staff of the Florida State University NMR Facility (J. Vaughn, R. Rosanske, and T. Gedris) for their maintenance, service, and modifications of the spectrometers, as well as the staff of the FSU Bioanalytical Synthesis and Services Facility (H. Henricks and U. Goli) for their expertise and maintenance of the ABI 430A peptide synthesizer and HPLC equipment. This effort was supported by NIH grant AI-23007.

References

Aubin, Y. and Prestegard, J.H. (1993) *Biochemistry 32*, 3422-3428.

Bechinger, B., Zasloff, M., and Opella, S.J. (1993) *Protein Sci. 2*, 2077-2084.

Cornell, B.A., Separovic, F., Baldassi, A.J., and Smith, R. (1988) *Biophys. J. 53*, 67-76.

Evans, J.N.S., Appelyard, R.J., and Shuttleworth, R.J. (1993) *J. Am. Chem. Soc. 115*, 1588-1590.

Farrar, M.R., Lakshmi, K.V., Smith, S.O., Brown, R.S., Raap, J., Lugtenburg, J., Griffin, R.G., and Herzfeld, J. (1993) *Biophys. J. 65*, 310-315.

Fields, C.G., Fields, G.B., Noble R.L., and Cross, T.A. (1989) *Int. J. Pept. Prot. Res. 33*, 298-303.

Garbow, J.R. and McWherter, C.A. (1993) *J. Am. Chem. Soc. 115*, 238

Hing, A.W. and Schaefer, J. (1993) *Biochemistry 32*, 7593-7604.

Hu, W., Lee, K.-C., and Cross, T.A. (1993) *Biochemistry 32*, 7035-7047.

Ketchem, R.R., Hu, W., and Cross, T.A. (1993) *Science 261*, 1457-1460.

Killian, J.A., Taylor, M.J., and Koeppe, R.E. II (1992) *Biochemistry 31*, 11283-11290.

Koeppe, R.E., II, Killian, J.A., and Greathouse, D.V. (1994) *Biophys. J. 66*, 14-24.

Lazo, N.D., Hu, W., and Cross, T.A. (1992) *J. Chem. Soc. Chem Commun.* 1529-1531.

Lazo, N.D., Hu, W., Lee, K.-C., and Cross, T.A. (1993) *Biochem. Biophys. Res. Commun. 197,* 904-909.

Lee, K.-C., Hu, W., and Cross, T.A. (1993) *Biophys. J. 65,* 1162-1167.

Lomize, A.L., Orechov, V.Yu., and Arseniev, A.S. (1992) *Bioorg. Khim. 18,* 182.

McDonnell, P.A., Shon, K., Kim, Y., and Opella, S.J. (1993) *J. Mol. Biol. 233,* 447-463.

Nicholson, L.K. and Cross, T.A. (1989) *Biochemistry 28,* 9379-9385.

North, C.L. and Cross, T.A. (1993) *J. Magn. Reson. B101,* 35-43.

Opella, S.J., Stewart, P.L., and Valentine, K.G. (1987) *Quart. Rev. Biophys. 19,* 7-49.

Pascal, S.M. and Cross, T.A. (1992) *J. Mol. Biol. 226,* 1101-1109.

Prosser, R.S., Davis, J.H., Dahlquist, F.W., and Lindorfer, M.A. (1991) *Biochemistry 30,* 4687-4696.

Sanders, C.R. and Schwonek, J.P. (1993) *Biophys. J. 65,* 1460-1469.

Shon, K., Kim, Y., Colnago, L.A., Valentine, K.G., and Opella, S.J. (1991) *Science 252,* 1303-1305.

Smith, R., Thomas, D.E., Separovic, F., Atkins, A.R., and Cornell, B.A. (1989) *Biophys. J. 56,* 307-314.

Teng, Q., Nicholson, L.K., and Cross, T.A. (1991) *J. Mol. Biol. 218,* 607-619.

Ulrich, A.S., Heyn, M.P., and Watts, A. (1992) *Biochemistry 31,* 10390-10399.

Ulrich, A.S. and Watts, A. (1993) *Solid State NMR 2,* 21-36.

Urry, D.W. (1971) *Proc. Natl. Acad. Sci. USA 68,* 672-676.

Zhang, Z., Pascal, S.M., and Cross, T.A. (1992) *Biochemistry 31,* 8822-8828.

Isotopic Chirality

Heinz G. Floss

Department of Chemistry
BG-10
University of Washington
Seattle, WA 98195

This paper deals with compounds that are chiral—at least in part, due to isotope substitution—and their use in tracing the steric course of enzyme reactions *in vitro* and *in vivo*. There are other applications of isotopically chiral compounds (for example, in analyzing the steric course of nonenzymatic reactions and in probing the conformation of biomolecules) that are important but they will not be discussed in this context.

One of the remarkable features of enzymes is the fact that they catalyze chemical reactions with usually complete stereospecificity. This is a consequence of the fact that enzymes are made up entirely of optically active building blocks and are thus "perfect asymmetric catalysts." Their stereospecific action extends not only to chiral centers in substrate molecules, but also to centers that are not by nature chiral. The stereospecificity of reactions at the latter centers is cryptic (Hanson and Rose, 1975); it is hidden from the observer, and its elucidation requires rendering the particular center artificially chiral by isotope substitution. The application of isotopic chirality in biochemistry goes back to a seminal paper by Ogston (1948), in which he reinterpreted the conclusions drawn from earlier tracer studies on the Krebs cycle. The earlier workers had excluded citrate as an intermediate in the Krebs cycle because the tracer data did not show the randomization of label thought to result from involvement of a symmetrical intermediate like citrate. Ogston pointed out that the two carboxymethylene groups of citrate are distinct and distinguishable by an enzyme, indicating that citrate could well be an intermediate in the pathway.

Citrate, like many other biochemical compounds, represents a prochiral system (Hanson, 1966), a center of the *Caabd* type,

in which the two like substituents are intrinsically different; they are heterotopic because they occupy different positions in space. In the interaction of such a system with a chiral catalyst like an enzyme, two stereochemical features can be evident. First, the enzyme can distinguish the two heterotopic groups and can react with one to the complete exclusion of the other. Second, in reactions at such a center, the replacement or elimination of one of the groups can occur with a particular stereochemistry; for example, retention or inversion in a displacement reaction. Both features are usually observed in enzyme-catalyzed reactions at prochiral centers. Virtually hundreds of examples of the elucidation of the stereochemistry at such prochiral centers in the literature have contributed to our understanding of enzyme catalysis (Alworth, 1972; Bentley, 1969, 1970; Rétey and Robinson, 1982). In following sections, a few examples of our work will serve for purposes of illustration.

Many antibiotics contain 2,6-dideoxy-hexose moieties, which are crucial to their biological activity. These sugar moieties are generally derived from glucose through a series of transformations at the sugar nucleotide level. In the process, the oxygen functions at C2 and at C6 of glucose must be removed, and some time ago we became interested in the stereochemistry and mechanism of these deoxygenation reactions. One of the systems we studied is the antibiotic granaticin, a metabolite of *Streptomyces violaceoruber* Tü22 (Snipes *et al.*, 1979). To determine the steric course of deoxygenation at C2, we fed the organisms D-[2-$^2H_1$]glucose and analyzed the resulting granaticin by proton NMR spectroscopy, observing the signal for 3'H in the sugar moiety. As shown in Fig. 1, the signal for 3'H in the unlabeled granaticin shows a large coupling to 2'H_R and a small coupling to 2'H_S. In the deuterated sample from the feeding experiment, the large coupling was fully retained, but the small coupling had largely disappeared, indicating that the deuterium is located in the pro-*S* position. Therefore, replacement of the 2-oxygen of glucose by an unlabeled hydrogen has occurred in a retention mode. The corresponding experiment on the antibiotic chlorothricin, a metabolite of *S. antibioticus* Tü99 (Lee *et al.*, 1986), showed that the deuterium occupied the pro-2*R* position in the sugar moieties, which indicates replacement of the oxygen in an inversion mode. Thus, these two seemingly identical reactions proceed with different stereochemistry and are not catalyzed by the same type of enzyme. They may involve different mechanisms or, more likely, there is no mechanistic imperative for a particular stereochemistry, and the observed difference reflects independent evolution of the two enzymes.

Isotopic chirality is obviously not limited to labeling with hydrogen isotopes. A prochiral system of considerable biochemical utility is glycerol, in which the two hydroxymethyl groups have different metabolic fates. The pro-*S* hydroxymethyl

group is phosphorylated by glycerol kinase and the pro-*R* hydroxymethyl group in the further metabolism eventually ends up as the aldehyde function of phosphoglyceraldehyde. Thus, glycerol labeled stereospecifically with ^{13}C in one of the hydroxymethyl groups can be used to probe the orientation

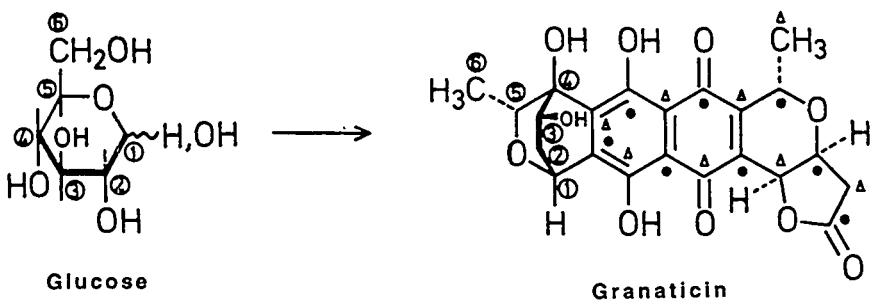
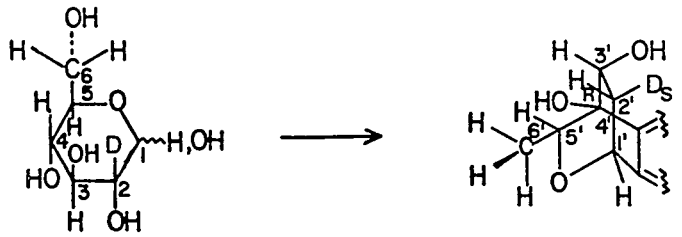
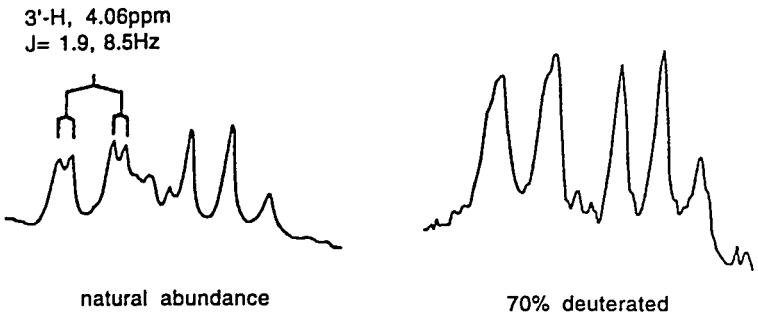


Fig. 1. Origin of the granaticin dideoxyhexose moiety from D*-glucose and the stereochemistry of the 2-deoxygenation reaction.*





in which triose phosphates are incorporated into various metabolites, as exemplified in a study on the biosynthesis of the antibiotic asukamycin in *Streptomyces nodoccus*, ssp. *asukaensis* (Cho *et al.*, 1993). The ^{13}C-NMR analysis of asukamycin derived from 2*R*-[2-$^{13}C_1$]glycerol showed that the label in the central C_7N unit was located exclusively at the carbonyl carbon (Fig. 2), defining the mode of incorporation of the triose phosphate into this structural moiety. This information was useful in deriving a hypothetical pathway for the formation of this moiety.

Asukamycin contains another unusual structural element: a cyclohexane ring attached to a short polyolefin chain. Such cyclohexyl moieties are rare in nature but are found in a few other antibiotics, particularly as cyclohexanecarboxylic acid in ansatrienin, a metabolite of *S. collinus*, and in the ω-cyclohexyl fatty acids of thermoacidophilic and mesophilic bacteria.

Cyclohexanecarboxylic acid, the precursor of all these cyclohexyl moieties, is formed from shikimic acid in a remarkable series of reactions involving a sequence of reductions and dehydrations that avoid aromatization of the system. While studying this notable transformation, we determined the fate of every hydrogen atom in the shikimate molecule by deuterium labeling. In addition, we established that

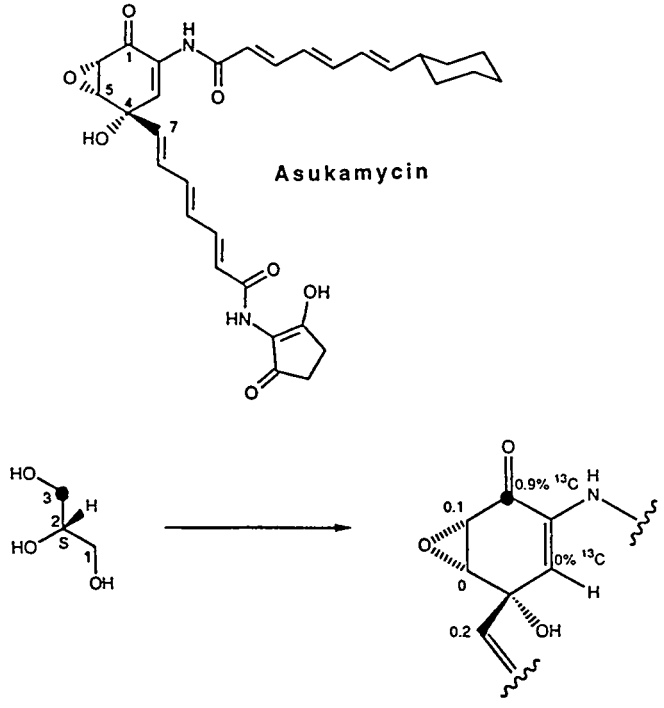


Fig. 2. Mode of incorporating sn[3-^{13}C]glycerol into the antibiotic asukamycin.

the two sides of the shikimate ring remain distinct throughout the reaction sequence. Because cyclohexanecarboxylic acid is achiral, the two heterotopic edges of the molecule cannot be distinguished by direct NMR observation, but only after derivatization of the molecule with a chiral auxiliary. This established that carbon 2 of shikimate gives rise to C6 in the pro-*S* edge of cyclohexanecarboxylic acid. With this information at hand, further analysis of the deuterated products allowed an unequivocal assignment of the stereochemistry at each of the deuterated chiral centers of the molecule (Fig. 3). From this and some additional information, we were able to write a detailed reaction sequence for this transformation and could define the stereochemistry of all but one of the reactions.

Whereas, in the previous example, the elucidation of the steric course of the reaction sequence required synthesis of numerous labeled shikimic acid samples, in another instance, we were able to generate a wealth of stereochemical information from just two biosynthetic experiments. This work dealt with the biosynthesis of the antibiotic thiostrepton in *S. laurentii*. This antibiotic is derived entirely from a number of amino acids; serine particularly is incorporated numerous times into the molecule, both directly and after conversion into cysteine or tryptophan. Many reactions in this biosynthesis involve transformations at C3 of serine, and the steric course of these reactions was probed by feeding two samples of serine, one dideuterated at C3 and the other stereospecifically monodeuterated with 3*S* configuration. In addition, each of the two serine samples, which were synthesized for us by Dr. Cliff Unkefer of the Stable Isotope Resource at Los Alamos National Laboratory, contained a ^{13}C-label at C3. This permitted analysis of the resulting thiostrepton samples by ^{13}C-NMR with simultaneous proton and deuterium broadband decoupling. One observes the enriched carbon signals for all the positions labeled by C3 of serine and, if these carbons retain one deuterium, a deuterium isotope-shifted signal ~0.25 ppm upfield from the nondeuterated carbon resonance. The deuterium isotope shift is additive; that is, two deuteriums shift by ~0.5 ppm. In this way, one can directly determine the number of deuterium

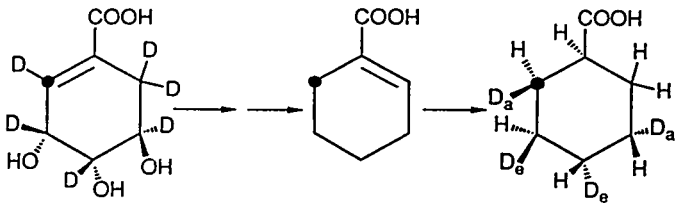


Fig. 3. Labeling pattern of the cyclohexane- and cyclohexene-carboxylic acid moieties of ansatrienin antibiotics from various ^{13}C- and deuterium-labeled shikimic acids.

atoms retained at each of the positions derived from C3 of serine. In instances where a methylene group carries only one deuterium, the configuration can be determined either by repeating the experiment with single-frequency deuterium decoupling or by running a proton-carbon correlation spectrum with deuterium decoupling (Reese *et al.*, 1986). From this experiment (Fig. 4), we learned (Mocek *et al.*, 1993) that the dehydration of serine to achieve dehydroalanine moieties proceeds with *anti* stereochemistry. The formation of the quinaldic acid moiety by ring expansion of a molecule of tryptophan proceeds with stereospecific loss of the pro-3*S* hydrogen of the sidechain, as does the formation of the thiazole rings from cysteine. Both deuterium atoms from C3 of serine are retained at the single dihydrothiazole ring in the molecule, indicating that the latter is not derived by reducing a thiazole, but rather must be an intermediate in the formation of the thiazole. Because the configuration at this dihydrothiazole ring corresponds to that of a D-cysteine, it also follows that at least that amino acid in the peptide chain—and possibly others—must have a D configuration. Finally, the results also establish the stereochemistry of the rather unique formation of a tetrahydropyridine ring from a tail-to-tail condensation of two molecules of serine,

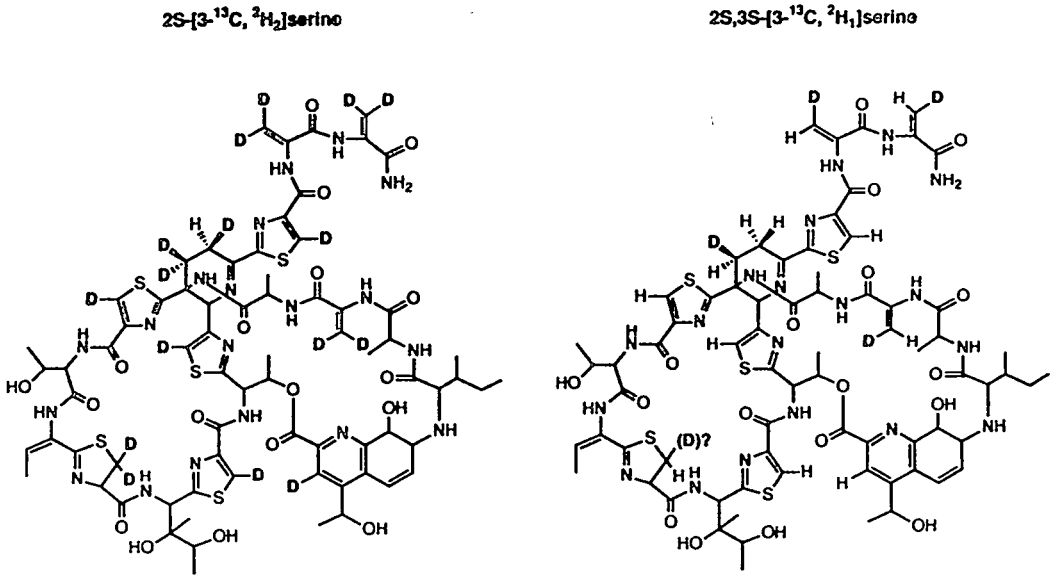


Fig. 4. Deuterium distribution in the antibiotic thiostrepton biosynthesized from (2S)-[3-^{13}C,$^{2}H_2$]-serine and (2S,3S)-[3-^{13}C,$^{2}H_1$]serine.

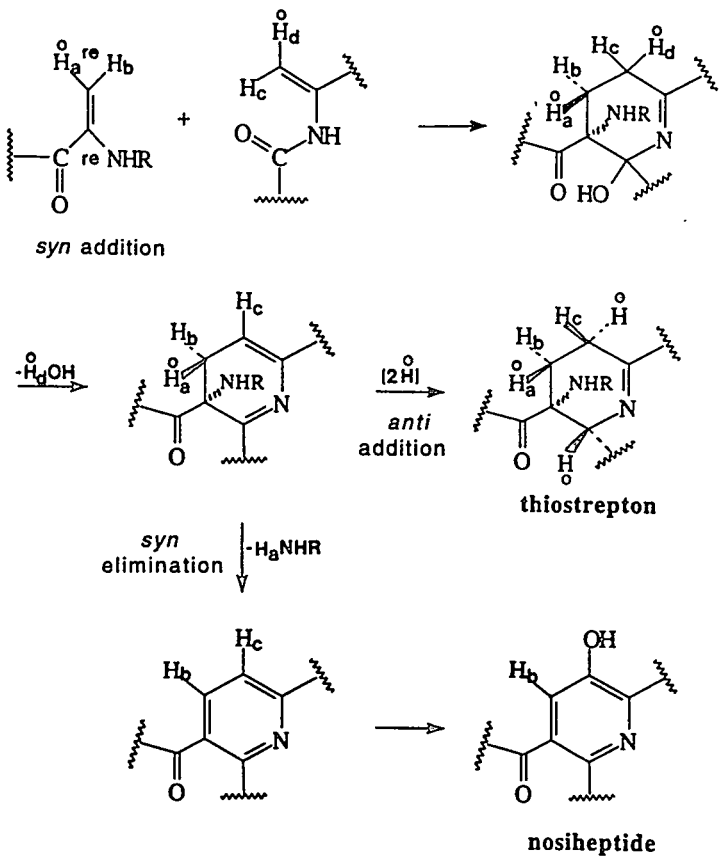


Fig. 5. Stereochemistry and proposed mechanism of formation for the pyridine and tetrahydropyridine moieties in the antibiotics nosiheptide and thiostrepton.

as shown in Fig. 5. These two simple feeding experiments have generated stereochemical information on an entire range of different reactions in this biosynthesis.

Isotopic chirality is not limited to achiral systems containing heterotopic ligands. Even a system in which the ligands are homotopic (that is, indistinguishable even to a chiral agent; for example, a methyl group attached to some moiety R) can be rendered chiral by virtue of isotope substitution. Obviously, in the case of such a *Caaab* system, one needs three different isotopes to generate a chiral version; for instance, ^{1}H, ^{2}H, and ^{3}H in the case of a methyl group, or ^{16}O, ^{17}O, and ^{18}O in the case of a phosphate group (Floss *et al.*, 1984). In such systems, although the enzyme cannot distinguish the three homotopic substituents, it can still replace any one of them in a certain stereochemical mode, retention or inversion, or it can transfer the entire group to a different ligand in a stereospecific

manner. The difficulty in working with such systems lies not so much in the synthesis of an isotopically chiral version—this can be done more or less by conventional chemistry—but in devising a method to distinguish an *R* from an *S* configuration. For the chiral methyl group, this problem was solved quite elegantly by the research teams of Arigoni and Cornforth (Cornforth *et al.*, 1970; Lüthy *et al.*, 1969) using a primary kinetic isotope effect in an enzymatic replacement of one of the hydrogens at the chiral methyl group. This "chiral methyl group" methodology has since been employed to analyze the stereochemistry of a wide range of biochemical reactions (Floss and Tasi, 1979). One example is methyltransfer reactions, in which the methyl group is transferred from the sulfur of *S*-adenosylmethionine to various other nucleophiles. Analyses of a large number of such reactions (Arigoni, 1978; Floss and Lee, 1993) has shown that they overwhelmingly proceed with inversion of configuration of the methyl group, indicating a single direct S_N2-type transfer of the methyl group from the sulfur to the acceptor nucleophile. Only a very few reactions show retention stereochemistry (Floss and Lee, 1993) and thus involve a different mechanism. Overall, however, there seems to be a strong mechanistic imperative favoring the inversion stereochemistry.

There is a second type of a pro-prochiral center, a type *Caabb*, containing two paired ligands, which again are homotopic. The biochemically most prominent example of this type of molecule is malonic acid, a precursor of fatty acids and various polyketide natural products. Again, a chiral version of this molecule can be synthesized; for example, by placing ^{13}C into one carboxyl group and replacing one methylene hydrogen by deuterium. Both our laboratory and that of Peter Jordan in England accomplished the synthesis of (2*R*)- and (2*S*)-[1-$^{13}C_1$,2-$^{2}H_1$]malonate by very similar routes (Fig. 6) (Huang *et al.*, 1986; Jordan and Spencer, 1991). The experimental challenge here lies in the fact that malonic acid undergoes hydrogen exchange at C2 at appreciable rates, and thus loses its stereochemical integrity when kept for extended periods of time in aqueous solution. Therefore, both groups synthesized the stable chiral precursor malate, which was then converted to malonate under carefully controlled conditions immediately before use. Both groups demonstrated that the material was indeed optically active. Our analysis involved reduction to 1,3-propanediol followed by derivatization with a chiral auxiliary, followed by proton NMR analysis (Huang *et al.*, 1986). The Jordan group converted their malonate samples into fatty acids, which were then analyzed by mass spectrometry (Jordan and Spencer, 1991). Both laboratories then used these samples to examine the steric course of polyketide biosyntheses, but there is currently some controversy on the results of these analyses (Wood *et al.*, 1989; Spencer and Jordan, 1992). The principle of this analysis is based on the fact that one isomer with

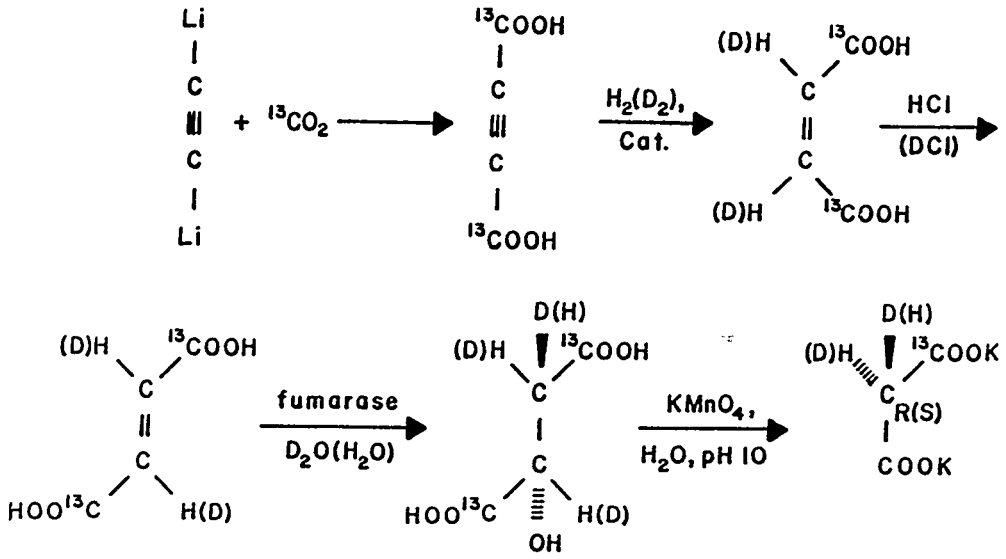


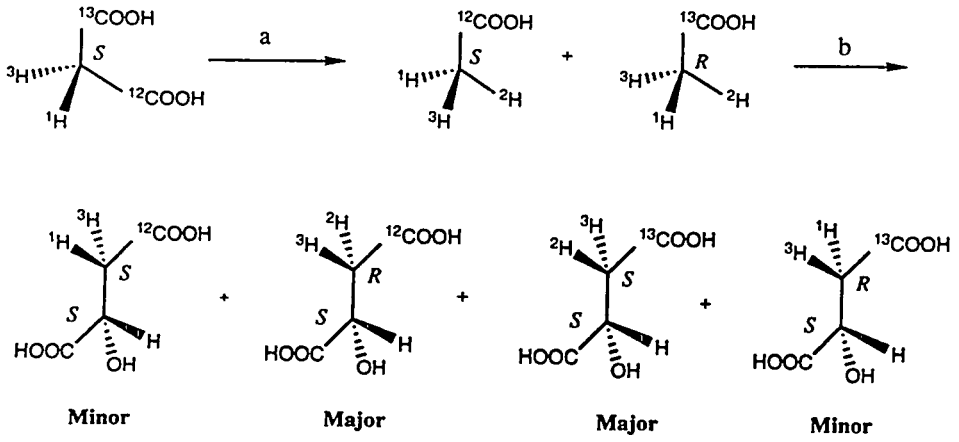
Fig. 6. Synthesis of potassium (R)- and (S)-[1-^{13}C, 1,2-$^{2}H_2$]malonate.

a given steric course of the reactions gives "malonate units" in the polyketide, which contain either ^{13}C or deuterium, whereas from the opposite enantiomer of malonate, one-half of the units contain both isotopes and the other one-half contain neither isotope. These patterns can be distinguished by mass spectral analysis of the products. However, the mass spectral patterns obtained from the two enantiomers are not very much different, which led the two groups to different interpretations. A more definitive analysis should be possible by NMR spectroscopy, and therefore we have recently repeated the syntheses of chiral malonate, incorporating tritium instead of deuterium, which will allow product analysis by tritium NMR spectroscopy.

These tritiated chiral malonate samples are also being used to study another interesting reaction, the decarboxylation of malonic acid to acetic acid by an enzyme from *Malonomonas*. This reaction is intriguing, not only because of its biochemical mechanism, but also because it involves the conversion of one pro-prochiral system of the *Caabb* type into another of the *Caaab* type. This again presents the challenge of devising a suitable method of analysis. The approach we used is shown in Fig. 7. The mixture of malate samples obtained after conversion of the resulting acetate into acetyl CoA and reaction with glyoxylate catalyzed by malate synthase is distinguishable from the set arising from the opposite enantiomer of malonate by tritium NMR

spectroscopy. Analysis showed that the decarboxylation of malonate proceeds stereospecifically in a retention mode (Micklefield *et al.*, unpublished results).

Finally, it should also be possible to generate a chiral version of a system carrying four identical ligands on a carbon atom; that is, a *Caaaa* system. A prominent example is inorganic phosphate that has been prepared in a chiral form, $H_3PO(^{17}O,^{18}O)S$ (Tsai and Chang, 1980; Webb, 1980). The simplest such system, of course, is methane. This is more than just an academic curiosity because there are some important biochemical reactions in which methane is generated or used, and their stereochemical course is of interest. The obvious problem, however, is that four isotopes of hydrogen would be required to generate a chiral version of methane, but nature has only provided three. Fortunately, the enzyme we wished to study (methane monooxygenase (MMO) from *Methylosinus trichosporium*, which catalyzes the hydroxylation of methane to methanol) will use ethane as an alternate substrate. Thus, we can use an additional unlabeled methyl group in place of the fourth isotope of hydrogen; that is, carry out the reaction with ethane containing one chiral methyl group. Although the reaction superficially resembles a P450-catalyzed hydroxylation, the enzyme contains iron in the active site as an oxo bridged diiron cluster, not as a heme iron. In addition, the complete system consists of two additional proteins, an oxidoreductase and another



a) Malonate decarboxylase from *M. rubra*, in D_2O. b) i) Acetate kinase, ATP.

ii) Phosphotransacetylase, CoASH. iii) Glyoxylic acid, malate synthase.

Fig. 7. Stereochemical analysis of the decarboxylation of chiral malonate to chiral acetate by **Malonomonas rubra** ***malonate decarboxylase.***

component of presently unknown function. These latter two proteins can be replaced by hydrogen peroxide as the source of oxygen and reducing equivalents.

Synthesis of the labeled substrates was carried out in collaboration with the National Tritium Labeling Facility in Berkeley by using carrier-free tritium. The radioactive ethane samples were then incubated with MMO and the resulting ethanol as well as the water of the reaction mixture recovered by lyophilization and analyzed by tritium NMR spectroscopy. The signal for the methylene protons consisted of two lines, one from tritium with ^1H as a neighbor and the other—deuterium upfield shifted—from tritium with deuterium as a neighbor. Integration of these two lines gives directly the primary kinetic deuterium isotope effect of the reaction (k_H/k_D 4.2 ± 0.2). The ethanol samples were then derivatized with optically active O-acetylmandelic acid and the tritium NMR spectra of the derivatives were examined. The two methylene hydrogens now have different chemical shifts and each again appears as a two-line pattern as a result of tritium with ^1H and tritium with ^2H as a neighbor. Integration of these four lines gave the relative abundance of the four tritiated species resulting from hydroxylation at the chiral methyl group and, thus, the stereochemistry of the reaction. The results (Priestley *et al.*, 1992) indicate that the reaction, both with the complete system and with the hydroxylase alone plus H_2O_2, proceeds with predominant (64 to 68%) retention of configuration, paralleling the stereochemistry of P450 catalyzed hydroxylations (Fig. 8). However, in every case, approximately one-third inversion was observed, which must have resulted from a reorientation of a reaction intermediate relative to the presumed iron oxo species delivering the hydroxyl oxygen. In principle, the reaction could proceed either with homolytic or heterolytic CH bond cleavage or by a concerted mechanism that avoids any free alkyl intermediates. The latter option can be excluded from these data because it would not explain the observed inversion. Cationic or anionic intermediates would very likely have an appreciable barrier to reorientation relative to the iron oxo species; the intermediate most likely to undergo such reorientation would be an ethyl radical that could undergo facile rotation around the carbon-carbon bond. Thus, we conclude that the MMO catalyzed hydroxylation most likely proceeds through an alkyl radical. If rotation around the carbon-carbon bond of such an ethyl radical is the only mechanism for "flipping" of the intermediate, then from the degree of inversion observed and from the known rate of this carbon-carbon bond rotation, one can estimate the rate of the subsequent rebound reaction. To rule out the possibility that other processes contribute to the observed inversion (for example, detachment and reattachment of the ethyl radical from its binding site), we repeated the reaction with butane carrying a chiral methyl group. This substrate should provide

more binding interaction with the enzyme and should result in less inversion if a process other than bond rotation contributes to the reorientation of the intermediate. The observed retention-to-inversion ratio was identical to that seen with ethane as a substrate. Hence, we believe that bond rotation is the only mechanism by which the intermediate radical reorients itself and, thus, this experiment provides a new radical clock for measuring the rate of the rebound reaction in such enzymatic

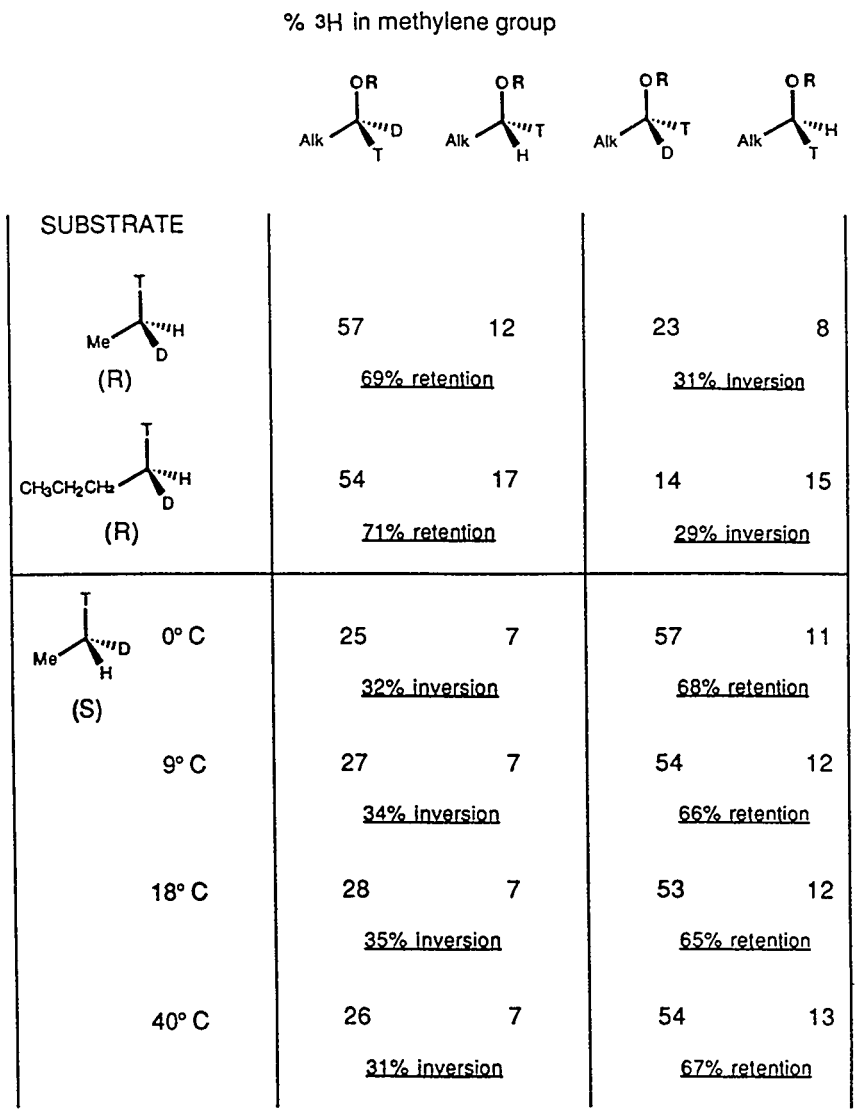


Fig. 8. Results of the stereochemical analysis of the methane monooxygenase reaction with the enzyme from Methylosinus trichosporium.

processes. Experiments are now underway to examine MMO from another source and possibly P450 enzymes with this stereochemical probe.

The examples above illustrate some of the many applications of stable and radioactive isotopes in probing biochemical mechanisms. Information generated in studies of this kind is useful in many fields, ranging from very fundamental understanding of enzyme mechanisms to biomedical and pharmaceutical applications in the design of new biologically active molecules. Most of this work would not have been possible without the help of the Los Alamos Stable Isotope Resource and the National Tritium Labeling Facility in Berkeley. These two federally supported facilities serve a vital role in furthering frontier research, and it is to be hoped that they will be able to continue to do so in the future.

Acknowledgements

The author wishes to acknowledge the contributions of a talented and enthusiastic group of coworkers as well as fruitful interactions with a number of collaborating laboratories. These are listed in the various references cited. In addition, financial support of this work by the National Institutes of Health (GM20264 and AI32333) is gratefully acknowledged, as are the services of the Los Alamos Stable Isotope Resource and of the National Tritium Labeling Facility, Berkeley.

References

Alworth, W. L. (1972) *Stereochemistry and Its Application in Biochemistry*, Wiley-Interscience, New York.

Arigoni, D. (1978) *Ciba Found. Symp. 60*, 243.

Bentley, R. (1969, 1970) *Molecular Asymmetry in Biology, Vols. 1 and 2*, Academic Press, New York.

Cho, H., Sattler, J., Beale, J.M., Zeeck, A., and Floss, H.G. (1993) *J. Org. Chem. 58*, 7925.

Cornforth, J.W., Redmond, J.W., Eggerer, H., Buckel, W., and Gutschow, C. (1970) *Eur. J. Biochem. 14*, 1.

Floss, H.G. and Lee, S. (1993) *Accts. Chem. Res. 26*, 116.

Floss, H.G. and Tasi, M.-D. (1979) *Adv. Enzymol. 50*, 243.

Floss, H.G., Tsai, M.-D, and Woodard, R.W. (1984) *Topics in Stereochem. 15*, 253.

Hanson, K.R. (1966) *J. Am. Chem. Soc. 88*, 2731.

Hanson, K.R. and Rose, I.A. (1975) *Accts. Chem. Res. 8*, 1.

Huang, S., Beale, J.M., Keller, P.J., and Floss, H.G. (1986) *J. Am. Chem. Soc. 108*, 1100.

Jordan, P.M. and Spencer, J.B. (1991) *Tetrahedron 47*, 6015.

Lee, J.J., Lee, J.P., Keller, P.J., Cottrell, C.E., Chang, C.-J., Zähner, H., and Floss, H.G. (1986) *J. Antibiot. (Tokyo) 39*, 1123.

Lüthy, J., Rétey, J., and Arigoni, D. (1969) *Nature (London) 221*, 1213.

Mocek, U., Zeng, Z., O'Hagan, D., Zhou, P., Fan, L.-D.G., Beale, J.M., and Floss, H.G. (1993) *J. Am. Chem. Soc. 115*, 7992.

Moore, B.S., Cho, H., Casati, R., Kennedy, E., Reynolds, K.A., Mocek, U., Beale, J.M., and Floss, H.G. (1993) *J. Am. Chem. Soc. 115*, 5254

Ogston, A.B. (1948) *Nature (London) 162*, 963.

Priestley, N.D., Floss, H.G., Froland, W.A., Lipscomb, J.D., Williams, P.G., and Morimoto, N. (1992) *J. Am. Chem. Soc. 114*, 7561.

Reese, P.B., Trimble, P.B., and Vederas, J.C. (1986) *Can. J. Chem. 64*, 1427.

Rétey, J. and Robinson, J.A. (1986) *Stereospecificity in Organic Chemistry and Enzymology*, Verlag Chemie, Weinheim, Germany.

Snipes, C.E., Chang, C.-J., and Floss, H.G. (1979) *J. Am. Chem. Soc. 101*, 701

Spencer, J.B. and Jordan, P.M. (1992) *J. Chem. Soc., Chem. Comm.* 646.

Tsai, M.-D. and Chang, T.T. (1980) *J. Am. Chem. Soc. 102*, 5416.

Webb, M.R. and Trentham, D.R. (1980) *J. Biol. Chem. 255*, 1775.

Wood, E.-R., Fujii, I., Ebizuka, Y., Sankawa, U., Kawaguchi, A., Huang, S., Beale, J.M., Shibuya, M., Mocek, U., and Floss, H.G. (1989) *J. Am. Chem. Soc. 111*, 5498.

Structures of Larger Proteins in Solution: Three- and Four-Dimensional Heteronuclear NMR Spectroscopy

Angela M. Gronenborn and G. Marius Clore

National Institute of Diabetes and Kidney Diseases
National Institutes of Health
Bethesda, MD 20892

Complete understanding of a protein's function and mechanism of action can only be achieved with a knowledge of its three-dimensional structure at atomic resolution. At present, there are two methods available for determining such structures. The first method, which has been established for many years, is x-ray diffraction of protein single crystals. The second method has blossomed only in the last 5 years and is based on the application of nuclear magnetic resonance (NMR) spectroscopy to proteins in solution. This review paper describes three- and four-dimensional NMR methods applied to protein structure determination and was adapted from Clore and Gronenborn (1991b). The review focuses on the underlying principals and practice of multidimensional NMR and the structural information obtained.

The advent of two-dimensional (2D) NMR methods (Jeener, unpublished; Aue *et al.*, 1976; Jeener *et al.*, 1979; Ernst *et al.*, 1987; Bax and Lerner, 1986) set the stage for determination of the first low-resolution structures of small proteins in the mid-1980s (Williamson *et al.*, 1985; Kaptein *et al.*, 1985; Clore *et al.*, 1985; 1986b). Subsequent improvements led to such an increase in the precision and accuracy of protein structure determinations that it is now possible, by using 2D NMR methods, to determine structures of proteins up to 100 residues that are comparable in quality to 2- to 2.5-Å-resolution x-ray structures (reviewed in Wüthrich, 1986, 1989, 1990; Clore and Gronenborn, 1987, 1989, 1991a; Kaptein *et al.*, 1988; Bax, 1989; Markley, 1989; Gronenborn and Clore, 1990). For proteins larger than 100 residues, however, heteronuclear 3D and 4D NMR with isotope labeling are necessary to overcome the limitations imposed by the spectral complexity inherent to larger proteins (Clore and Gronenborn, 1991b).

This review describes the applications of such methods, which culminated in 1991 in the determination of the high-resolution NMR structure of a protein greater than 150 residues (Clore *et al.*, 1991b).

Structural Data from NMR

The principal source of geometric information used in NMR protein structure determination lies in short, approximate interproton distance restraints (r) derived from nuclear Overhauser enhancement (NOE) measurements (Overhauser, 1953a,b; Solomon, 1955; Noggle and Schirmer, 1971). The physical basis for the NOE effect lies in the fact that each proton spin possesses a property known as magnetization. This magnetization is exchanged between the spins by a process termed *cross relaxation*, and the rate constant for this process is directly related to r^{-6}. Cross relaxation can be measured by perturbing the magnetization of a particular spin and observing the resulting change in magnetization (the NOE) of the other spins as the equilibrium is re-established. If the NOE is observed only a short time after the perturbation, the size of the NOE is proportional to the cross-relaxation rate—and, hence, to r^{-6}. Because of the r^{-6} dependence, the magnitude of these effects decreases rapidly as the interproton distances increase, so that NOEs are generally not observable beyond 5 Å. The interproton distance restraints derived from NOE measurements may also be supplemented by backbone and sidechain torsion-angle restraints derived from three-bond coupling constants and appropriate NOEs (Wagner *et al.*, 1987; Hyberts *et al.*, 1987; Arseniev *et al.*, 1988; Nilges *et al.*, 1990; Kraulis *et al.*, 1989; Güntert *et al.*, 1989).

With approximate interproton distance and torsion-angle restraints in hand, one can apply a number of computational strategies to locate the minimum of a target function that comprises terms for the experimental restraints, covalent geometry (that is, bonds, angles, planes, and chirality), and nonbonded contacts (such as a van der Waals repulsion term to prevent atoms from coming too close together). The types of algorithms used operate either in n-dimensional distance space, followed by projection into real space [such as metric matrix distance geometry (Crippen and Havel, 1978; Havel *et al.*, 1983; Havel and Wüthrich, 1985)] or directly in real space [such as minimization in torsion-angle space with a variable target function (Braun and Go, 1985), dynamical simulated annealing (Nilges *et al.*, 1988a,b,c), and restrained molecular dynamics (Kaptein *et al.*, 1988, Clore *et al.*, 1985, 1986a,b; Brünger *et al.*, 1986; Nilsson *et al.*, 1986; Gronenborn and Clore, 1989)]. All of the real-space methods require initial structures, which can be random structures with correct covalent geometry, structures very far from the final structure (such as extended strand), structures made up of a completely random array of atoms, or structures generated by distance-space methods. The key aspect of these methods

is that they have large radii of convergence and they fully sample, in an unbiased fashion, the conformational space consistent with the experimental, geometrical, and van de Waals restraints (Clore and Gronenborn, 1989). These various methods have been described in detail in a number of reviews (Braun, 1987; Crippen and Havel, 1988; Clore *et al.*, 1989b; Clore and Gronenborn, 1989, 1991b; Kuntz *et al.*, 1989; Scheek *et al.*, 1989) and have all been successfully applied to NMR structure determinations.

To assess the accuracy and precision of an NMR structure determination, it is essential to calculate a large number of structures independently by using the same experimental data set with different starting structures or conditions. The spread of structures that are consistent with the experimental data can be assessed *qualitatively* by visual inspection of the best fit superposition of a series of computed conformers and *quantitatively* by calculating the average atomic root-mean-square (rms) distribution of the individual structures about the mean coordinate positions. The quality and accuracy of NMR structures determined using heteronuclear 3D and 4D methods and evaluated by these parameters have dramatically improved as more experimental constraints have been applied in structure determinations. The most recent generation of NMR structures has been obtained using stereospecific assignments and loose torsion-angle restraints, derived by carrying out conformational grid searches of ϕ,ψ,χ_1 torsion-angle space on the basis of the $^3J_{NH,\alpha H}$ and $^3J_{\alpha H,\beta H}$ coupling constants and intraresidue and sequential interresidue NOEs involving the NH, αH, and βH protons (Wagner *et al.*, 1987; Hyberts *et al.*, 1987; Arseniev *et al.*, 1988; Kraulis *et al.*, 1989). This approach leads to 16–20 experimental restraints per residue, a backbone atomic rms distribution of ~0.4 Å, and an atomic rms distribution of ~0.5 Å for ordered sidechains. The errors in the atomic coordinates of these most recent generation structures, like that of interleukin-8 [a dimer with 72 residues per subunit (Clore *et al.*, 1989a)], are similar to those of 2- to 2.5-Å-resolution x-ray structures (Luzzatti, 1952; Cothia and Lesk, 1986; Kuryan *et al.*, 1986; Blundell *et al.*, 1987; Brünger, 1988, 1991a,b; Clore and Gronenborn, 1993). Indeed, the solution structure of interleukin-8 was used to solve the x-ray structure by molecular replacement because traditional methods based on heavy atom derivatives had proved unsuccessful despite several years of effort (Baldwin *et al.*, 1991). Other examples of these most recent generation structures include the COOH-terminal domain of cellobohydrolase (Kraulis *et al.*, 1989), the homeodomain of the *Antennapedia* protein (Qian *et al.*, 1989; Billeter *et al.*, 1990), a zinc finger domain from a human enhancer binding protein (Omichinski *et al.*, 1989), human thioredoxin (Forman-Kay *et al.*, 1991), and interleukin-1β (IL-1β) (Clore *et al.*, 1991b).

Deriving Experimental Restraints from NMR Data

Experimental restraints derived from NMR data require the identification of specific interactions between proton pairs, which may be either through-space (through the NOE) or through-bonds (through coupling constants). The power of the NMR method compared to other spectroscopic techniques lies in the fact that each proton gives rise to a specific resonance in the spectrum. Thus, a key aspect of any NMR structure determination is the requirement to assign each resonance to an individual proton and then to uniquely identify each pairwise through-space/NOE interaction. In principle, this can be accomplished in a relatively straightforward manner by using correlation experiments to identify resonances that belong to different amino acid types. The identification process employs through-bond connectivities and NOE experiments to subsequently link these residues in a sequential manner along the polypeptide chain on the basis of sequential and short-range interresidue NOEs that involve the NH, αH, and βH protons (Wüthrich, 1986, 1989, 1990; Clore and Gronenborn, 1987, 1989, 1991a; Gronenborn and Clore, 1990; Kaptein *et al.*, 1988; Bax, 1989; Markley, 1989). The types of NOE interactions that are most instructive for this purpose include NH^{i}-$NH^{i+1,2}$, αH^{i}-$NH^{i+1,2,3,4}$, βH^{i}-NH^{i+1}, and αH^{i}-CH^{i+3} connectivities. The pattern of observed NOEs provides a very good indication of the different secondary structure elements along the polypeptide chain. With the resonance assignments in hand, one can begin to identify long-range NOE interactions between protons that belong to residues far apart in the sequence but close together in space—a process that yields crucial information for determining the tertiary structure of the protein.

Although the principles of sequential resonance assignment are simple, the practice is difficult. Even for a small protein of ~50 residues, there are likely to be 300 to 400 protons for which resonances must be uniquely assigned. This number increases linearly with the number of residues, so for a 150-residue protein there will be 900 to 1200 protons. Because of the large number of protons, there is an extensive degree of resonance overlap and chemical shift degeneracy. As a result, 1D NOE and decoupling experiments can only be applied with any degree of confidence to peptides up to about 10 residues, and even then there may be serious difficulties.

The major conceptual advance in the application of NMR as a method for determining protein structure was the introduction of 2D NMR (Jeener, unpublished; Aue *et al.*, 1976; Jeener *et al.*, 1979; Ernst *et al.*, 1987; Bax and Lerner, 1986). Spreading out the correlations in two ^{1}H-frequency dimensions ensures that each interaction is labeled by two chemical shifts: the frequencies of the originating and destination protons. This approach not only results in a tremendous increase

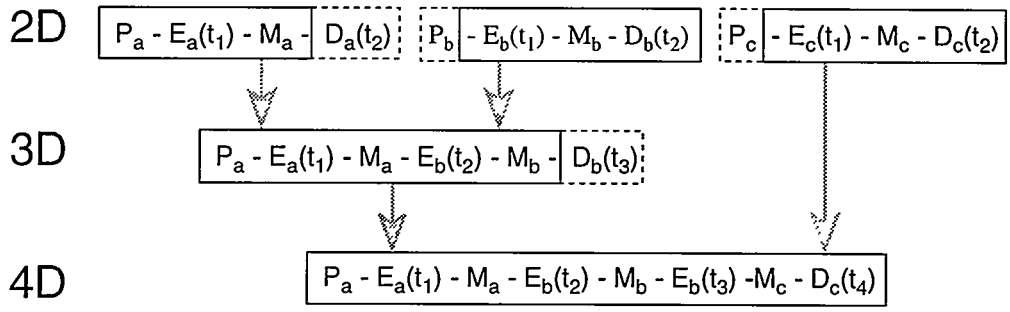


Fig. 1. This general representation of pulse sequences used in multidimensional NMR illustrates the relation between the basic schemes used to record 2D, 3D, and 4D NMR spectra. Note the way 3D and 4D experiments are constructed by the appropriate linear combination of 2D spectra (P = preparation; E = evolution; M = mixing; and D = detection). In 3D and 4D NMR, increments for the evolution periods are calculated independently.

in spectral resolution but, equally importantly, enables one to detect and interpret effects that would not have been possible in 1D.

All 2D experiments can be reduced to the basic conceptual scheme shown in Fig. 1 (Ernst *et al.*, 1987; Bax and Lerner, 1986). This scheme comprises a preparation pulse, an evolution period (t_1) during which the spins are labeled according to their chemical shifts, a mixing period (M_1) during which the spins are correlated with one another, and finally, a detection period (t_2). The experiment is repeated several times with successive linearly incremental values of the evolution period t_1 to yield a data matrix $s(t_1,t_2)$. Fourier transformation in the t_2 dimension yields a set of n 1D spectra, in which the intensities of the resonances are sinusoidally modulated as a function of t_1. Subsequent Fourier transformation in the t_1 dimension yields the desired 2D spectrum $S(\omega_1,\omega_2)$.

Such 2D methods have proved extremely powerful for determining the structure of small proteins. The largest proteins (in terms of number of residues) for which this approach has been successfully applied to achieve a complete 3D structure determination are thioredoxins from *E. coli* (Dyson *et al.*, 1990) and humans (Forman-Kay *et al.*, 1991), which have 108 and 105 residues, respectively. Beyond this limit of 100 residues, 2D methods soon break down because of (a) extensive spectral overlap caused by the larger number of resonances, and (b) the sharp decrease in efficiency with which magnetization can be transferred through the small, three-bond ^1H-^1H J couplings (3 to 12 Hz) because linewidths

become larger than the couplings with increasing rotational correlation time. This latter effect leads to incomplete delineation of spin systems (that is, amino acid types) in through-bond correlation experiments.

Solutions to both of these problems are obtained by extending the dimensionality of the NMR spectra to remove resonance overlap and degeneracy and by making use of through-bond correlations through heteronuclear couplings that are larger than the linewidths. This method necessitates using proteins that are uniformly labeled with ^{15}N, ^{13}C, or both. In proteins that can be over-expressed in bacterial systems, such labeling can be readily achieved by growing the organism in minimal medium, supplemented by $^{15}NH_4Cl$ or D-[$^{13}C_6$]glucose (or both) as the sole nitrogen and carbon sources, respectively.

Principals of 3D and 4D NMR

The design and implementation of higher dimensionality NMR experiments can be carried out by the appropriate combination of 2D NMR experiments (Fig. 1) (Griesinger *et al.*, 1987; Oschkinat *et al.*, 1988). A 3D experiment is constructed from two 2D pulse schemes by omitting the detection period of the first experiment and the preparation pulse of the second. This results in a pulse train comprising two independently incremented evolution periods t_1 and t_2, two corresponding mixing periods M_1 and M_2, and a detection period t_3. Similarly, a 4D experiment is obtained by combining three 2D experiments in an analogous fashion. Thus, conceptually n-dimensional NMR can be conceived as a straightforward extension of 2D NMR. However, the real challenge of 3D and 4D NMR is twofold: first, to ascertain which 2D experiments could be combined to achieve the best advantage; and second, to design the pulse sequences in such a way that undesired artifacts are removed.

The first application of 3D NMR to the small protein α_1-purothionin was presented in 1988 (Oschkinat *et al.*, 1988). The experiment was of the proton homonuclear variety in which a through-bond correlation experiment (HOHAHA; homonuclear Hartmann Hahn spectroscopy) was combined with a through-space experiment (NOESY; nuclear Overhauser enhancement spectroscopy). Although this experiment demonstrated the potential of the methodology, it suffered from a number of drawbacks that severely limited its application to larger proteins. First, the correlation portion of the experiment relied on small ^1H-^1H couplings. Second, all homonuclear 3D spectra are substantially more difficult to interpret than equivalent 2D versions because the number of cross peaks present in the former far exceeds that in the latter.

Fortunately, heteronuclear 3D and 4D NMR experiments do not suffer from any of these disadvantages and yield important additional information in the form of ^{15}N and ^{13}C chemical shifts.

They exploit a series of large one-bond heteronuclear couplings for magnetization transfer through-bonds, as is summarized in Fig. 2. This advantage, coupled with the fact that the ^1H nucleus is always detected, renders these experiments very sensitive. Indeed, high-quality 3D and 4D heteronuclear-edited spectra can easily be obtained on samples of 1- to 2-mM uniformly labeled protein in a time frame that is limited solely by the number of increments that must be collected for appropriate digitization and the number of phase cycling steps needed to reduce artifacts to an acceptably low level. Typical measurement times are 1.5–3 days for 3D experiments and 2.5–5 days for 4D experiments.

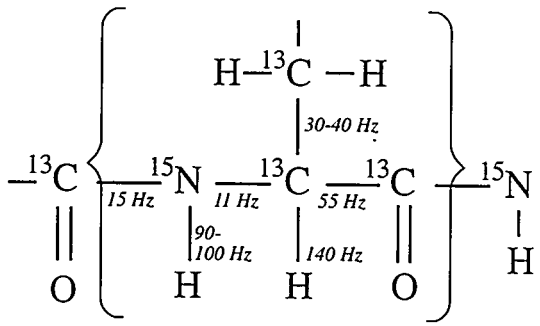


Fig. 2. Summary of the one-bond heteronuclear couplings along the polypeptide chain used in 3D and 4D NMR experiments. The backbone torsion angles ϕ and ψ involve rotations about the N^i-$C\alpha_i$, and $C\alpha$-C^i bonds, respectively; whereas the sidechain torsion angle χ_1 involves a rotation about the $C\alpha^i$- β^i bond.

Many of the 3D and 4D experiments are based on heteronuclear editing of ^1H-^1H experiments so that the general appearance of conventional 2D experiments is preserved and the total number of crosspeaks present is the same as that in the 2D equivalents. The progression from a 2D spectrum to 3D and 4D heteronuclear edited spectra is depicted schematically in Fig. 3. Consider, for example, the crosspeaks involving a particular ^1H frequency in a 2D NOESY spectrum, a 3D ^{15}N- or ^{13}C-edited NOESY spectrum, and finally a 4D ^{15}N- or ^{13}C-edited NOESY spectrum. In the 2D spectrum, a series of crosspeaks are seen from the originating proton frequencies in the F_1 dimension to the single destination ^1H frequency along the F_2 dimension. It is impossible to ascertain from the 2D experiment whether these NOEs involve only a single destination proton or several destination protons with identical chemical shifts. When the spectrum is spread into a third dimension, according to the chemical shift of the heteronucleus attached to the destination proton or protons, NOEs that involve different destination protons appear in distinct ^1H-^1H planes of the 3D spectrum. Thus, each interaction is simultaneously labeled by three chemical shift coordinates along three orthogonal axes of the spectrum. The projection of all of these planes onto a single plane yields the corresponding 2D spectrum. For the purposes of sequential assignment, heteronuclear-edited 3D spectra are often sufficient for analysis. However, when the goal of the analysis is to assign NOEs between protons far apart in the sequence, a 3D ^{15}N- or ^{13}C-edited NOESY

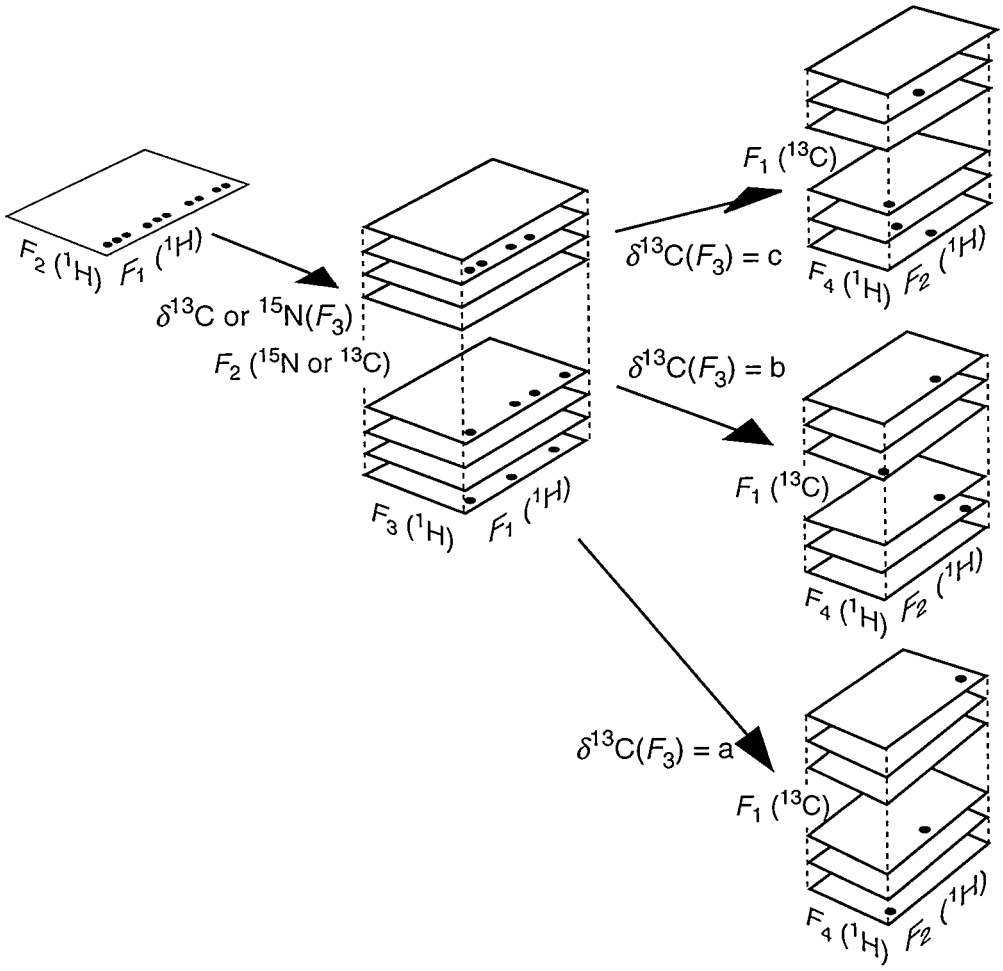


Fig. 3. Schematic of the progression and relation between 2D, 3D, and 4D heteronuclear NMR experiments (closed circles represent NOE cross peaks). In the example shown, 11 NOEs originate from 11 different protons in the F_1 dimension to a single frequency position in the F_2 dimension. In the 2D spectrum, it is impossible to ascertain whether there is only one destination proton or several in the F_2 dimension. By spreading the spectrum into a third dimension (labeled F_3) according to the chemical shift of the heteronucleus attached to the destination proton, it is possible to see that the NOEs lie in three distinct $^1H(F_1)$-$^1H(F_3)$ planes, indicating that three different destination protons are involved. However, the ^1H chemical shifts still provide the only means of identifying the originating protons. Hence, the problem of spectral overlap still prevents the unambiguous assignment of these NOEs. By extending the dimensionality of the spectrum to four, each NOE interaction is labeled by four chemical shifts along four orthogonal axes. Thus, the NOEs in each plane of the 3D spectrum are now spread over a cube in the 4D spectrum according to the chemical shift of the heteronucleus directly attached to the originating protons.

spectrum often proves inadequate because the originating protons are only specified by their ^{1}H chemical shifts and—more often than not—there are several protons that resonate at the same frequencies. For example, in the case of the 153-residue protein IL-1β, there are 60 protons that resonate in a 0.4-ppm interval between 0.8 and 1.2 ppm (Clore *et al.*, 1990a). Such ambiguities can then be resolved by spreading out the 3D spectrum still further into a fourth dimension, according to the chemical shift of the heteronucleus attached to the originating protons. In this manner, each NOE interaction is simultaneously labeled by four chemical shift coordinates along four orthogonal axes (those of the originating and destination protons and those of the corresponding heteronuclei directly bonded to these protons) (Kay *et al.*, 1990; Clore *et al.*, 1991a; Zuiderweg *et al.*, 1991). The result is a 4D spectrum in which each plane of the 3D spectrum constitutes a cube in the 4D spectrum.

For illustrative purposes, it is also useful to compare the type of information that can be extracted from a very simple system with the use of 2D, 3D, and 4D NMR. Consider a molecule with only two NH and two aliphatic protons, in which only one NH proton is close to an aliphatic proton. In addition, the chemical shifts of the NH protons are degenerate, as are those of the aliphatic protons, so that only two resonances are seen in the lD spectrum. In the 2D NOESY spectrum, an NOE would be observed between the resonance position of the NH protons and the resonance position of the aliphatic protons, but it would be impossible to ascertain which one of the four possible NH-aliphatic proton combinations gives rise to the NOE. By spreading the spectrum into a third dimension (for example, using the chemical shift of the ^{15}N atoms attached to the NH protons), the number of possibilities would be reduced to two—provided, of course, that the chemical shifts of the two nitrogen atoms are different. Finally, when the fourth dimension corresponding to the chemical shift of the ^{13}C atoms attached to the aliphatic protons is introduced, it is possible to make an unique assignment of the NH-aliphatic proton pair that gives rise to the NOE.

Applying 3D and 4D NMR to Larger Proteins

The test of multidimensional NMR is its success in solving the problem it was originally designed to tackle: the determination of high-resolution 3D structures of larger proteins in solution. This goal was attained in 1991 with the solution of the structure of IL-1β (17.4 kDa), which plays a central role in immune and inflammatory responses (Clore *et al.*, 1991b). At the time, this protein was 50% larger than any other protein whose 3D structure had been determined by NMR.

Despite extensive analysis of 2D spectra obtained at different pH values and temperatures, as well as examination of 2D spectra of mutant proteins, it was not feasible to obtain unambiguous

^{1}H assignment for more than 30% of the residues in IL-1β (Driscoll *et al.*, 1990a,b). Further progress could only be made by resorting to higher dimensional heteronuclear NMR. Figure 4 provides a summary of the strategy we used to determine the structure of this larger protein. The initial step involved the complete assignment of the ^{1}H, ^{15}N, and ^{13}C resonances of the backbone and sidechains by using the entire gamut of double- and triple-resonance 3D experiments listed in the top left-hand panel of Fig. 4 (Driscoll *et al.*, 1990a,b; Clore *et al.*, 1990a). In the second step, backbone and sidechain torsion-angle restraints, as well as stereospecific assignments for 1β-methylene protons, were obtained by means of a 3D grid search of ϕ,ψ,χ_1 space in which the experimental data for the $^3J_{NH,\alpha H}$ and

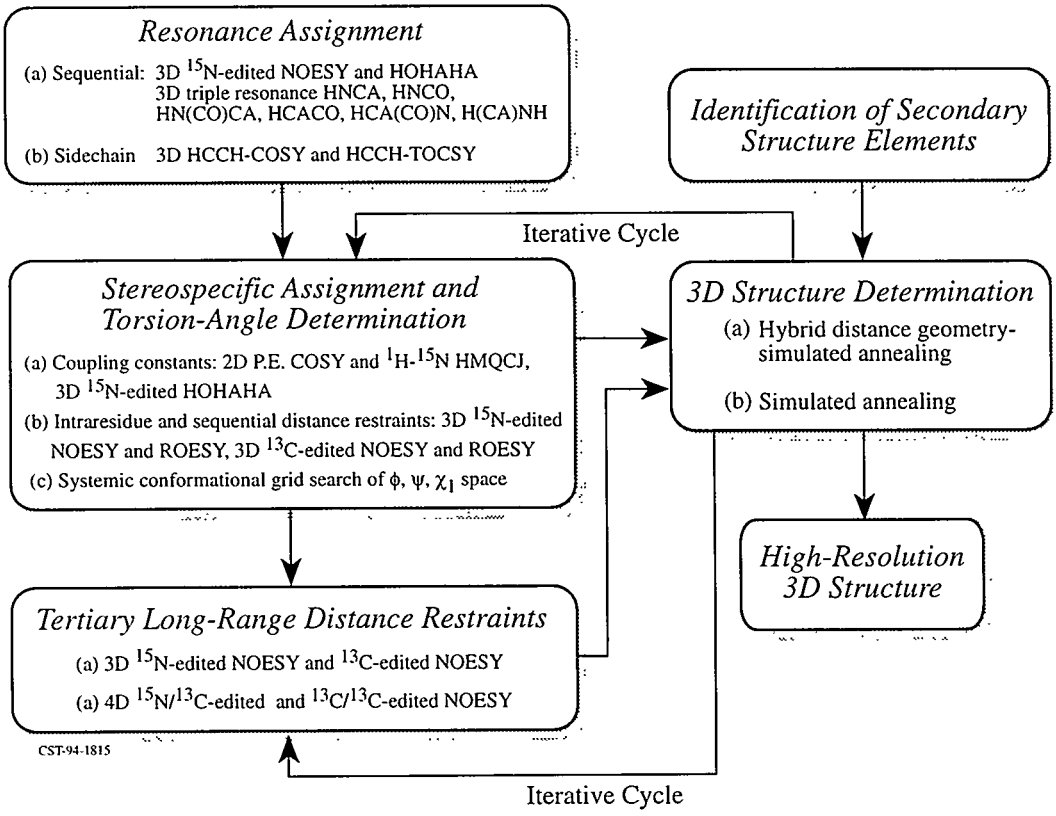


Fig. 4. Outline of the general strategy used in our laboratory to determine the 3D structure of larger proteins such as IL-1β by 3D and 4D NMR.

$^3J_{\alpha H,\beta H}$ coupling constants and the approximate intraresidue and sequential distance restraints between the NH, αH, and βH protons were compared to values present in an extensive database (Nilges *et al.*, 1990; Kraulis *et al.*, 1989). In the third step, approximate interproton distance restraints between nonadjacent residues were derived from analysis of 3D and 4D heteronuclear-edited NOESY spectra. Analysis of the 3D heteronuclear-edited NOESY spectra alone was sufficient to derive a low-resolution structure on the basis of a small number of NOEs involving solely NH, αH, and βH protons (Clore *et al.*, 1990c). However, further progress using 3D NMR was severely hindered by the numerous ambiguities still present in these spectra—in particular, for NOEs arising from the large number of aliphatic protons. Thus, the 4D heteronuclear-edited NOESY spectra proved to be absolutely essential for successful completion of this task. In addition, the proximity of backbone NH protons to bound structural water molecules was ascertained from a 3D ^{15}N-separated ROESY (rotating frame Overhauser spectroscopy) spectrum that makes it possible to distinguish specific protein-water/NOE interactions from chemical exchange with bulk solvent (Clore *et al.*, 1990b). We should emphasize again that, in our laboratory, all NOE data are interpreted in as conservative a manner as possible and are simply classified into three distance ranges: 1.8–2.7, 1.8–3.3, and 1.8–5.0 Å, which correspond to strong- , medium- , and weak-intensity NOEs, respectively.

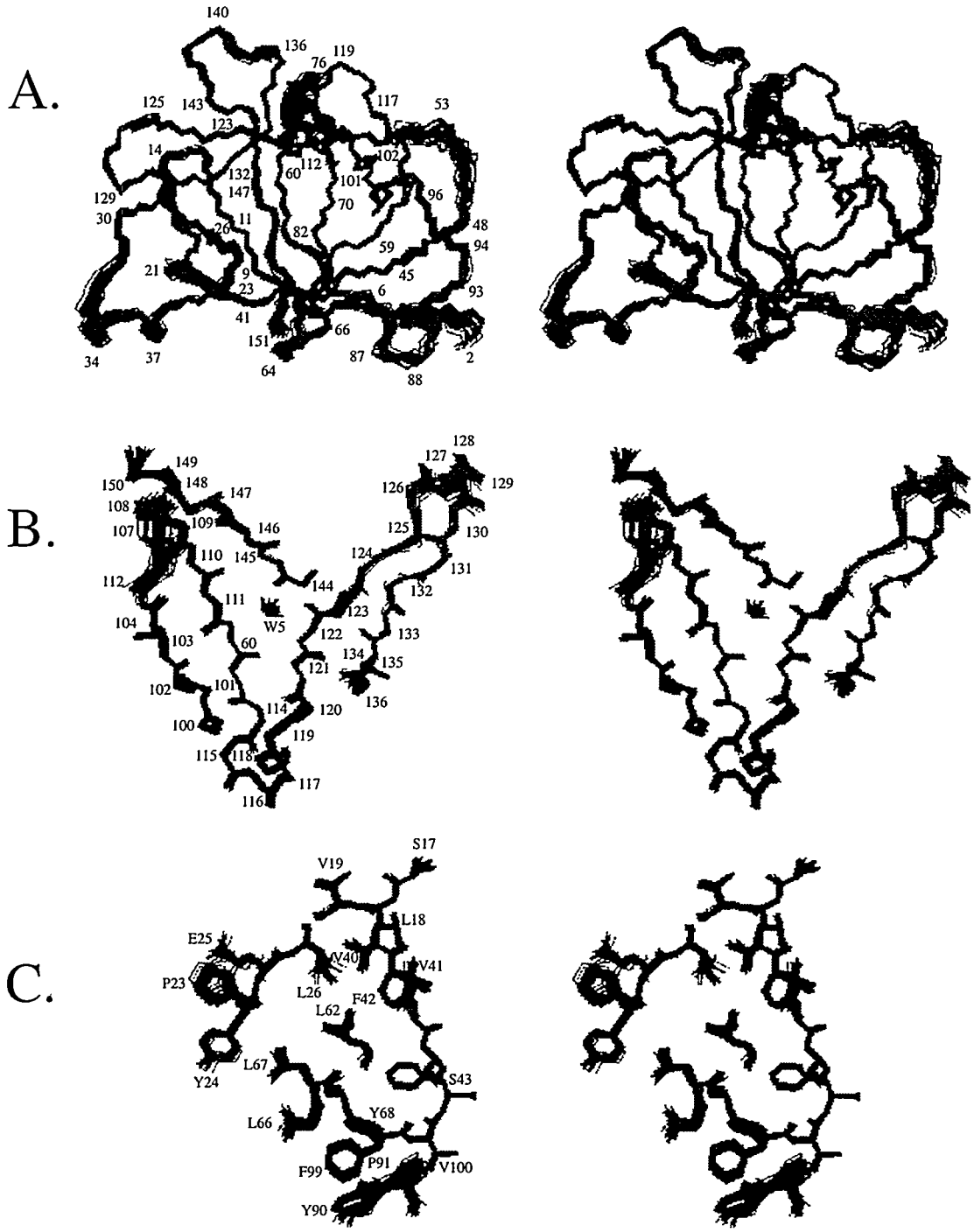
With an initial set of experimental restraints in hand, we initiated 3D structure calculations. Typically, we use the method of hybrid distance geometry-dynamical simulated annealing, in which an approximate polypeptide fold is obtained by projecting a subset of atoms from n-dimensional distance space into cartesian coordinate space; this is followed by simulated annealing that includes all of the atoms (Nilges *et al.*, 1988a). Alternatively, we use simulated annealing from either random structures with intact covalent geometry or from a completely random array of atoms (Nilges *et al.*, 1988b,c). All of these simulated annealing protocols involve solving Newton's equations of motion subject to a simplified target function comprising terms for the experimental restraints, covalent geometry, and nonbonded contacts. The underlying principle consists of raising the temperature of the system and then slowly cooling it to (a) overcome false local minima and large potential energy barriers along the path toward the global minimum region of the target function, and (b) sample efficiently and comprehensively the conformational space consistent with the experimental restraints. A key aspect of the overall strategy lies in using an iterative approach, whereby experimental data is re-examined in the light of the initial set of calculated structures. This method is employed to resolve ambiguities in NOE assignments, to obtain more torsion-angle restraints and stereospecific assignments (such as the α-methylene protons of Gly

and the methyl groups of Val and Leu), and to assign backbone hydrogen bonds that are associated with slowly exchanging NH protons as well as with bound water molecules. The iterative cycle comes to an end when all of the experimental data have been interpreted.

The final experimental data set for IL-1β comprised a total of 3146 approximate and loose experimental restraints made up of 2780 distance and 366 torsion-angle restraints (Clore *et al.*, 1991b). These data represent an average of 21 experimental restraints per residue. If one takes into account that interresidue NOEs affect two residues, whereas intraresidue NOE and torsion-angle restraints only affect individual residues, the average number of restraints influencing the conformation of each residue is 33. Figure 5 shows a superposition of 32 independently calculated structures. All 32 structures satisfy the experimental restraints within their specified errors, display very small deviations from idealized covalent geometry, and have good nonbonded contacts. Both the backbone and the ordered sidechains are exceptionally well defined. Indeed, the atomic rms distribution about the mean coordinate positions is 0.4 Å for the backbone atoms, 0.8 Å for all atoms, and 0.5 Å for side chains with ~40% of their surface (relative to that in a tripeptide Gly-X-Gly) accessible to solvent (Clore *et al.*, 1991b).

The structure of IL-1β itself resembles a tetrahedron and displays threefold internal pseudosymmetry. There are 12 β–strands arranged in an exclusively antiparallel structure, and 6 of the strands form a barrel (see Fig. 5A) that is closed off at the back of the molecule by the other 6 strands. Each repeating topological unit is composed of 5 strands arranged in an antiparallel manner with respect to each other (see Fig. 5B). Water molecules occupy similar positions in all three topological units, as well as at the interface of the three units, and are involved in bridging backbone hydrogen bonds. Thus, in the case of the topological unit shown in Fig. 5B, the water molecule labeled W5 accepts a hydrogen bond from the NH of Phe^{112} in strand IX and donates two hydrogen bonds to the backbone carbonyls of Ile^{122} in strand X and Thr^{144} in strand XII. The packing of some internal residues, as well as the excellent definition of internal sidechains, is illustrated in Fig. 5C. Because of the high resolution

Fig. 5. Stereoviews of the 3D structure of IL-1β determined by 3D and 4D heteronuclear NMR spectroscopy on the basis of a total of 3146 approximate and loose experimental NMR restraints (Clore* et al., *1991b). Best fit superpositions of the backbone (N, Cα, and C) atoms of residues 2 to 151, the backbone (N, Cα, C, and O) atoms of one of the three repeating topological units including a water molecule (W5), and selected sidechains of 32 simulated annealing structures are shown in (A), (B), and (C), respectively. The NH_2-terminal residue and the two COOH-terminal ones (residues 152 and 153) are ill-defined.


A.
140
136
76
119
125
143
123
117
53
14
102
112
60
132
101
147
70
96
129
30
11
26
82
48
59
94
21
9
45
23
6
93
41
66
151
87
34
37
64
2
88
B.
149
128
127
150
148
147
126
129
108
109
146
125
107
130
110
145
124
131
112
144
132
111
123
W5
104
122
133
103
60
134
135
121
102
101
136
114
120
100
119
115
118
117
116
C.
S17
V19
L18
E25
V40
V41
P23
L26
F42
L62
L67
S43
Y24
Y68
L66
P91
F99
V100
Y90

of the IL-1β structure, it was possible to analyze in detail the sidechain/sidechain interactions that stabilize the structure. In addition, our examination of the structure in the light of mutational data permitted us to propose the presence of three distinct sites involved in the binding of IL-1β to its cell surface receptor (Clore *et al.*, 1991b).

Conclusions

This review summarizes recent developments in heteronuclear 3D and 4D NMR that have been designed to extend NMR methodology to medium-size proteins in the 15- to 30-kDa range. The underlying principle of this approach relies on extending the dimensionality of the spectra to obtain dramatic improvements in spectral resolution and exploiting large heteronuclear couplings to circumvent problems associated with larger linewidths. A key feature of all of these experiments is that they do not result in any increase in the number of observed crosspeaks relative to their 2D counterparts. Hence, the improvement in resolution is achieved without raising the spectral complexity, which renders data interpretation straightforward. Thus, for example, in 4D heteronuclear-edited NOESY spectra, the NOE interactions between proton pairs are not only labeled by the ^{1}H chemical shifts but also by the corresponding chemical shifts of their directly bonded heteronuclei in four orthogonal axes of the spectrum. For practical applications, the high sensitivity of these experiments makes it feasible to obtain high-quality spectra in a relatively short timeframe on 1- to 2-mM protein samples uniformly labeled with ^{15}N or ^{13}C or both.

Just as 2D NMR expanded applications of NMR to structure determination of small proteins of less than 100 residues, 3D and 4D heteronuclear NMR provide the means for extending the methodology to medium-size proteins in the 150- to 300-residue range. The determination of the high-resolution structure of IL-1β using 3D and 4D heteronuclear NMR (Clore *et al.*, 1991b) demonstrates that the technology is available for determining the structures of such medium-sized proteins at a level of accuracy and precision that is comparable to the best results attainable for small proteins.

Acknowledgements

This work was supported by the AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institute of Health.

References

Arseniev, A., Schultze, P., E., Braun, W., Wagner, G., Vasák, M., Kägi, J.H.R., and Wüthrich, K. (1988) *J. Mol. Biol. 201,* 637.

Aue, W.P., Bartholdi, E., and Ernst, R.R., (1976) *J. Chem. Phys. 64,* 2229.

Baldwin, E.T., Weber, I.T., St. Charles, R., Xuan, J.-C., Apella, E., Yamada, M., Matsushima, K., Edwards, B.F.P., Clore, G.M., Gronenborn, A.M., and Wlodower, A. (1991) *Proc. Natl. Acad. Sci. USA 88,* 502.

Bax, A. (1989) *Annu. Rev. Biochem. 58,* 223.

Bax, A., Clore, G.M., and Gronenborn, A.M. (1990) *J. Magn. Reson. 88,* 425.

Bax, A. and Lerner, L.E. (1986) *Science 232,* 960.

Billeter, M. (1990) *J. Mol. Biol. 214,* 183.

Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E., and Thornton, J.M. (1987) *Nature 326,* 347.

Braun, W. (1987) *Q. Rev. Biophys. 19,* 115.

Braun, W. and Go, N., (1985) *J. Mol. Biol. 186,* 611.

Brünger, A.T. (1988) *J. Mol. Biol. 203,* 803.

Brünger, A.T. (1991a) in *Molecular Dynamics: Applications to Molecular Biology,* J. M. Goodfellow, ed., Macmillan, London, pp. 137.

Brünger, A.T. (1991b) *Annu. Rev. Phys. Chem. 42,* 197.

Brünger, A.T., Clore, G.M., Gronenborn, A.M., and Karplus, M. (1986) *Proc. Natl. Acad. Sci. USA 83,* 3801.

Chothia, C. and Lesk, A. (1986) *EMBO J. 5,* 823.

Clore, G.M. and Gronenborn, A.M., (1987) *Protein Eng. 1,* 275.

Clore, G.M. and Gronenborn, A.M., (1989) *CRC Crit. Rev. Biochem. Mol. Biol. 24,* 479.

Clore, G.M. and Gronenborn, A.M., (1991a) *Ann. Rev. Biophys. Biophys. Chem. 21,* 29.

Clore, G.M. and Gronenborn, A.M. (1991b) *Science 252,* 1390.

Clore, G.M. and Gronenborn, A.M. (1993) *J. Mol. Biol. 233,* 331.

Clore, G.M., Gronenborn, A.M., Brünger, A.T., and Karplus, M. (1985) *J. Mol. Biol. 186,* 435.

Clore, G.M., Brünger, A.T., Karplus, M., and Gronenborn, A.M. (1986a) *J. Mol. Biol. 191,* 523.

Clore, G.M., Nilges, M., Sukumaran, D.K., Brünger, A.T., Karplus, M., and Gronenborn, A.M. (1986b) *EMBO J. 5,* 2729.

Clore, G.M., Appella, E., Yamada, M., Matsushima, K., and Gronenborn, A.M. (1989a) *Biochemistry 29,* 1689.

Clore, G.M., Nilges, M., and Gronenborn, A.M. (1989b) in *Computer-Aided Molecular Design,* W. G. Richards, ed., IBC Technical Services, London, pp. 203-219.

Clore, G.M., Bax, A., Driscoll, P.C., Wingfield, P.T., and Gronenborn, A.M. (1990a) *Biochemistry 29,* 8172.

Clore, G.M., Bax, A., Wingfield, P.T., and Gronenborn, A.M. (1990b) *Biochemistry 29,* 5671.

Clore, G.M., Driscoll, P.C. Wingfield, P.T., and Gronenborn, A.M. (1990c) *J. Mol. Biol. 214*, 811.

Clore, G.M., Bax, A., and Gronenborn, A.M. (1991) *J. Biomol. NMR 1(1)*, 13.

Clore, G.M., Kay, L.E., Bax, A., and Gronenborn, A.M. (1991a) *Biochemistry 30*, 12.

Clore, G.M., Wingfield, P.T., and Gronenborn, A.M. (1991b) *Biochemistry 30*, 2315.

Crippen, G.M. and Havel, T.F. (1978) *Acta Crystallogr. A34*, 282.

Crippen, G.M. and Havel, T.F. (1988) Distance Geometry and Molecular Conformation, Wiley, New York.

Driscoll, P.C., Clore, G.M., Marion, D., Wingfield, P.T., and Gronenborn, A.M. (1990) *Biochemistry 29*, 3542.

Driscoll, P.C., Gronenborn, A.M., Wingfield, P.T., and Clore, G.M. (1990b) *Biochemistry 29*, 4468.

Dyson, H.J., Gippert, G.P., Case, D.A., Holmgren, A., and Wright, P.E. (1990) *Biochemistry 29*, 4129.

Ernst, R.R., Bodenhausen, G., and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon, Oxford.

Forman-Kay, J.D., Clore, G.M., Wingfield, P.T., and Gronenborn, A.B. (1991) *Biochemistry 30*, 2685.

Griesinger, C., Sorensen, O.W., and Ernst, R.R. (1987) *J. Magn. Reson. 73*, 574.

Gronenborn, A.M. and Clore, G.M. (1989) *Biochemistry 28*, 5978.

Gronenborn, A.M. and Clore, G.M. (1990) *Anal. Chem. 62*, 2.

Güntert, P., Braun, W., Wider, W., and Wüthrich, K. (1989) *J. Am. Chem. Soc. 111*, 3997.

Havel, T.F., Kuntz, I.D., and Crippen, G.M. (1983) *Bull. Math. Biol. 45*, 665.

Havel, T.F. and Wüthrich, K. (1985) *J. Mol. Biol. 182*, 281.

Hyberts, S., Marki, W., and Wagner, G. (1987) *Eur. J. Biochem. 164*, 625.

Jeener, J., Meier, B.H., Bachmann, P., and Ernst, R.R. (1979) *J. Chem. Phys. 71*, 4540. (1979).

Kaptein, R., Boelens, R., Scheek, R.M., and van Gunsteren, W.F. (1988) *Biochemistry 27*, 5389;

Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R., and van Gunsteren, W.F. (1985) *J. Mol. Biol. 182*, 179.

Kay, L.E., Clore, G.M., Bax, A., and Gronenborn, A.M. (1990) *Science 249*, 411.

Kraulis, P.J., Clore, G.M., Nilges, M., Jones, T.A., Pettersson, G., Knowles, J., and Gronenborn, A.M. (1989) *Biochemistry 28*, 7241.

Kuntz, D., Thomason, J.F., and Oshiro, C.M. (1989) *Methods Enzymol. 177*, 159.

Kuryan, J., Petzko, G.A., Levy, R., and Karplus, M. (1986) *J. Mol. Biol. 190*, 227.

Luzzati, V. (1952) *Acta Crystallogr. 5*, 802.

Markley, J. (1989) *Methods Enzymol. 176*, 12.

Nilges, M., Clore, G.M., and Gronenborn, A.M. (1988a) *FEBS Lett. 239*, 129.

Nilges, M., Clore, G.M., and Gronenborn, A.M. (1990) *Biopolymers 29*, 813.

Nilges, M., Gronenborn, A.M., Brünger, A.T., and Clore, G.M. (1988b) *Protein Eng. 2*, 27.

Nilges, M., Clore, G.M., and Gronenborn, A.M. (1988c) *FEBS Lett. 229*, 317.

Nilges, M., Clore, G.M., and Gronenborn, A.M., Brünger, A. T., and Karplus, M. (1986) *J. Mol. Biol. 188*, 455.

Noggle, J.H. and Schirmer, R.E. (1971) *The Nulear Overhauser Effect—Chemical Applications*, Academic Press, New York.

Omichinski, J.G., Clore, G.M., Appella, E., Sakaguchi, K., and Gronenborn, A.M. (1989) *Biochemistry 29*, 9324.

Oschkinat, H., Griesinger, C., Kraulis, P.J., Sorensen, D.W., Ernst, R.R., Gronenborn, A.M., and Clore, G.M. (1988) *Nature 332*, 374.

Overhauser, A. (1953a) *Phys. Rev. 89*, 689.

Overhauser, A. (1953b) *Phys. Rev. 92*, 411.

Qian, Y.Q., Billeter, M., Otting, G., Müller, M., Gehring, W.J., and Wüthrich, K. (1989) *Cell 59*, 573

Scheek, R.M., van Gunsteren, W.F., and Kaptein, R. (1989) *Methods Enzymol. 177*, 204.

Solomon, I. (1955) *Phys. Rev. 99*, 559.

Wagner, G., Braun, W., Havel, T.I., Schaumann, T., Go, N., and Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611.

Williamson, M.P., Havel, T.F., and Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295.

Wüthrich, K. (1986) *NMR of Proteins*, Wiley, New York.

Wüthrich, K. (1989) *Science 243*, 45.

Wüthrich, K. (1990) *J. Biol. Chem. 265*, 22059.

Zuiderweg, E.R.P., Petros, A.M., Fesik, S.W., and Olejniczak, E.T. (1991) *J. Am. Chem. Soc. 113*, 370.

SYNTHESIS AND NMR OF ^{15}N-LABELED DNA FRAGMENTS

ROGER A. JONES

Department of Chemistry
Rutgers, The State University of New Jersey
Piscataway, NJ 08855

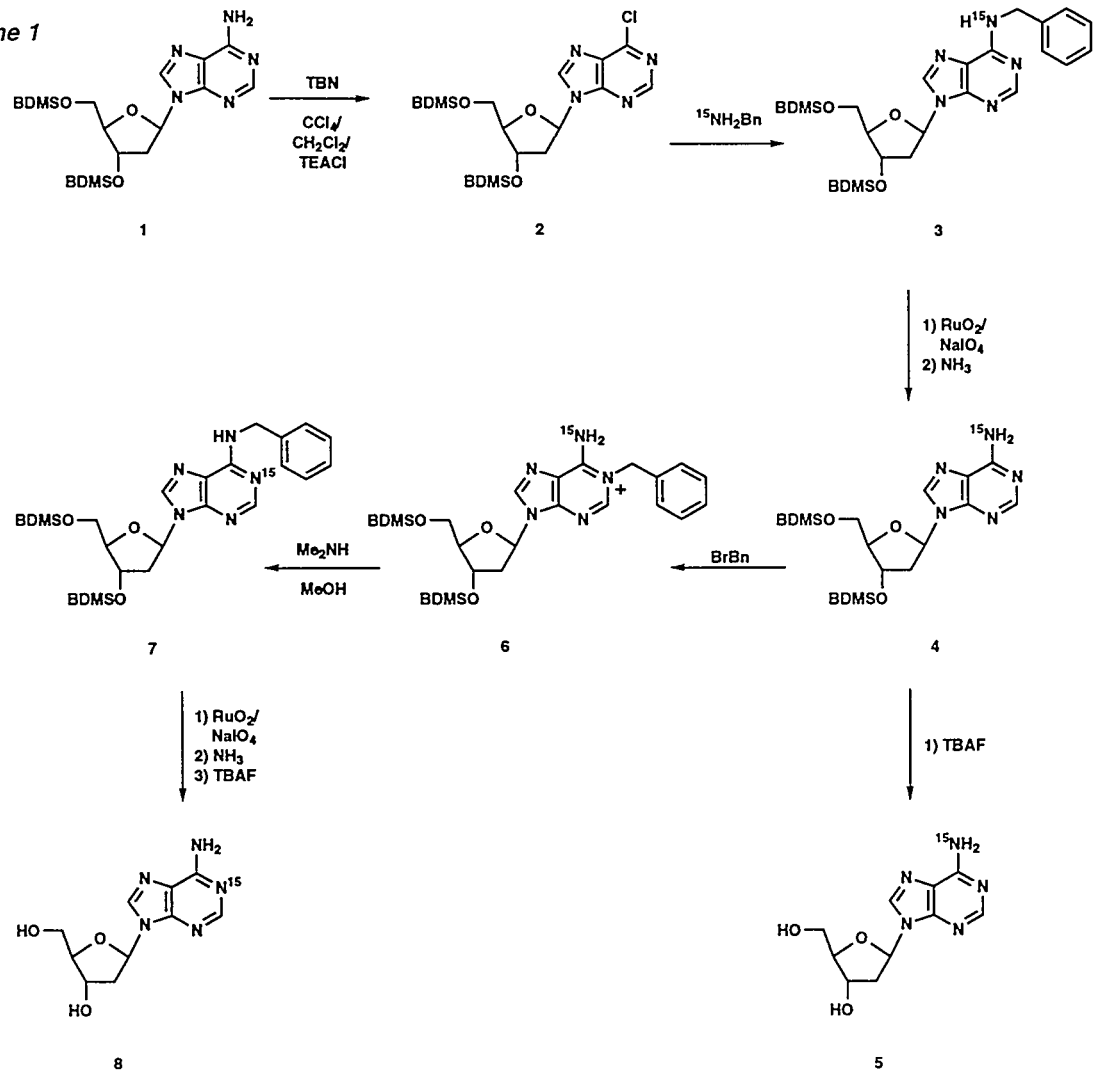
DNA fragments labeled with ^{15}N at the ring nitrogens and at the exocyclic amino groups can be used to obtain novel insight into interactions such as base pairing, hydration, drug binding, and protein binding. A number of synthetic routes to ^{15}N-labeled pyrimidine nucleosides (DeGraw and Lawson, 1978; Poulter and Livingston, 1979; Niu, 1984), purines (Leonard and Henderson, 1975; Bario *et al.*, 1981; Sethi *et al.*, 1982), and purine nucleosides (Golding *et al.*, 1986; Gao and Jones, 1987; Kupferschmitt *et al.*, 1987; Massefski *et al.*, 1990; Gaffney *et al.*, 1990; Rhee and Jones, 1990; Goswami and Jones, 1991) have been reported. Moreover, many of these labeled bases or monomers have been incorporated into nucleic acids, either by chemical synthesis (Kupferschmitt *et al.*, 1987; Massefski *et al.*, 1990; Gao and Jones, 1987b; Wang *et al.*, 1991a, 1991b; Gaffney *et al.*, 1992) or by biosynthetic procedures (Griffey *et al.*, 1982; Roy *et al.*, 1984; Gewirth *et al.*, 1987; Davis *et al.*, 1989; Davis and Poulter, 1991). The focus of this chapter will be on the preparation of ^{15}N-labeled purine 2′-deoxynucleosides, their incorporation into DNA fragments by chemical synthesis, and the results of NMR studies using these labeled DNA fragments.

Synthesis of ^{15}N-Labeled 2′-Deoxyadenosine and 2′-Deoxyguanosine

2′-Deoxy-[1-^{15}N]- and 2′-Deoxy-[amino-^{15}N]adenosine

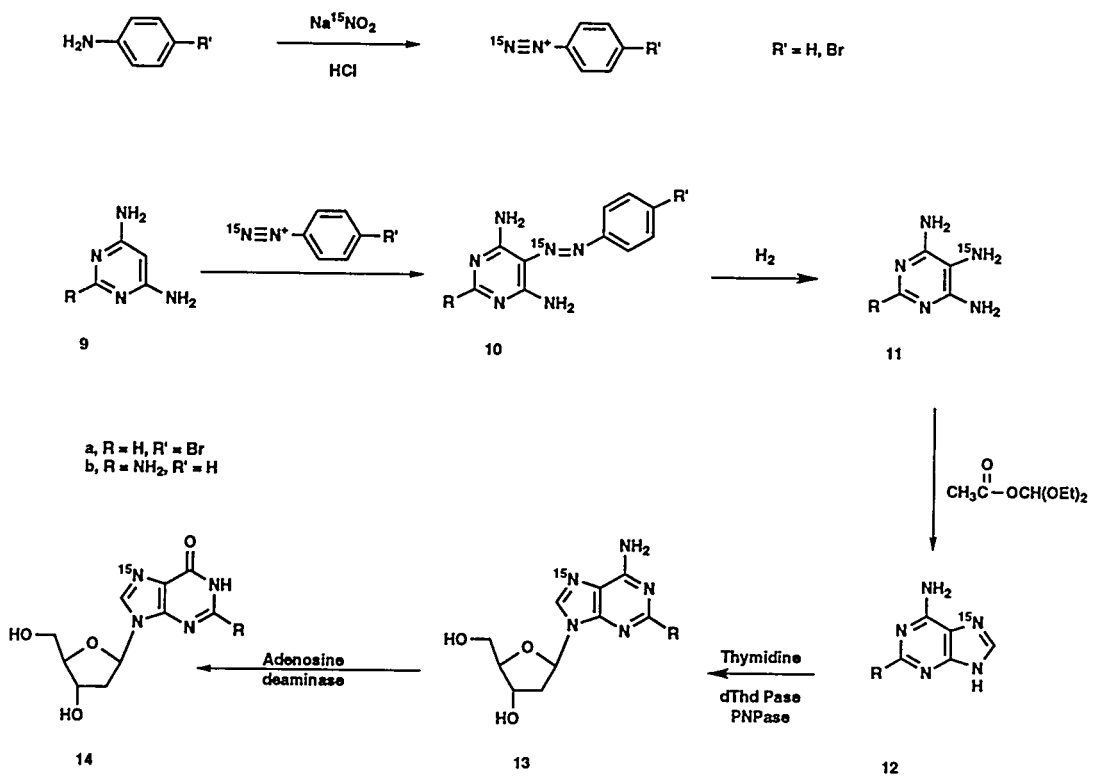
The transformation of 2′-deoxyadenosine to the 2′-deoxy-[*amino*-^{15}N]adenosine derivative (**4**) and from **4** to 2′-deoxy-[1-^{15}N]adenosine (**8**) is shown in Scheme 1 (Gao and Jones, 1987a). Starting with the 3′- and 5′-**tert**-butyldimethylsilyl (BDMS) derivative (**1**), a nonaqueous diazotization reaction gives the corresponding 6-chloro derivative (**2**). Reaction of **2** with [^{15}N]-benzylamine then gives the [6-^{15}N]benzyl derivative (**3**). Although all attempts at

Scheme 1



reductive debenzylation of 3 proved unsatisfactory, oxidative conversion to the corresponding benzoyl derivative proceeds readily. Ammonolysis then gives 4, which can be desilylated using tetra-n-butylammonium fluoride (TBAF) in THF to yield 2′-deoxy-[*amino*-^{15}N]-adenosine (5). Alternatively, alkylation of 4 with benzyl bromide ((Robins and Trip, 1973), followed by treatment with methanolic dimethylamine effects Dimroth rearrangement to give the [1-^{15}N]derivative (7). Debenzylation and desilylation of 7, carried out exactly as for the conversion of 3 to 5, then gives 2′-deoxy-[1-^{15}N]adenosine (8).

Scheme 2



2'-Deoxy-[7-^{15}N]adenosine and 2'-Deoxy-[7-^{15}N]guanosine

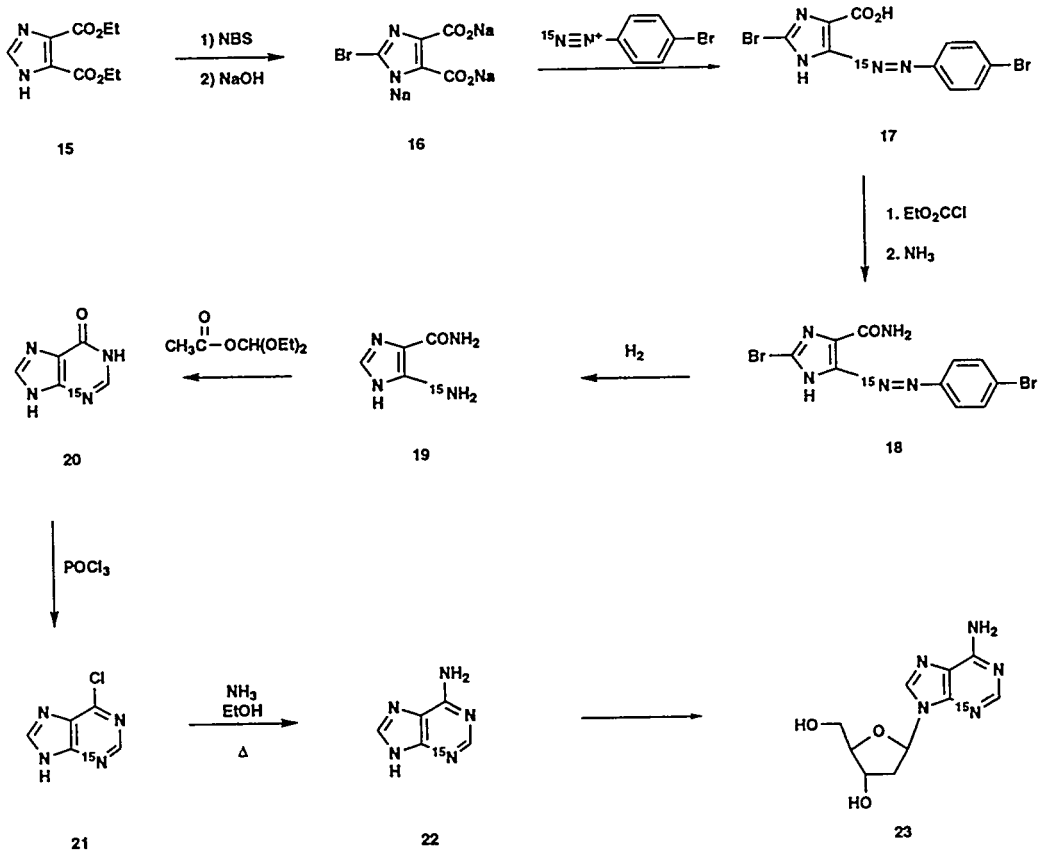
Labeling the 7 position of deoxyadenosine and deoxyguanosine was carried out by construction of the [7-^{15}N]heterocycle followed by an enzymatic transglycosylation reaction, as shown in Scheme 2 (Gaffney *et al.*, 1990). This procedure uses ^{15}N-sodium nitrite as an inexpensive ^{15}N source. Diazotization of aniline or 4-bromoaniline using $Na^{15}NO_2$ gives the corresponding [β-^{15}N]diazonium ion, which is then used immediately in an azo coupling reaction with 4,6-diaminopyrimidine (**9a**) or 2,4,6-triaminopyrimidine (**9b**). Rearrangement of the ^{15}N to the α position does not occur to a detectable extent under these conditions. The crystalline azo derivatives (**10a/b**) are obtained in high yield, without need of further purification. Hydrogenolysis of the azo linkage then gives the corresponding ^{15}N-labeled tri- or tetraaminopyrimidines (**11a/b**). These aminopyrimidines will discolor readily, and are best immediately converted to [7-^{15}N]adenine (**12a**) or 2,6-diamino-[7-^{15}N]purine (**12b**) by treatment with diethoxy-methyl

acetate. The corresponding 2′-deoxynucleosides are then obtained by an enzymatic transglycosylation procedure in which thymidine is used as the glycosyl donor. The thymidine is cleaved by thymidine phosphorylase, and the resultant 2-deoxy-α-D-ribose-1-phosphate is coupled to the purine by purine nucleoside phosphorylase (Krenitsky *et al.*, 1981, 1986, 1989). These reactions do not employ protecting groups, and the only chromatography is a low-resolution ion-exchange column after the transglycosylation step to remove the excess thymidine, thymine, and 2-deoxy-α-D-ribose-1-phosphate. The hypoxanthine (**14a**) or guanine (**14b**) nucleosides are obtained by deamination of **13a/b** using adenosine deaminase.

2′-Deoxy-[3-^{15}N]adenosine

Synthesis of [3-^{15}N]adenine also uses an azo coupling reaction for introduction of the ^{15}N atom (Scheme 3) (Rhee and Jones, 1990). In this case, coupling of a [β-^{15}N]-diazonium ion with 2-bromoimidazole-4,5-dicarboxylate (**16**) occurs with concomitant decarboxylation to give **17**. The bromine is used here as a protecting group, to direct the coupling reaction to the 4/5 position, rather than the 2 position. The bromine is cleaved during hydrogenolysis of the azo linkage, so that no

Scheme 3



additional steps are required for its removal. Treatment of **17** with cold ethyl chloroformate followed by ammonia gives the corresponding carboxamide (**18**). Hydrogenolysis of **18** then gives 5-[^{15}N]-amino-4-imidazolecarboxamide (**19**), which is converted to [3-^{15}N]hypoxanthine (**20**) by reaction with diethoxymethyl acetate in DMF at reflux. The conversion of **20** to [3-^{15}N]adenine (**22**) is then carried out in two steps by first reaction of **20** with $POCl_3$ to give 6-chloro-[3-^{15}N]purine (**21**) followed by treatment with ethanolic ammonia at 120°C to give **22**. The same enzymatic transglycosylation procedure then gives 2′-deoxy-[3-^{15}N]adenosine (**23**). The only protecting group used in these reactions is the bromine atom, which is introduced before the ^{15}N atom and which is cleaved during the hydrogenolysis of the azo linkage. The only chromatography is the low-resolution ion-exchange column used after the transglycosylation step. The 5-[^{15}N]amino-4-imidazolecarboxamide (**19**) intermediate is a versatile intermediate that can be used to prepare other ^{15}N3-labeled bases and nucleosides (Golding *et al.*, 1986; Chern and Townsend, 1985; Groziak and Townsend, 1986; Chern *et al.*, 1991).

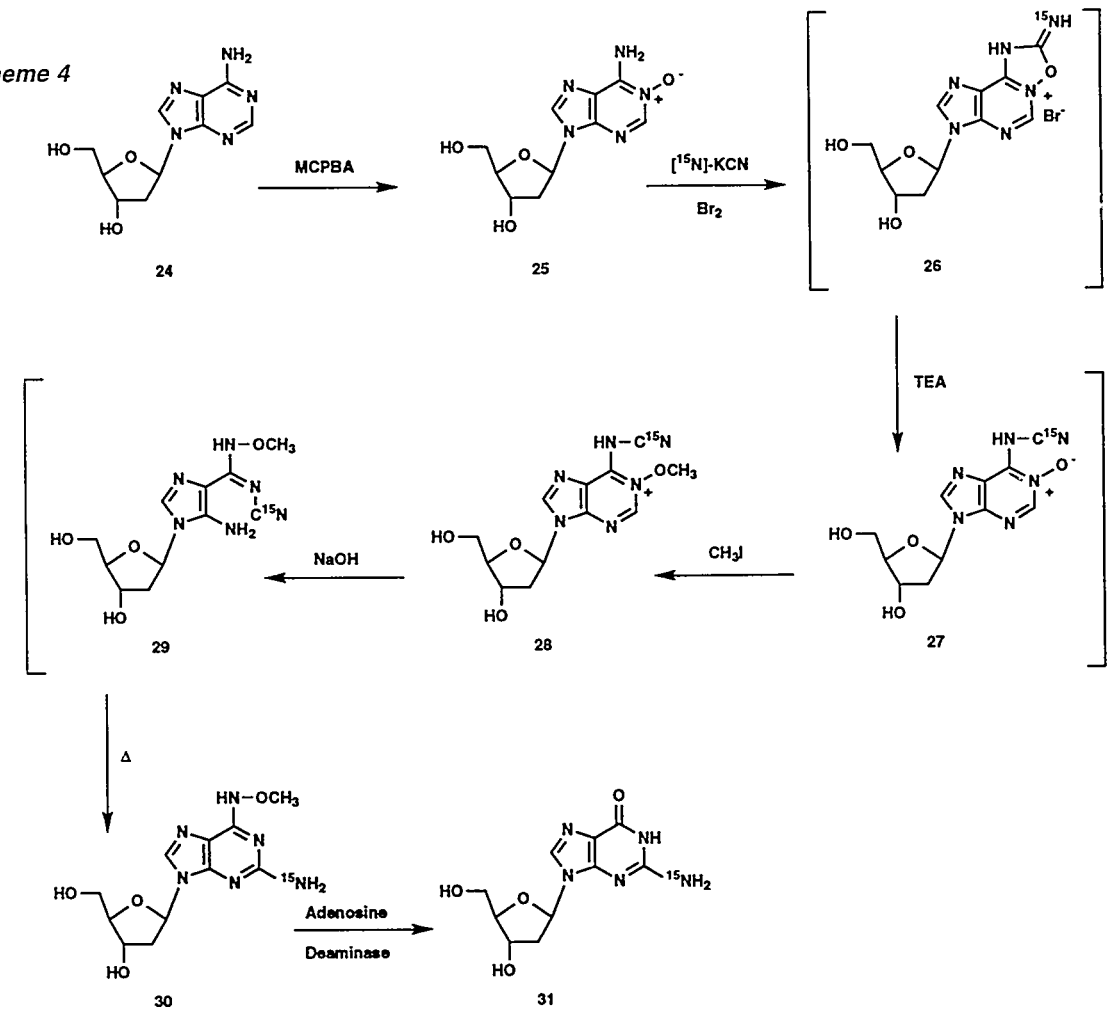
2′-Deoxy-[1-^{15}N]- and 2′-Deoxy-[2-^{15}N]guanosine

The syntheses of 2′-deoxy-[1-^{15}N]guanosine and 2′-deoxy-[2-^{15}N]guanosine (**31**) make use of an adenine-to-guanine transformation that is based on a route developed by Ueda for synthesis of 6-thioguanine and 2,6-diaminopurine nucleosides (Ueda *et al.*, 1978). The synthesis of 2′-deoxy-[2-^{15}N]guanosine (**31**) is shown in Scheme 4 (Goswami and Jones, 1991). In this case, the ^{15}N source is KC^{15}N, which is used in an *in situ* generation of cyanogen bromide. In the first step, 2′-deoxyadenosine is converted to the crystalline N1-oxide (**25**) by oxidation with m-chloroperbenzoic acid (MCPBA) (MacCoss *et al.*, 1980; Zhao and Jones, unpublished results) or with magnesium monoperoxyphthalate (MMPP). Then a series of five reactions is carried out without purification, beginning with reaction of **25** with the [^{15}N]cyanogen bromide. Opening of the oxadiazoline ring of **26** by treatment with triethylamine is followed by methylation of the N1-oxide using methyl iodide, and Dimroth rearrangement promoted by treatment with sodium hydroxide. The 6-methoxyamino derivative (**30**) then is purified by reversed-phase or ion-exchange chromatography and 2′-deoxy-[2-^{15}N]guanosine (**31**) is obtained by treatment with adenosine deaminase. There are no protecting groups used in this transformation, and there is only one chromatographic purification. Synthesis of 2′-deoxy-[1-^{15}N]guanosine is carried out using the same reactions starting from 2′-deoxy-[*amino*-^{15}N]adenosine and using unlabeled cyanogen bromide.

Oligonucleotide Synthesis

Both the phosphoramidite (Mateucci and Caruthers, 1981) and H-phosphonate (Froehler *et al.*, 1986; Froehler and Mateucci, 1986; Garegg *et al.*, 1986)

Scheme 4



methods have been used to prepare ^{15}N-labeled DNA fragments (Kupferschmitt *et al.*, 1987; Massefski et al, 1990; Gao and Jones, 1987a; Wang *et al.*, 1991a, 1991b; Gaffney *et al.*, 1992, 1993; Rhee *et al.*, 1993; Goswami *et al.*, 1993). An advantage of the H-phosphonate method is that it allows recovery and reuse of monomers (Gao *et al.*, 1991; Seliger and Rösch, 1990). This is particularly important with valuable ^{15}N-labeled monomers.

An alternative approach, recently reported by Seliger, is to combine the two methods, using the H-phosphonate method for any rare or valuable monomer and commercially available phosphoramidites for the rest (Geiger *et al.*, 1993).

At present, we use the H-phosphonate method and carry out the oligonucleotide synthesis on a tentagel polystyrene/polyethylene glycol (PEG-PS) support

(Gao *et al.*, 1991). The PEG-PS support allows a two- to threefold increase in scale for a given size reactor, relative to the more commonly used controlled-pore-glass (CPG) supports. This makes it possible to do a 30–35 μmole scale synthesis in a standard 10–15 μmole cartridge. We generally use adamantoyl chloride (Andrus *et al.*, 1988) as the condensing agent, although pivaloyl chloride (Froehler *et al.*, 1986; Froehler and Mateucci, 1986) also works well, along with cyanoethyl-H-phosphonate for capping (Gaffney and Jones, 1988). Other simple H-phosphonates also can be used for capping (Andrus *et al.*, 1988; Froehler, 1993). In addition, we have constructed devices to automate the trityl assay and to better regulate the amount of monomer used in the coupling steps. These devices use glass columns with optical liquid-level sensors to control the appropriate solenoid valves in the synthesizer (Gao *et al.*, 1991).

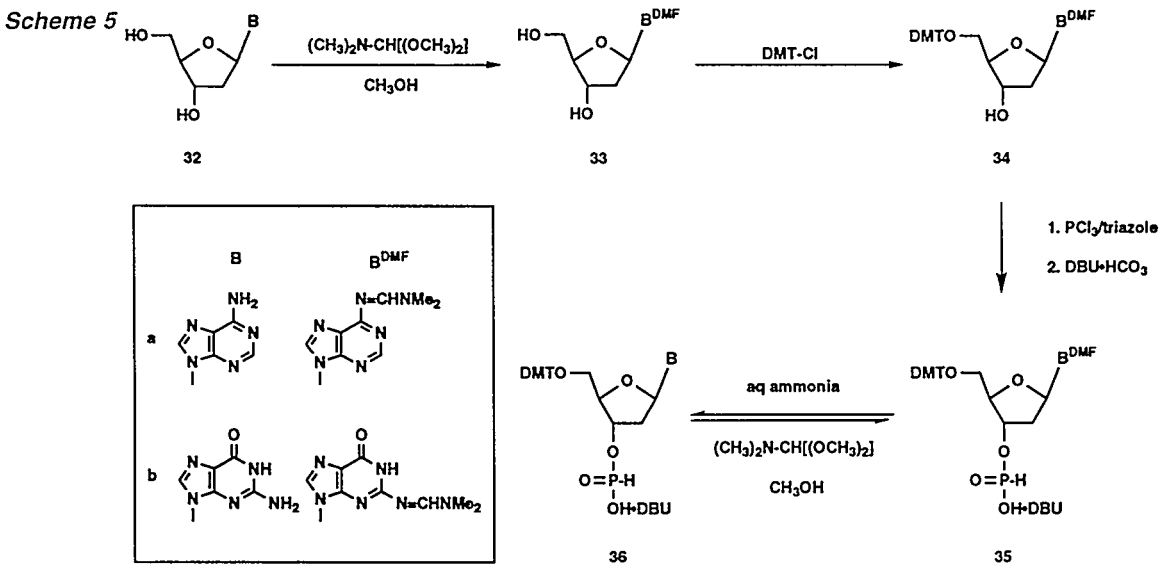
H-Phosphonate DNA Synthesis Without Amino Protecting Groups

We recently have found that the nucleoside amino groups do not react with activated nucleoside 3'-H-phosphonates (Kung and Jones, 1992). This makes it potentially feasible to carry out H-phosphonate DNA synthesis without amino protecting groups on any of the monomers. In practice, we have found that unprotected dC and dG monomers generally give lower coupling yields than do their amino protected derivatives while, in contrast, the unprotected dA monomer gives excellent coupling yields (Zhao and Jones, unpublished results). One advantage of using amino-unprotected dA is that it is much less susceptible to depurination than is benzoylated dA. In addition, the amino-unprotected monomer can be prepared easily in high yield (Scheme 5). The *N,N*-dimethyl-aminomethylene group serves as a transient amino protecting group during the tritylation reaction because 4,4'-dimethyloxytrityl chloride reacts with unprotected nucleoside amino groups. The *N,N*-dimethylamino-methylene group is removed at the end of this synthesis. The same route can be used for dG, but in this case the *N,N*-dimethyl-aminomethylene group is left on because it appears thus far that this DMF-protected monomer gives coupling yields that are comparable to the standard isobutyryl-protected monomer. The advantage for ^{15}N-labeled monomers is that preparation of the *N,N*-dimethy-laminomethylene derivatives proceeds cleanly in quantitative yield.

Oligonucleotide Deprotection and Purification

We purify the ^{15}N-labeled oligonucleotides by HPLC both before and after detritylation (Jones, 1994; Gaffney and Jones, 1989). The first purification cleanly separates tritylated product from untritylated failure sequences and also is used to fractionate carefully the tritylated product. There are always a few impurities that are actually better resolved at this stage than they are after detritylation.

Scheme 5



Further, because any product that is inadvertently detritylated before the first HPLC purification will be lost, it is important to use extra care in handling the tritylated material. For this reason, the ammonia deprotection step is carried out for 2 to 3 days at room temperature rather than overnight at 65°C. The ammonia solution is then concentrated to remove most of the ammonia, so that it can be lyophilized. Simple evaporation of the ammonia solution to dryness is likely to result in detritylation.

Fractions from the first HPLC purification are detritylated and purified a second time by HPLC. The detritylation can be effected simply by evaporation to dryness. However, use of 0.1 M acetic acid allows the detritylation to be carried out under controlled conditions where depurination can be minimized (Gaffney and Jones, 1989). The use of dilute acetic acid gives much less depurination than the standard 80% acetic acid treatment. Finally, the pure oligonucleotide is freed of traces of triethylamine and acetic acid by a third HPLC procedure using 0.1 M ammonium bicarbonate buffer. This buffer is much more volatile than is the triethylammonium acetate buffer we generally use for the high-resolution purifications. Finally, the product is converted to the sodium form by ion exchange. Homogeneity of the product is checked by HPLC on both C_3 and C_{18} analytical columns, and the product is characterized by analysis of the ratio of 2'-deoxynucleosides produced upon enzymatic degradation of a 1 OD sample using a combination of a phosphodiesterase and a phosphatase. We generally use a combination of venom phosphodiesterase with an alkaline phosphatase (Gaffney and Jones,

1989; Kuzmich *et al.*, 1983), although in some cases using venom alone (Gaffney *et al.*, 1982) or, alternatively, nuclease P1 (Gao *et al.*, 1992) can be beneficial.

^{15}N NMR Studies

Protonation of A•C and A•G Mispairs

We used 2'-deoxy-[1-^{15}N]adenosine to probe protonation of A•C and A•G mispairs in {d(CG[1-^{15}N]AGAATTCCCG)}$_2$ and {d(CGGGAATTC[1-^{15}N]ACG)}$_2$, respectively (Wang *et al.*, 1991b). Protonation of the adenine N1 is known to bring about a large (~70 ppm) upfield shift (Büchner *et al.*, 1978). We monitored the ^{15}N chemical shifts of these duplexes over a pH range of ~5 to 8. In this case, the ^{15}N chemical shift was obtained by a heteronuclear multiple quantum correlation experiment via the coupling to 2H. For the molecule with the A•C mispairs, we obtained a sigmoidal plot of chemical shift *vs* pD (Fig. 1), which gave an apparent pK_D of 6.6. This value is ~2.5 units above the monomer pK and demonstrates unequivocally that the low-pH form of this A•C mispair is protonated on the adenine N1. For the molecule with the A•G mispairs, we again demonstrated that the low-pH form is protonated on the adenine N1 (Fig. 2), but we were not able to determine the pK because of line broadening at intermediate pD values. In this case, a *syn/anti* conformational change for the guanine residue that accompanies the protonation may be the cause of the line broadening. These experiments provided the first direct evidence for protonation of a specific nitrogen in a mispair.

Proton NMR studies of the same molecules had been unable to detect this adenine N1-H$^+$ proton (Gao and Patel, 1987, 1988; Kalnik *et al.*, 1988)—perhaps because of rapid exchange—and the proton is too small to show up in x-ray studies.

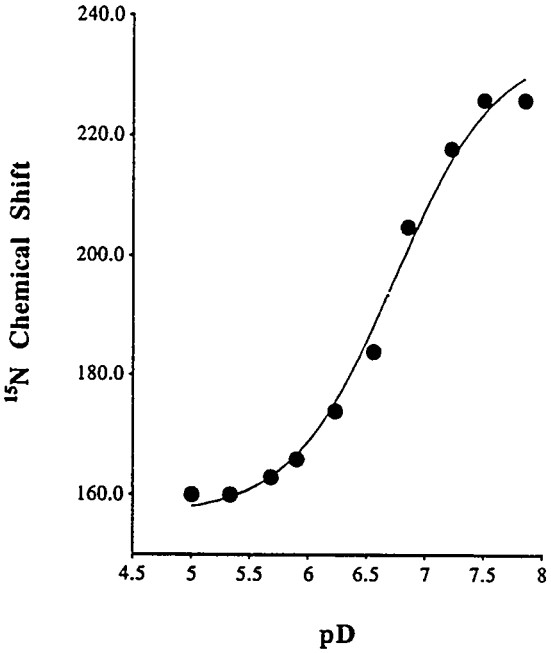


Fig. 1. A plot of ^{15}N chemical shift vs pD for d(CG[1-^{15}N]AGAATTCCCG). The chemical shifts are reported relative to NH_3. The spectra were obtained by a 2D ^1H-^{15}N COSY experiment at 31°C in D_20 containing 0.1 M NaCl, 10 mM phosphate, and 0.1 mM EDTA. The spectra were recorded at a ^1H frequency of 500 MHz with a resolution of 7.5 ppm/point (equivalent to a t_1 acquisition time of 2.7 ms) for the ^{15}N dimension. Four transients for each FID were acquired at a t_2 acquisition time of 128 ms, an experimental recycle time of 2.2 s, a τ delay equivalent to J-coupling of 14.5 Hz, and a total acquisition time of 10 min. To more accurately determine the ^{15}N chemical shift, higher resolution spectra were also obtained, at a similar total acquisition time, by reducing the spectral window.

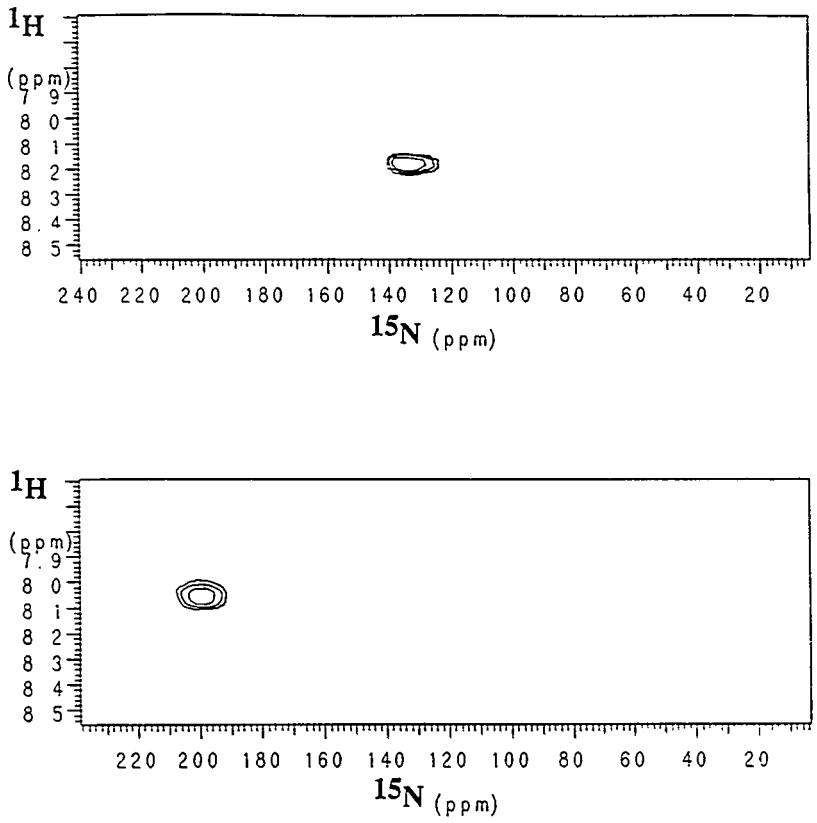


Fig. 2. 2D ^{1}H-^{15}N COSY spectra of d(CGGGAATTC[1-^{15}N]ACG) at 31°C in D_2O containing 0.1 NaCl, 10 mM phosphate, and 0.1 mM EDTA, at pD 4.5 (top) and at pD 6.5 (bottom).

H-Bonding in an A•T Pair

We used the self-complementary molecules d(CGT[1-^{15}N]ACG) and d(CGT[*amino*-^{15}N]ACG) to probe the ability of these ^{15}N labels to monitor Watson-Crick H-bonding in an A•T pair. We first followed the ^{15}N chemical shift of each molecule through the helix-to-coil transition (Fig. 3). In each case we obtained a sigmoidal plot of chemical shift as a function of temperature, consistent with fast exchange between the duplex and single strand. Further, in this and other molecules that we later studied, the overall direction of chemical shift change with increasing temperature is downfield for the sp^2 acceptor nitrogen and upfield for the sp^3 donor nitrogen. This is the directionality to be expected for thermal disruption of H-bonding for each type of nitrogen (Bachovchin, 1986). In addition, the thermodynamic values derived from this ^{15}N NMR data are in agreement with those obtained from optical melting

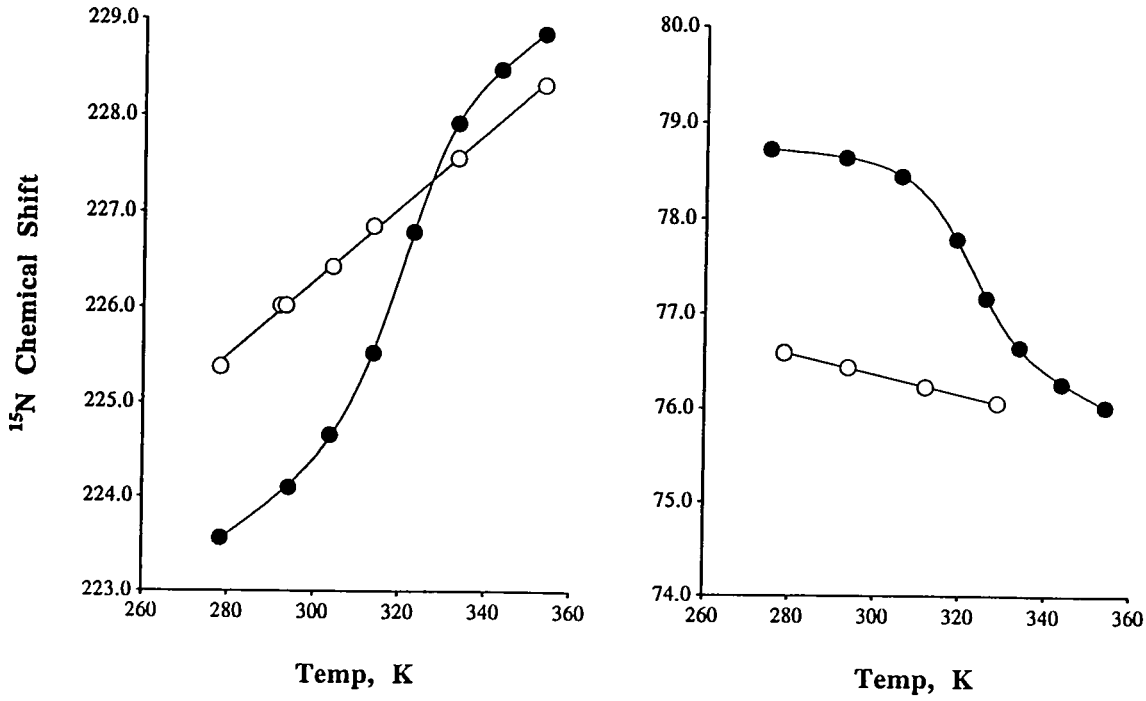


Fig. 3. Plots of ^{15}N chemical shift vs temperature for: (left) d(CGT[1-^{15}N]ACG), 4.8 mM (●) and [1-^{15}N]dA, 6 mM (○); (right) d(CGT[amino-^{15}N]ACG), 3.2 mM (●) and [amino-^{15}N]dA,6 mM (○); in 20% D_2O, 0.1 M NaCl, 10 mM phosphate, 0.1 M EDTA, and pH 6.5.

experiments (Kuzmich *et al.*, 1982). In a later study of the duplex containing the [1-^{15}N]adenine residue, we were able to demonstrate directly H-bonding between the adenine N1 and the thymine 3H (Wang *et al.*, 1991b). By monitoring the N1 chemical shift in mixtures of H_2O and D_2O, we showed that the N1 chemical shift is influenced by the hydrogen isotope present in the H-bond. In this case, the ^{15}N-D interaction is downfield of the ^{15}N-H interaction by ~0.34 ppm.

H-Bonding in an 6-O-Methyl Guanasine (O^6MeG) Thymidine Mispair

We recently studied two molecules of the same sequence—d(CGTGAATTC O^6Me[1-^{15}N]GCG) and d(CGTGAATTC O^6Me[2-^{15}N]GCG)—containing an O^6MeG•T mispair (Fig. 4), in which the O^6MeG residue was labeled at either the N1 or N2 position (Goswami *et al.*, 1993). The N2 chemical shift showed a sigmoidal temperature dependence,

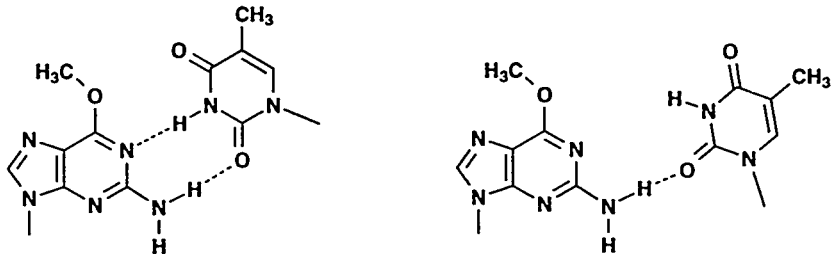


Fig. 4. A rotamer of an O⁶MeG•T pair.

from which we obtained thermodynamic values consistent with those obtained by UV melting experiments (Fig. 5) (Gaffney *et al.*, 1984). In contrast, the N1 chemical shift showed only a linear temperature dependence, identical to that of the monomer. These results demonstrate that there is base•base H-bonding only at the N2. This result is consistent with an earlier ^{1}H NMR study of the same molecule (Patel *et al.*, 1986), but differs from the results of an x-ray study, albeit of a different molecule (Leonard *et al.*, 1990). We had shown previously that, despite the known preferential incorporation of T opposite O⁶MeG, the O⁶MeG•T pair is the least stable of the O⁶MeG•N pairs (Gaffney and Jones, 1989; Kuzmich *et al.*, 1983; Gaffney *et al.*, 1984). These ^{15}N data have provided the only unambiguous solution information on the H-bonding present in this carcinogenic base mispair.

H-Bonding in an O⁶MeG•C Mispair

As in the previous study, we used two molecules of the same sequence, d(CGCGAATTC O⁶Me[1-^{15}N]GCG) and d(CGCGAATTC O⁶Me[2-^{15}N]GCG), containing an O⁶MeG•C mispair, in which the O⁶MeG residue was labeled at either the N1 or N2 position (Gaffney *et al.*, 1993). By monitoring the ^{15}N chemical shifts as functions of both pH and temperature, we showed that there is a pH-dependent conformational transition between a wobble-type pair at high pH and a Watson-Crick-type pair at lower pH, with a pK of 5 for the transition (Figs. 6 and 7). This is about 0.7 unit above the 2'-deoxycytidine value of 4.3, which must be the site of protonation since the change in chemical shift of the O⁶MeG N1 is too small for protonation to have occurred there. Moreover, in marked contrast to our results for the O⁶MeG•T pair, the temperature-dependent ^{15}N chemical shift data show that the O⁶MeG•C pair is H-bonded at both the O⁶MeG N1 and N2 in both the protonated and unprotonated forms. In each case (Fig. 8) we obtained sigmoidal plots of the ^{15}N chemical shift as a function of temperature, with thermodynamic values identical to those from optical melting experiments (Gaffney *et al.*, 1984).

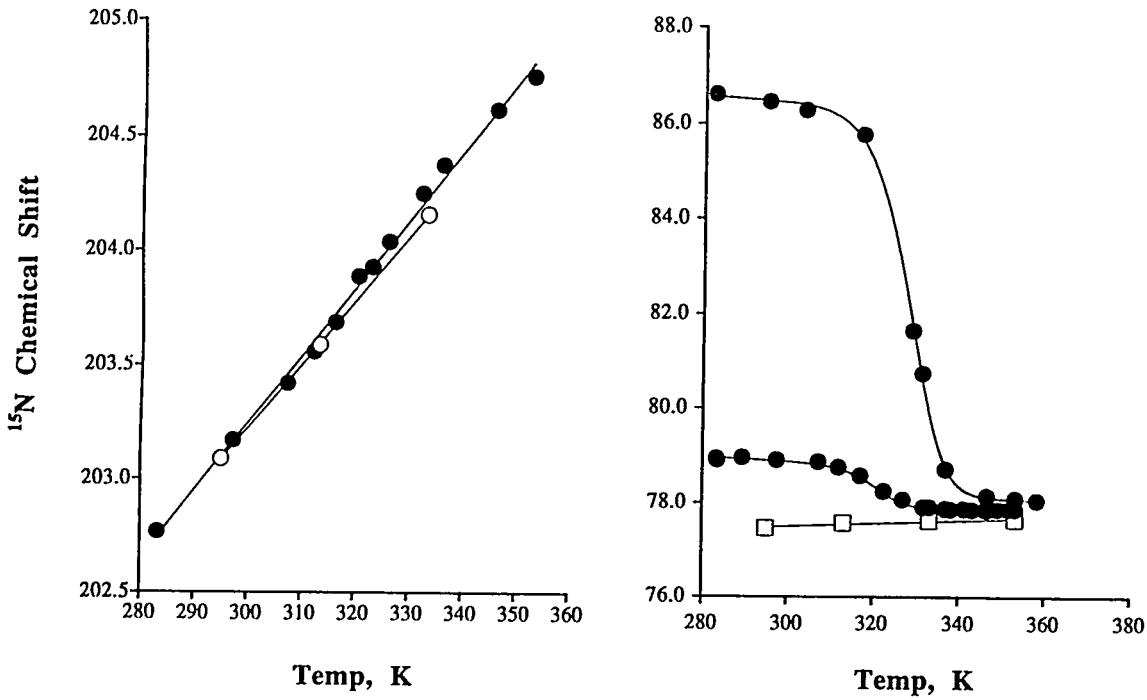


Fig. 5. Plots of ^{15}N chemical shift vs temperature for (left) d(CGTGAATTC O^6Me[1-^{15}N]GCG), 11.2 mM (●), and 6-O-methyl-2'-deoxy-[1-^{15}N]guanosine, 12 mM (○); (right) d(CGTGAATTC O^6Me[2-^{15}N]GCG), 10.7 mM (◑) and 6-O-methyl-2'-deoxy-[2-^{15}N]guanosine, 12 mM (○), in 10–20% D_2O, 0.1 M NaCl, 10 mM phosphate, 1 mM EDTA, pH 6.8–7.3. A single-pulse ^{15}N experiment with a flip angle of ~90° and an experimental recycle delay of 8.8 s was used.

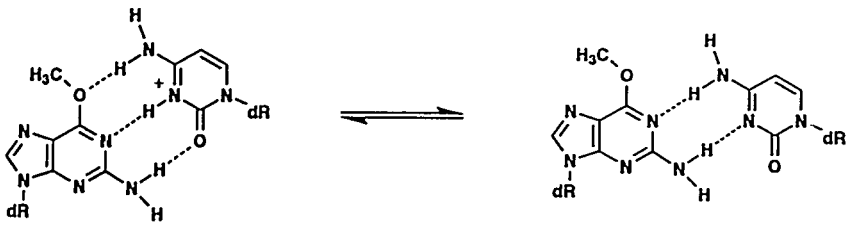


Fig. 6. A resonance form of a protonated O^6MeG•C pair (left) and a rotamer of a wobble O^6MeG•C pair (right).

The influence of the particular H-bond donor and acceptor groups on the magnitude of the chemical shift changes that result from hydrogen bonding is evident. The 3 ppm* upfield shift of the O^6MeG $^{15}N1$ at pH 7 is similar to the 2.6 ppm upfield shift of the adenine N1 we found in an A•T pair. In each case, the labeled nitrogen is an sp^2 H-bond acceptor (Gao and Jones, 1987b). In contrast, the 8.3-ppm-downfield shift of the O^6MeG $^{15}N2$ (pH 7) is much larger for this sp^3 H-bond donor to sp^2 nitrogen acceptor than is either the 1.2 ppm shift of the adenine 6N in an A•T pair or the 1.1 ppm shift of the O^6MeG $^{15}N2$ in an O^6MeG•T pair, both of which are NH_2-O interactions (Goswami *et al.* , 1993). In the O^6MeG•C case, protonation brings about a 4.5-ppm-upfield shift for the $^{15}N1$, which then is involved in an sp^2 N to protonated sp^2 N interaction,

*These chemical shift changes are relative to a base line drawn by extrapolation of the high temperature (single strand) chemical shifts to room temperature. The single strands would normally be hydrated, with the amino groups H-bonded to water oxygen atoms.

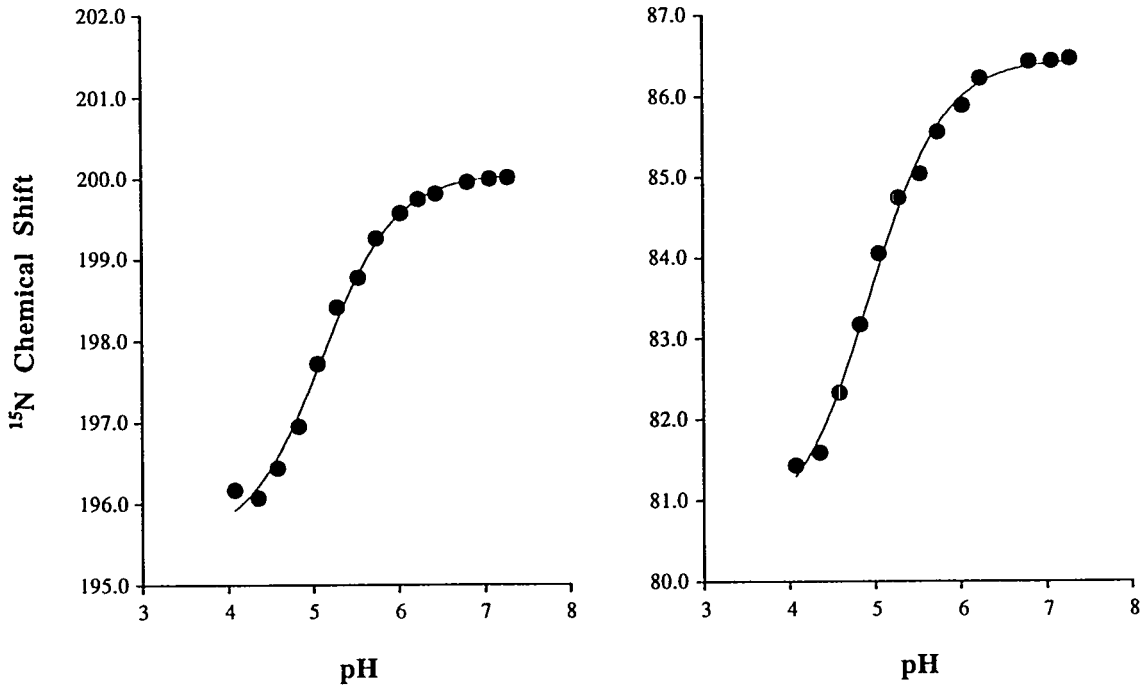


Fig. 7. Plots of the ^{15}N chemical shift vs pH for: (left) d(CGCGAATTC $O^6Me[1\text{-}^{15}N]GCG$) and (right) d(CGCGAATTC $O^6Me[2\text{-}^{15}N]GCG$). The sample consisted of a 1:1 mixture of each molecule at a total strand concentration of 11 mM in 90% H_2O/10% D_2O, 0.1 M NaCl, 10 mM EDTA. The pH was adjusted by addition of NaOH or HCl. A nonlinear least-squares fit to the data gives the curve shown, from which pK values of 5.11 ± 0.04 (left) and 4.94 ± 0.04 (right) were obtained.

and a 5.8-ppm-upfield shift for the ^{15}N2, which then is involved in an NH_2-O interaction. This brings the latter resonance to ~2.5 ppm downfield of that of the single-strand chemical shift, which is then more like the differences we found for the A ^{15}N6 and the O^{6}MeG ^{15}N2 when H-bonded to a carbonyl oxygen. Thus, it appears that the chemical shift change of an amino nitrogen for H-bonding with an imino nitrogen is much greater than that for H-bonding with a carbonyl oxygen, as would be expected.

Guanine Tetrads and Hydration

We used two molecules containing 2′-deoxy-[7-^{15}N]guanosine, d(T[7-^{15}N]GGGT) and d(G[7-^{15}N]GTTTTTGG), as well as the corresponding duplexes with their Watson-Crick complementary strands, to probe the ability of these ^{15}N labels to monitor major groove interactions (Gaffney *et al.*, 1992). In this case, we again used an HSQC experiment, detecting the N7 chemical shift via the 8H atom. In all of the duplexes we have examined to date, the single strand and duplex

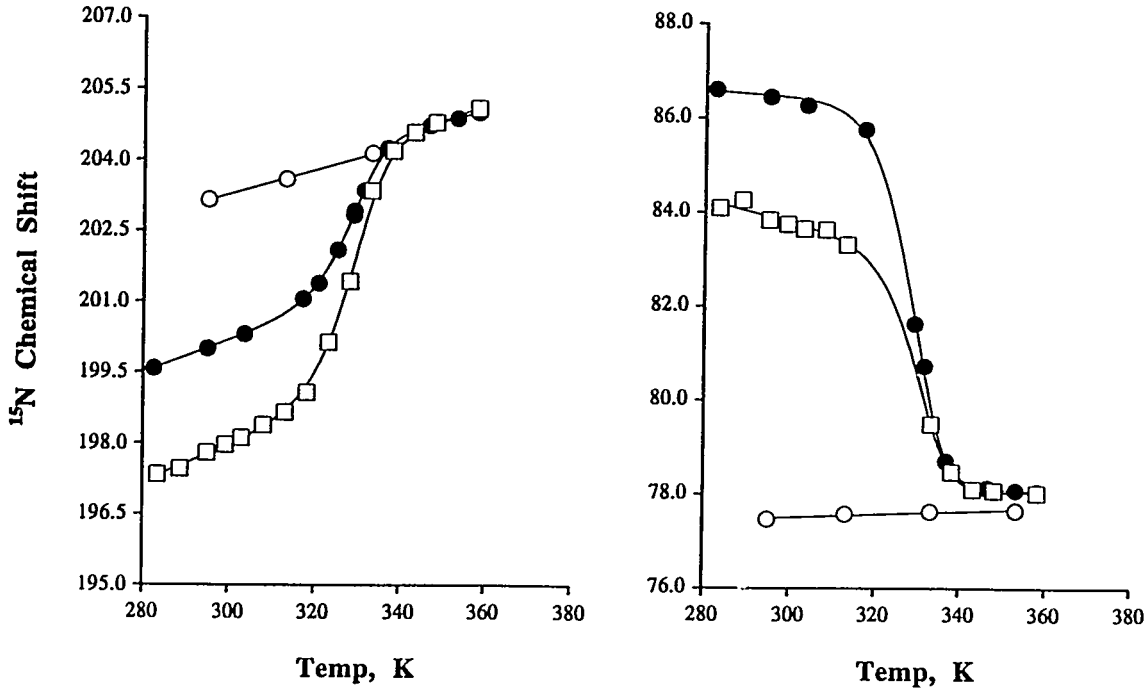


Fig. 8. Plots of the ^{15}N chemical shift vs temperature at pH 7 (●) and at pH 5 (○) for (left) d(CGCGAATTC O^{6}Me[1-^{15}N]GCG) and (right) d(CGCGAATTC O^{6}Me[2-^{15}N]GCG), and of dO^{6}Me[1-^{15}N]G (left) and dO^{6}Me[2-^{15}N]G (right) at pH 7 (□). The sample conditions were as described in Fig. 7. A nonlinear least-squares fit to the data gives the curves shown.

forms appear to be in fast exchange, giving averaged chemical shifts. This is also true for these duplexes (Fig. 9), but the tetrads are in slow exchange. Thus, we observed separate resonances and the tetrad resonance diminishing in intensity with increasing temperature. At 80°C, only the single-strand resonance remains. At low temperatures, the chemical shifts of the tetrad and single strand are nearly identical, indicative of a similar degree of H-bonding to the N7 in each case, regardless of whether the H-bond donor is water or a Hoogsteen interaction with a guanine amino group. The N7 resonance in the duplex d(G[7-^{15}N]GTTTTTGG)•d(CCAAAAACC), however, is ~1.5 ppm downfield, which is indicative of some restriction of hydration

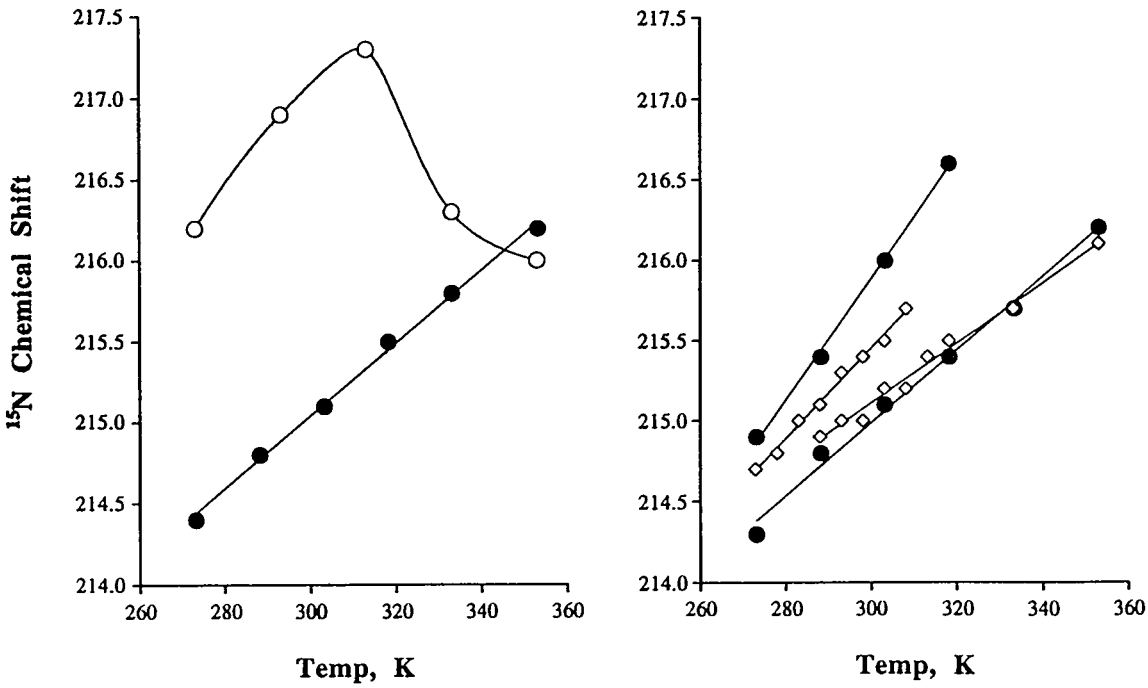


Fig. 9. Plots of the ^{15}N chemical shift vs temperature for: (left) d(G[7-^{15}N]GTTTTTGG)•d(CCAAAAACC), 7.8 mM (○), and d(T[7-^{15}N]GGGT)•d(ACCCA), 9.0 mM (●); (right) d(G[7-^{15}N]GTTTTTGG), 10.1 mM (◊), and d[T[7-^{15}N]GGGT], 5.0 mM (●), in D_2O containing 0.1 M NaCl, 10 mM phosphate, 0.1 mM EDTA at pH 6.9–7.0, over a temperature range of 0–80°C. The data were obtained by using an antiphase version of the heteronuclear single quantum correlation experiment: 90° (H_x) - 1/(4 $^2J_{NH}$) - 180°(H), 180°(N) - 1/(4 $^2J_{NH}$) - 90° (H_y), 90°($N_{\pm x}$) - t_1/2 - 180° (H) - t_1/2 - 90° (H_x) , 90° (N_x) - Acq.(±). The spectra were recorded at a ^{1}H frequency of 500 MHz with a resolution of 0.3 ppm/point for the ^{15}N dimension. Two to four transients for each FID were acquired at a total acquisition time of 4 to 8 min.

in this duplex. In contrast, the duplex d(G[7-^{15}N]GGGT)•d(ACCCA), which has a 5′ terminal A•T pair that would be substantially frayed, showed an N7 chemical shift identical to that of the single strand. The ability of these ^{15}N labels to monitor the most ubiquitous ligand interaction, hydration, may prove to be particularly useful.

Binding of Netropsin and Distamycin to {d[CGCGAATTCGCG]}$_2$

To probe the ability of ^{15}N labeling to monitor minor groove interactions, we used two duplexes containing 2′-deoxy-[3-^{15}N]adenosine, {d(CGCG[3-^{15}N]AATTCGCG)}$_2$ and {d(CGCGA[3-^{15}N]ATTCGCG)}$_2$, with either netropsin or distamycin bound to the duplex (Fig. 10) (Rhee *et al.*, 1993). These drug•DNA complexes have been well studied by both x-ray (Sriram *et al.*, 1992; Kopka *et al.*, 1985; Coll et al,, 1987, 1989) and ^1H NMR (Patel, 1982; Patel and Shapiro, 1986; Klevit *et al.*, 1986; Pelton and Wemmer, 1988, 1989; Ashcroft *et al.*, 1991; Fagan and Wemmer, 1992). In agreement with these earlier studies, our results show H-bonding at the propyl-amidinium end of each drug, as well as at the formamide terminus of distamycin, and show exclusion of water in the center of the binding site. This is seen (Fig. 11) by the upfield shifts of the A^{17} N3 resonances

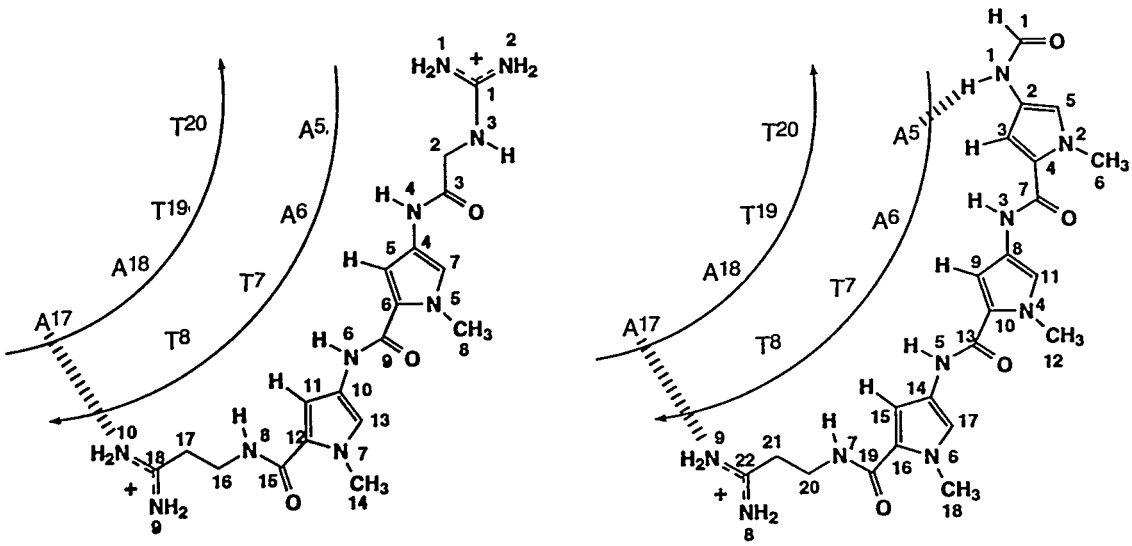


Fig. 10. Diagrammatic representation of (left) the netropsin•{d(CGCGAATTCGCG)}$_2$ complex and (right) the distamycin•{d(CGCGAATTCGCG)}$_2$ complex, showing the drug numbering and the numbering of the AATT binding sites. The dashed lines indicate hydrogen bonding interactions between the indicated adenine N3 atoms and the drug.

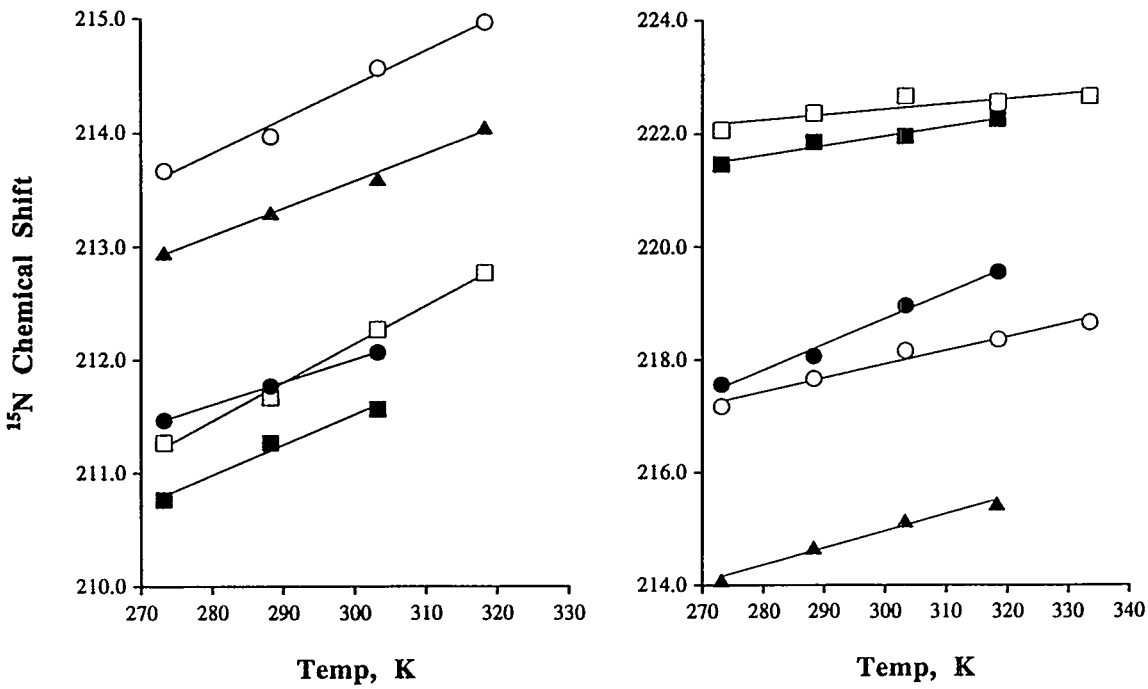


*Fig. 11. Plots of the ^{15}N chemical shift vs temperature for: (left) the indicated resonances of d(CGCG[3-^{15}N]AATTCGCG), A^{17} N3 in the 1•distamycin complex (■), A^{17} N3 in the 1•netropsin complex (□), A^{5} N3 in the 1•distamycin complex (●), A^{5} N3 in the 1•netropsin complex (○), the average of the chemical shifts of the N3 in the absence of drug (▲); (right) the indicated resonances of d(CGCGA[3-^{15}N]ATTCGCG), A^{18} N3 in the 2•distamycin complex (■), A^{18} N3 in the 2•netropsin complex (□), A^{6} N3 in the 2•distamycin complex (●), A^{6} N3 in the 2•netropsin complex (○), the average of the chemical shifts of the N3 in the absence of drug (▲). The spectra were obtained at a drug:DNA ratio of 0.4:1 using an antiphase version of the HSQC experiment: 90° (H_x) - 1/(4 * $^{2}J_{NH}$) - 180° (H), 180° (N) - 1/(4 * $^{2}J_{NH}$) - 90° (H_y), 90° ($N_{\pm x}$) - t_1/2 - 180° (H) - t_1/2 - 90° (H_x), 90° (N_x) - Acq.(±). The spectra were recorded at a ^{1}H frequency of 500 MHz with a resolution of 0.6 ppm/point for the ^{15}N dimension.Two to four transients for each FID were acquired at a total acquisition time of 2–15 min.*

in each complex, along with the upfield shift of the A^5 resonance in the distamycin complex, and the downfield shifts of the A^6 and A^{18} resonances in both complexes. Thus, this type of experiment is able to define sites of water exclusion as well as sites of hydrogen bonding.

Summary

We have developed synthetic routes to ^{15}N-labeled 2′-deoxyadenosine and 2′-deoxyguanosine, as well as some modified purine deoxynucleosides, by which ^{15}N can be introduced readily at any but the glycosidic nitrogen. Thus, all of the nitrogens that are involved in base pairing, either Watson-Crick or Hoogsteen, or other ligand•DNA interactions, can be labeled. These ^{15}N-labeled monomers have been incorporated into DNA fragments by chemical synthesis both by phosphoramidite and H-phosphonate methods. The latter method has the advantage that the excess monomer used in the coupling reaction can be recovered and reused. The ^{15}N labels have proved to be sensitive monitors of the hydrogen bonding present in a Watson-Crick pair as well as that in O^6MeG mispairs. Further, ^{15}N labels in the major and minor grooves are sensitive probes of ligand interactions such as drug-binding and hydration.

References

Andrus, A., Efcavitch, J.W., McBride, L.J., and Giusti, B. (1988) *Tetrahedron Lett. 29*, 861-864.

Ashcroft, J., Live, D.H., Patel, D.J., and Cowburn, D. (1991) *Biopolymers 31*, 45-55.

Bachovchin, W.W. (1986) *Biochemistry 25*, 7751-7759.

Barrio, M.D.C.G., Scopes, D.I.C., Holtwick, J.B., and Leonard, N.J. (1981) *Proc. Natl. Acad. Sci. USA 78*, 3986-3988.

Büchner, P., Maurer, W., and Rüterjans, H. (1978) *J. Magn. Reson. 29*, 45-63.

Chern, J.-W. and Townsend, L.B. (1985) *Tetrahedron Lett. 26*, 6419-6422.

Chern, J.-W., Lin, G.-S., Chen, C.-S., and Townsend, L.B. (1991) *J. Org. Chem 56*, 4213-4218.

Coll, M., Frederick, C.A., Wang, A.H.-J., and Rich, A. (1987) *Proc. Natl. Acad. Sci. USA 84*, 8385-8389.

Coll, M., Aymami, J., van der Marel, G.A., van Boom, J.A., Rich, A., and Wang, A.H.-J. (1989) *Biochemistry 28*, 310-320.

Davis, D.R., Yamaizume, Z., Nishimura, S., and Poulter, C.D. (1989) *Biochemistry 28*, 4105-4108.

Davis, D.R. and Poulter, C.D. (1991) *Biochemistry 30*, 4223-4231.

DeGraw, J.I. and Lawson, J.A. (1978) in *Nucleic Acid Chemistry, Vol. 2*, L.B. Townsend and R.S. Tipson, eds., Wiley Interscience: New York, pp. 921-926.

Fagan, P. and Wemmer, D.E. (1992) *J. Am. Chem. Soc. 114*, 1080-1081.

Froehler, B.C. and Mateucci, M.D. (1986) *Tetrahedron Lett. 27*, 469-472.

Froehler, B.C., Ng, P.G., and Matteucci, M.D. (1986) *Nucleic Acids Res. 14*, 5399-5407.

Froehler, B.C. (1993), *Analogs, Vol. 20*, S. Agrawal, ed., Humana Press, Totowa, New Jersey, pp. 63-80.

Gaffney, B.L., Marky, L.A., and Jones, R.A. (1982) *Nucleic Acids Res. 10*, 4351-4361.

Gaffney, B.L., Marky, L.A., and Jones, R.A. (1984) *Biochemistry 23*, 5686-5691.

Gaffney, B.L. and Jones, R.A. (1988) *Tetrahedron Lett. 29*, 2619-2622.

Gaffney, B.L. and Jones, R.A. (1989) *Biochemistry 28*, 5881-5889.

Gaffney, B.L., Kung, P.-P., and Jones, R.A. (1990) *J. Am. Chem. Soc. 112*, 6748-6749.

Gaffney, B.L., Wang, C., and Jones, R.A. (1992) *J. Am. Chem. Soc. 114*, 4047-4050.

Gaffney, B.L., Goswami, B., and Jones, R.A. (1993) *J. Am. Chem. Soc. 115*, 12607-12608.

Gao, X. and Jones, R.A. (1987a) *J. Am. Chem. Soc. 109*, 1275-1278.

Gao, X. and Jones, R.A. (1987b) *J. Am. Chem. Soc. 109*, 3169-3171.

Gao, X. and Patel, D.J. (1987) *J. Biol. Chem. 262*, 16973-16984.

Gao, X. and Patel, D.J. (1988) *J. Am. Chem. Soc. 10*, 5178-5182.

Gao, H., Gaffney, B.L., and Jones, R.A. (1991) *Tetrahedron Lett. 32*, 5477-5480.

Gao, H., Fathi, R., Gaffney, B.L., Goswami, B., Kung, P.-P., Rhee, Y., Jin, R., and Jones, R.A. (1992) *J. Org. Chem 57*, 6954-6959.

Garegg, P.J., Lindh, I., Regberg, T., Stawínski, J., and Strömberg, R. (1986) *Tetrahedron Lett. 27*, 4051-4054.

Geiger, A., Seliger, H., and Nehls, P. (1993) *Nucleosides and Nucleotides 12*, 463-477.

Gewirth, D.T., Arbo, S.R., Leontis, N.B., and Moore, P.B. (1987) *Biochemistry 26*, 5213-5220.

Golding, B.T., Slaich, P.K., and Watson, W.P. (1986) *J. Chem. Soc., Chem. Commun.* 901-902.

Goswami, B. and Jones, R.A. (1991) *J. Am. Chem. Soc. 113*, 644-647.

Goswami, B., Gaffney, B.L., and Jones, R.A. (1993) *J. Am. Chem. Soc. 115*, 3832-3833.

Griffey, R.H., Poulter, C.D., Yamaizumi, Z., Nishimura, S., and Hurd, R.E. (1982) *J. Am. Chem. Soc. 104*, 5810-5811.

Groziak, M.P. and Townsend, L.B. (1986) *J. Org. Chem 51*, 1277-1282.

Jones, R.A. (1994) in *Protocols for Oligonucleotide Conjugates, Vol. 26*, S. Agrawal, ed., Humana Press, Totowa, New Jersey, pp. 207-231.

Kalnik, M.W., Kouchakdjian, M., Li, B.F.L., Swann, P.F., and Patel, D.J. (1988) *Biochemistry 27*, 100-108.

Klevit, R.E., Wemmer, D.E., and Reid, B.R. (1986) *Biochemistry 25*, 3296-3303.

Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, R.E. (1985) *Proc. Natl. Acad. Sci. USA 82*, 1376-1380.

Krenitsky, T.A., Koszalka, G.W., and Tuttle, J.V. (1981) *Biochemistry 20*, 3615-3621.

Krenitsky, T.A., Rideout, J.L., Chao, E.Y., Koszalka, G.W., Gurney, F., Crouch, R.C., Cohn, N.K., Wolberg, G., and Vinegar, R. (1986) *J. Med. Chem. 9*, 138-143.

Krenitsky, T.A., Hall, W.W., Selph, J.L., Truax, J.F., and Vinegar, R. (1989) *J. Med. Chem. 32*, 1471-1475.

Kung, P.P. and Jones, R.A. (1992) *Tetrahedron Lett. 33*, 5869-5872.

Kupferschmitt, G., Schmidt, J., Schmidt, T., Fera, B., Buck, R., and Rüterjans, H. (1987) *Nucleic Acids Res. 15*, 6225-6241.

Kuzmich, S., Marky, L.A., and Jones, R.A. (1982) *Nucleosides and Nucleotides 10*, 6265-6271.

Kuzmich, S., Marky, L.A., and Jones, R.A. (1983) *Nucleic Acids Res. 11*, 3393-3404.

Leonard, N.J. and Henderson, T.R. (1975) *J. Am. Chem. Soc. 97*, 4990-4999.

Leonard, G.A., Thomson, J., Watson, W.P., and Brown, T. (1990) *Proc. Natl. Acad. Sci. USA 87*, 9573-9576.

MacCoss, M., Ryu, E.K., White, R.S, and Last, R.L. (1980) *J. Org. Chem 5*, 788-794.

Massefski, W., Jr., Redfield, A., Sarma, U.D., Bannerji, A., and Roy, S. (1990) *J. Am. Chem. Soc. 112*, 5350-5351.

Mateucci, M.D. and Caruthers, M.H. (1981) *J. Am. Chem. Soc. 103*, 3185-3191.

Niu, C.-H. (1984) *Anal. Biochem. 139*, 404-407.

Patel, D.J. (1982) *Proc. Natl. Acad. Sci. USA, 79*, 6424-6428.

Patel, D.J., Shapiro, L., Kozlowski, S., Gaffney, B.L., and Jones, R.A. (1986) *Biochemistry 25*, 1036-1042.

Patel, D.J. and Shapiro, L. (1986) *J. Biol. Chem. 261*, 1230-1240.

Pelton, J.G. and Wemmer, D.E. (1988) *Biochemistry 27*, 8088-8096.

Pelton, J.G. and Wemmer, D.E. (1989) *Proc. Natl. Acad. Sci. USA 86*, 5723-5727.

Poulter, C.D. and Livingston, C.L. (1979) *Tetrahedron Lett. 9*, 755-758.

Rhee, Y.S. and Jones, R.A. (1990) *J. Am. Chem. Soc. 112*, 8174-8175.

Rhee, Y., Wang, C., Gaffney, B.L., and Jones, R.A. (1993) *J. Am. Chem. Soc. 115*, 8742-8746.

Robins, M. J. and Trip, E. M. (1973) *Biochemistry 12*, 2179-2187.

Roy, S., Papastavros, M.Z., Sanchez, V., and Redfield, A.G. (1984) *Biochemistry 23*, 4395-4400.

Seliger, H. and Rösch, R. (1990) *DNA and Cell Biology 9*, 691-696

Sethi, S.K., Gupta, S.P., Jenkins, E.E., Whitehead, C.W., Townsend, L.B., and McCloskey, J.A. (1982) *J. Am. Chem. Soc. 104*, 3349-3353.

Sriram, M., van der Marel, G.A., Roelen, H.L.P.F., van Boom, J.H., and Wang, A.H.-J. (1992) *Biochemistry 31*, 11823-11834.

Ueda, T., Miura, K., and Kasai, T. (1978) *Chem. Pharm. Bull. , 26*, 2122-2127.

Wang, C., Gao, X., and Jones, R.A. (1991a) *J. Am. Chem. Soc. 113*, 1448-1450.

Wang, C., Gao, H., Gaffney, B.L., and Jones, R.A. (1991b) *J. Am. Chem. Soc. 113*, 5486-5488.

New Strategy for Stable-Isotope-Aided, Multidimensional NMR Spectroscopy of DNA Oligomers

Okira Ono, Shin-Ichi Tate, and Masatsune Kainosho

Department of Chemistry
Tokyo Metropolitan University
1-1 Minamiohsawa, Hachioji, Tokyo, 1292-03 Japan

Nuclear Magnetic Resonance (NMR) is the most efficient method for determining the solution structures of biomolecules (Wüthrich, 1986). By applying multidimensional heteronuclear NMR techniques to $^{13}C/^{15}N$-labeled proteins, we can determine the solution structures of proteins with molecular mass of 20 to 30 kDa at an accuracy similar to that of x-ray crystallography (Ikura *et al.*, 1990; Clore and Gronenborn, 1991).

Improvements in NMR instrumentation and techniques as well as the development of protein engineering methods for labeling proteins have rapidly advanced multidimensional heteronuclear NMR of proteins. In contrast, multidimensional heteronuclear NMR studies of nucleic acids is less advanced because there were no efficient methods for preparing large amounts of labeled DNA/RNA oligomers. Consequently, early heteronuclear multidimensional NMR studies of DNA focused on the DNA oligomers labeled at specific atoms (Kupferschmitt *et al.* , 1987; Fera *et al.*, 1987; Kieper *et al.*, 1988; Gao and Jones, 1987; Gaffney *et al.*, 1990, 1992, 1993; Goswami and Jones, 1991; Wang *et al.*, 1991; Goswami *et al.*, 1993; Rhee *et al.*, 1993; Manoharan *et al.*, 1987, 1988; Live *et al.*, 1991; Massefski *et al.*,1990; Williamson and Boxer, 1988, 1989a,b; Lancelot *et al.*, 1993; Bornet *et al.*, 1994; Wu and Serianni, 1994; Kellenbach *et al.*, 1991, 1992).

However, the recent development of two novel methods for preparing labeled DNA/RNA oligomers has radically changed the research approach. In one of the methods, labeled NMPs are extracted from bacteria grown on labeled media containing ^{13}C-glucose and/or $^{15}NH_4Cl$. Nucleotide monophosphates (NMPs) are converted to nucleotide triphosphates (NTPs), which are then incorporated into the desired RNA oligomers through

enzymatic synthesis (Niconowicz *et al.*, 1992; Batey *et al.*, 1992; Michnicka *et al.*, 1993) on DNA templates with T7 RNA polymerase (Milligan *et al.*, 1987). This "enzymatic method" has been used to synthesize labeled RNA oligomers used in multidimensional heteronuclear NMR studies (Niconowicz and Pardi, 1992a,b, 1993; Pardi and Niconowicz, 1992; Farmer *et al.*, 1993,1994; Legault *et al.*, 1994; Sklenar *et al.*, 1993a,b, 1994; Battiste *et al.*, 1994; Marino *et al.*, 1994; Hines *et al.*, 1994; Heus *et al.*, 1994).

In "the chemical method," labeled phosphramidite nucleotides are incorporated into oligomers using a DNA/RNA synthesizer. In recent experiments, we have prepared labeled DNA oligomers by solid phase chemical methods (Ono *et al*, 1994a,b; Tat *et al.*, 1994a,b).

These enzymatic and chemical methods are complementary and will contribute to multidimensional heteronuclear NMR studies of nucleic acids and protein-nucleic acid complexes. In this paper, we focus on the chemical preparation of labeled DNA oligomers and discuss a strategy for chemical synthesis of labeled 2'-deoxynucleosides. We also describe several examples of chemical synthesis of labeled DNA oligomers and their application to multidimensional heteronuclear NMR studies.

Enzymatic and Chemical Methods

Before describing the chemical synthesis of labeled nucleosides, we should compare the chemical and enzymatic methods to clarify their properties and the scope of their applications, which are summarized in Table 1.

Table 1. Preparation Methods for Stable-Isotope-Labeled DNA/RNA Oligomers

	Enzymatic Method	Chemical Method
Isotope sources	^{13}C-glucose, $^{15}NH_4Cl$ relatively smaller amount	^{13}C-glucose, $^{15}NH_4Cl$ relatively larger amount
Labeled nucleotide/nucleoside	a mixture of labeled NMPs	isolated labeled nucleosides
Precursors (yield from glucose)	4–5%	18–20% (w/w)
Synthetic method	*in vitro* transcription by T7 RNA polymerase	chemical synthesis on a DNA/RNA synthesizer
Product		
RNA oligomers	+ (longer length)	+ (shorter length)
DNA oligomers	–	+
modified nucleotides	–	+
Labels		
Full label	+	+
Nucleoside-specific label	+	+
Residue-specific label	–	+

The enzymatic approach is suitable for preparing labeled RNA oligomers. With this method, it is possible to label RNA oligomers that are longer than those synthesized by solid-phase chemical methods. However, enzymatic methods cannot be used to synthesize DNA oligomers. In addition, enzymatic methods do not allow preparation of RNA oligomers that are labeled at specific residue(s) and/or contain modified nucleotide(s). In contrast, chemical methods can be used to synthesize both DNA and RNA oligomers as well as DNA/RNA oligomers labeled at specific residues and/or containing modified nucleosides. The DNA and RNA oligomers labeled at specific residues will be important for studies of protein-nucleic acids complexes, and consequently, the chemical method can be applied to a wider range of studies.

The enzymatic method requires relatively small amounts of isotopic precursors to synthesize sufficient amounts of labeled RNA oligomer for NMR studies because a mixture of labeled nucleotides extracted from microorganisms can be used without separation (Niconowicz *et al.*, 1992; Batey *et al.*, 1992; Michnicka *et al.*, 1993). In contrast, for the solid-phase chemical synthesis of oligonucleotides, each labeled nucleoside is separated and chemically converted into the corresponding phosphramidite. Because it is difficult to synthesize these compounds from a small amount of labeled nucleotide isolated from microorganisms, we developed a practical method for obtaining larger amounts of labeled nucleosides. In addition to our work preparing labeled 2′-deoxynucleosides and incorporating them into DNA oligomers, the following sections describe our studies of multidimensional heteronuclear NMR of the labeled oligomers.

Practical Method for Synthesizing Labeled DNA Oligomers

Various pyrimidine 2′-deoxynucleosides with ^{13}C-labeled ribose moieties are chemically synthesized in good yields, starting from ^{13}C-labeled glucose (Ono *et al.*, 1994a; Hardegger, 1962; Hoffer, 1960; Hubbard *et al.*, 1984). In contrast, chemical synthesis of purine 2′-deoxynucleosides is not practical because of the lower yields in coupling the purine bases to ribose (Kazimierczuk *et al.*, 1984; Kawakami *et al.*, 1989). Because of the availability of bacterial strains that overproduce purine nucleotides, they can be prepared in good yields by microbial fermentation. Fermentation methods are not practical for synthesis of pyrimidine nucleosides because the yields are much lower. The labeled purine nucleosides adenosine and guanosine, prepared by fermentation methods, can be chemically converted into corresponding 2′-deoxynucleosides (Robins *et al.*, 1983), which are incorporated into DNA oligomers by the solid-phase phosphoramidite method (Kawakami *et al.*, 1989). The following section describes syntheses of $^{13}C/^{15}N$-labeled 2′-deoxyadenosine and thymidine with labeled sugar moiety ([1′,2′,3′,4′,5′-$^{13}C_5$]-thymidine) and incorporation of the labeled nucleosides into DNA oligomers.

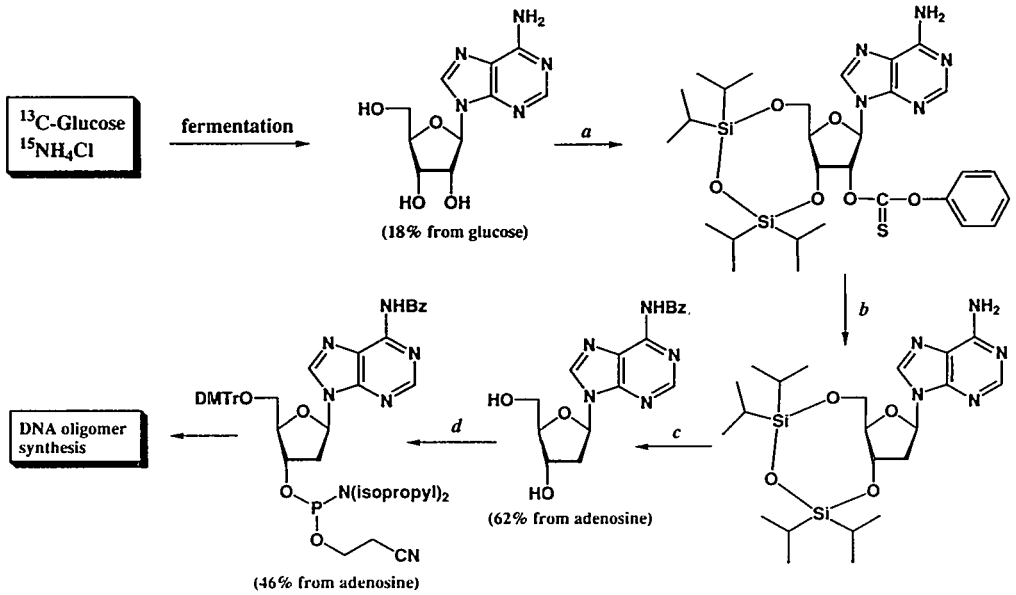


Fig. 1. A scheme for synthesizing labeled 2'-deoxyadenosine and its phosphoramidite derivative. (a) ***1,3-dichloro-1,1,3,3-tetraisopropyldisiloxane in pyridine and phenyl chlorothionoformate, 4-(dimethylamino)pyridine, in acetonitrile,*** (b) ***tributyltin hydride, AIBN in toluene, 65°C,*** (c) ***benzoyl chloride in pyridine and tetra-n-butylammonium fluoride in THF, and*** (d) ***DMTrCl in pyridine and 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite, N,N-diisopropylethylamine.***

Synthesis of $^{13}C/^{15}N$-labeled 2'-deoxyadenosine (Fig. 1)

Adenosine labeled uniformly with ^{13}C and ^{15}N was synthesized (Tate *et al.*, 1994a; Ono *et al.*, 1994b) by microbial fermentation using specific strains of *Bacillus subtilis* (AJ12519 FERMP-11474), which carry mutations that cause the accumulation of adenosine (Japanese Patent, 1993). D-[$^{13}C_6$]glucose and [^{15}N]ammonium chloride were used as labeled precursors. We obtained 18–20% (w/w) labeled adenosine from glucose. The labeled adenosine was converted into 2'-deoxyadenosine (Robins *et al.*, 1983), which was protected and phosphitylated (Sinha *et al.*, 1983) by the method reported by Atrinson and Smith (1985) to give a nucleoside 3'-phosphoramidite monomer in ~50% yield from adenosine. The amidite monomer was used for DNA oligomer synthesis (Beaucage and Caruthers, 1981).

Synthesis of [1',2',3',4',5'-$^{13}C_5$]thymidine (Fig. 2)

Labeled 2-deoxyribose was synthesized Ono *et al.*, 1994a) from uniformly^{13}C-labeled glucose (98% ^{13}C) in 33% yield, with minor modifications of a procedure described by Hardegger (1962). Subsequently, 2-deoxyribose was converted

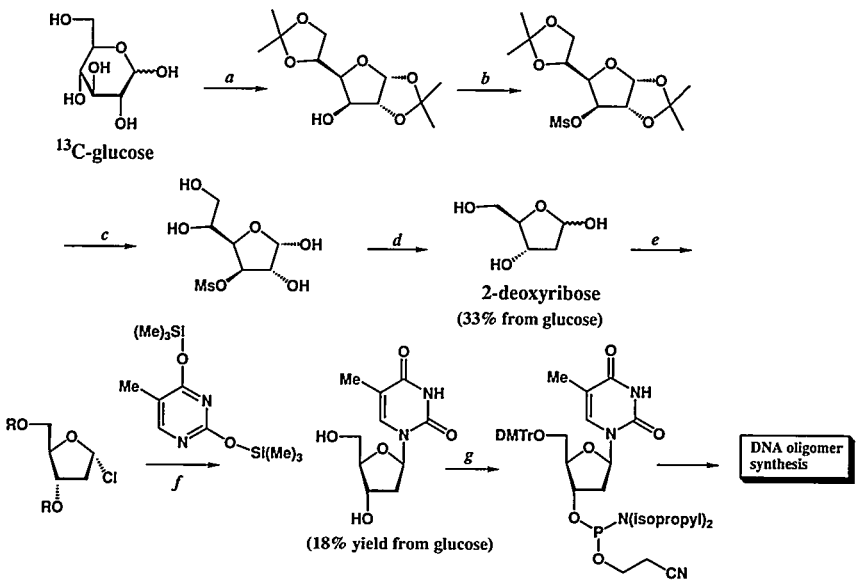


Fig. 2. Scheme for synthesizing labeled thymidine. **(a)** ***acetone, phosphoric acid,*** **(b)** ***methanesulfonyl chloride, triethylamine,*** **(c)** ***methanesulfonic acid in CH_2Cl_2,*** **(d)** ***Na_2CO_3-H_2O,*** **(e)** ***(1) MeOH, H_2SO_4, (2) toluoyl chloride, (3) acetylchloride, MeOH,*** **(f)** ***(1) silylated thymine, $CHCl_3$, (2) NaOH-H_2O, and*** **(g)** ***(1) DMTrCl in pyridine, (2) 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite, N,N-diisopropylethylamine.***

into α-2-deoxy-3,5-di-O-*p*-toluoyl-D-ribofuranosyl chloride(Hoffer, 1960), which was coupled with silylated thymine to yield 3′,5′-di-O-*p*-toluoylthymidine (Hubbard *et al.*, 1984). Deprotection of the toluoyl group produced labeled thymidine in an 18% yield from glucose. The labeled thymidine was protected and phosphitylated to give an amidite-monomer, which was used in DNA oligomer synthesis.

The yields for the labeled phosphoramidite precursors described above are considerably higher than those for the labeled NTPs isolated from microorganisms cultured on labeled glucose—which is reported to be only 4 to 5% from glucose for the mixture of the four nucleotides (Niconowicz *et al.*, 1992; Batey *et al.*, 1992; Michnicka *et al.*, 1993). This chemical method can be used for synthesis of 2′-deoxyuridine (Hubbard *et al.*, 1984), which then can be converted to 2′-deoxycytidine (Reese and Skone, 1984). We have developed a fermentation method that produces higher yields of labeled guanosine than adenosine (described elsewhere) (Ono *et al.*, 1994b). Thus, all four 2′-deoxynucleosides found commonly in DNA can be prepared in good yield.

Synthesis of DNA Oligomers That Contain Labeled Nucleosides

We synthesized DNA oligomers 5′-d(CGCG**AA**TTCGCG)-3′ (**A** = 2′-deoxy-[U-^{13}C,^{15}N]adenosine) and 5′-d(CGCGAA**T**TCGCG)-3′ (**T** = [1′,2′,3′,4′,5′-$^{13}C_5$]thymidine) using the standard protocol on an Applied Biosystems 392 DNA/RNA synthesizer. To conserve labeled amidite monomers, we used only two-thirds of the amount recommended for the standard coupling reaction. The yields of these modified coupling reactions were similar to those obtained using the basic protocol. After deprotection and purification, the amount of oligomer obtained from a 1-μmol synthesis is sufficient for the NMR studies described below.

Multidimensional Heteronuclear NMR Spectroscopy of the Labeled DNA Oligomers

Because methods for preparing labeled DNA/RNA oligomers have now been developed, multidimensional heteronuclear NMR spectroscopy of nucleic acids is progressing rapidly. Here, we discuss the methods for assigning NMR signals and analyzing sugar conformations that we reported recently. We used the DNA dodecamers 5′-d(CGCG**AA**TTCGCG)-3′ (**A** = 2′-deoxy-[U-^{13}C,^{15}N]adenosine) and 5′-d(CGCGAA**T**TCGCG)-3′ (**T** =[1′,2′,3′,4′,5′-$^{13}C_5$]thymidine) as samples for these NMR studies.

Methods for Assigning Signals

Several approaches to assigning NMR signals of nucleic acids are needed: (1) methods for defining the correlation between proton signals of neighboring sugar residues or the correlation between proton signals of a sugar and a base moiety in a nucleoside residue, and (2) a method for assigning all proton signals of a sugar residue. The sequential assignment processes for nucleic acids, unlike those for protein, have been heavily dependent on the through-space connectivities using NOEs (Wüthrich, 1986; Feigon *et al.*, 1983). However, these methods may not be applicable to nucleic acids whose structures deviate from the canonical A, B, and Z forms. Therefore, several alternative methods that use through-bond connectivities have been proposed (Pardi *et al.*, 1983; Sklenar *et al.*, 1986; Kellogg and Schweitzer, 1993; Chary *et al.*, 1993).

Recently, a novel strategy that uses ^{1}H, ^{13}C, ^{31}P triple-resonance experiments (HCP) was invented to create through-bond connectivities between two adjacent sugar residues of labeled RNA oligomers (Heus *et al.*, 1994; Marino *et al.*, 1994b). We have developed an alternative method, HCP-CCH-COSY (Tate *et al.*, 1994b), which results in three independent pathways, to make the sequential connectivities. Figure 3 depicts the pulse schemes used for the HCP correlation and the HCP-CCH-COSY. In the HCP correlation experiment (Fig. 3a), magnetization from the excited ^{1}H is transferred

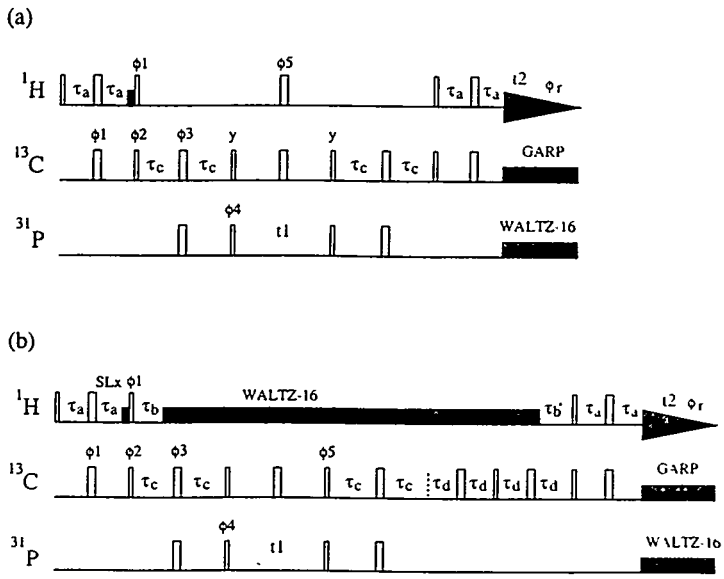


Fig. 3. The pulse sequences for (a) 2D HCP correlation and (b) 2D HCP-CCH-COSY experiments.

to its directly attached ^{13}C by INEPT (Morris and Freeman, 1979) and then to the ^{31}P nuclei (also by INEPT) through the ^{13}C-^{31}P scalar coupling interaction. After t_1-frequency labeling of the ^{31}P, the phosphorous magnetization is transferred back to the original ^{13}C and then is detected by the directly bonded ^{1}H. In the case of the HCP-CCH-COSY experiment (Fig. 3b), the ^{13}C magnetization is relayed to the neighboring ^{13}C through the large, single-bond ^{13}C-^{13}C scalar coupling via COSY before it is finally detected during the t_2 period.

Spectra obtained by these pulse sequences are shown in Fig. 4. Two adenosine residues of the DNA dodecamer 5′-d(CGCG**AA**TTCGCG)-3′, are fully labeled with ^{13}C and^{15}N (98%). (The portion of the dodecamer observed in the HCP and HCP-CCH-COSY spectra is printed in bold.) The three phosphorous atoms are numbered from the 5′-end as P^4, P^5, and P^6, respectively. In the HCP correlation spectrum (Fig. 4a), just two sets of intra- and interresidue cross peaks for H4′ afford the sequential connectivity (Heus *et al.*, 1994; Marino *et al.*, 1994b), as shown by the vertical dotted lines. In contrast, two extra, independent sequential connectivities between A^5 and A^6 are available for the H3′ and H5′,5″ signals (Fig. 4b), which appeared from the relayed coherence transfer via the CC-COSY step. Although it is not the case in the present sample, H4′ and H5′/5″ resonances often overlap. To solve this problem, we can suppress the cross peaks correlated to H5′/5″ by changing $\tau_{b'}$

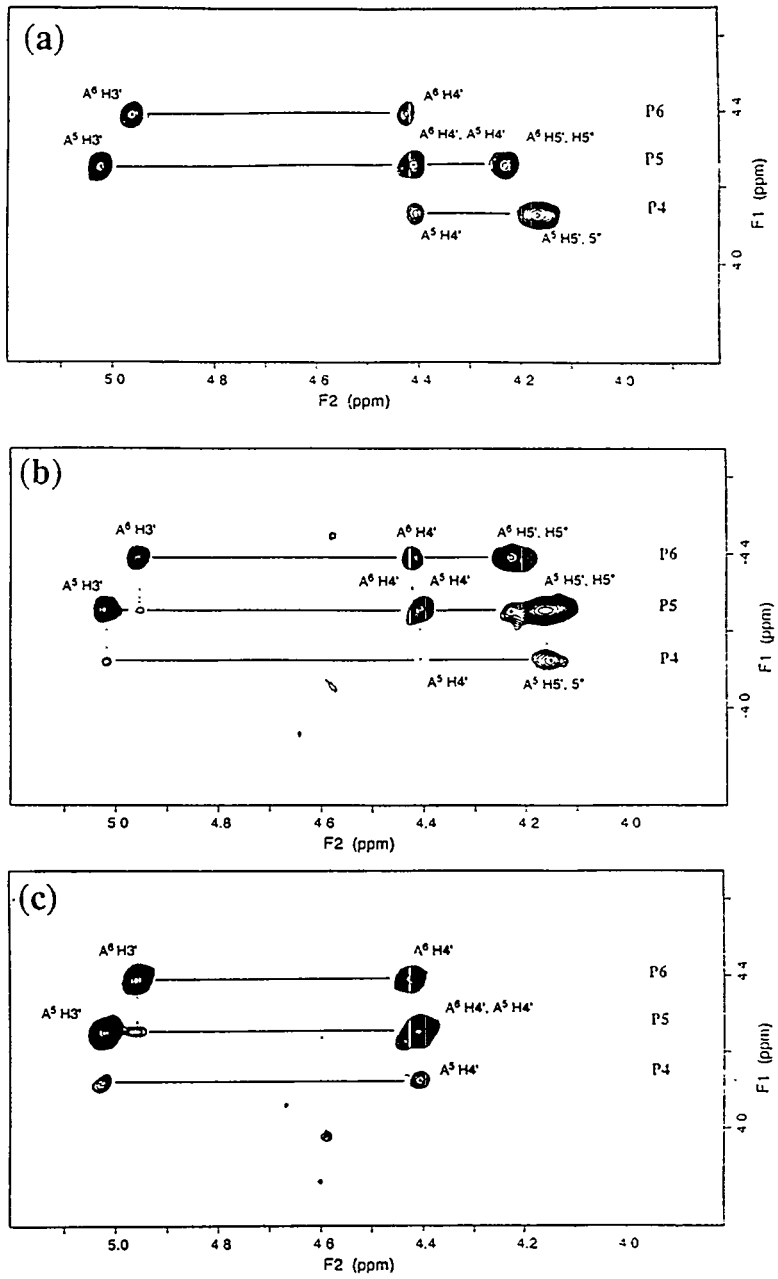


Fig. 4. NMR spectra of 5'-d(CGCGAATTCGCG)$_2$-3' at 40°C. (a) 2D HCP correlation spectrum; (b) and (c) 2D HCP-CCH-COSY spectra of 160 μℓ of a 1.5 mM solution of the labeled dodecamer dissolved in the D_2O buffer containing 0.1 M NaCl, 0.1 mM EDTA, and 0.01 M sodium phosphate at pD 7.0 (direct meter reading). A Shigemi micro cell was used.

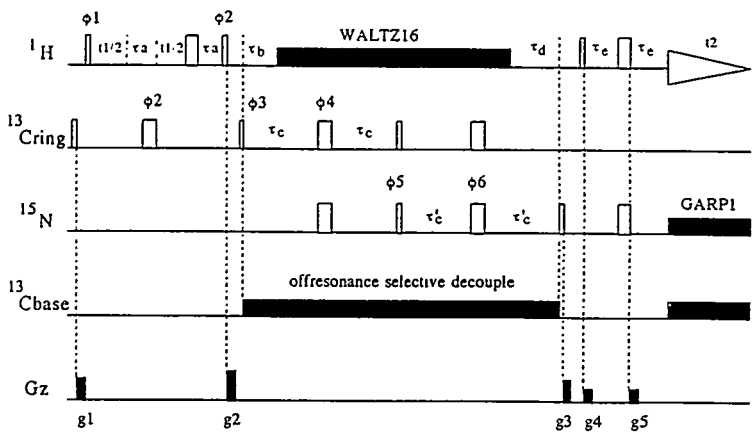


Fig. 5. The 2D ^{1}H, ^{13}C, and ^{15}N triple-resonance pulse sequence H1',C1',N9,8H (Tate et al., *1994a).*

from 1.1 to 3.4 ms, which concomitantly increases the peak intensities of H3′ and H4′ (Fig. 4c). Consequently, the HCP-CCH-COSY pulse sequence will work well in the case of nucleic acids whose NMR signals are severely overlapped.

Several methods for correlating the proton signals of a sugar and base moieties in a nucleoside residue have been proposed (Farmer *et al.*, 1993, 1994; Sklenar *et al.*, 1993a,b, 1994). We also proposed the H1′,C1′,N9,8H pulse sequence (Tate *et al.*, 1994a), by which H1′-8H correlation in purine nucleosides in labeled DNA/RNA oligomers can be identified more efficiently. The H1′,C1′,N9,8H pulse sequence is shown in Fig. 5. The H1′magnetization is transferred in three steps along the pathway shown in Fig.6. The feature that distinguishes the present sequence from other schemes is that it bypasses the magnetization transfer step to the purine C8, and thus eliminates the need to consider the complex heteronuclear *J*-coupling network among the carbon and nitrogen nuclei in the purine bases. Figure 7 shows a 2D H1′,C1′,N9,8H spectrum of 5′-d(CGCGAATTCGCG)-3′. Two H1′-8H correlation signals that correspond to two labeled 2′-deoxyadenosine

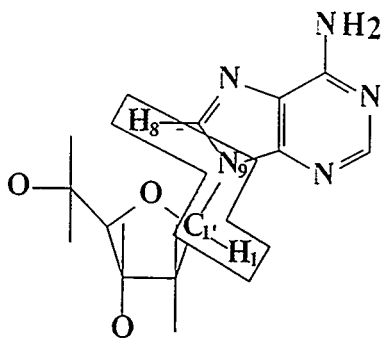


Fig. 6. The coherence transfer pathway of the H1′,C1′,N9,8H experiment for the purine deoxyribonucleotide moiety.

residues are clearly visible. The spectrum was measured over 2 hr using 160 uℓ of a 1.4-mM solution of the dodecamer. The high sensitivity of the method may be partly due to the smaller number of pulses in this sequence compared to other sequences in which selective and/or semiselective pulses are needed to suppress the attenuation of the magnetization caused by spin coupling between N9 and C4. In our case, the scalar coupling effects from $J_{N9,C4}$ and $J_{N9,C8}$ can be completely suppressed by selective decoupling of the base carbons. However, the method may not work as efficiently for pyrimidine nucleosides as it does for purine counterparts because $^2J_{N1,6H}$ (~7 Hz) of pyrimidine nucleosides is smaller than $^2J_{N9,8H}$ (10 Hz) of purine counterparts. Therefore, other sequences using the larger $^1J_{N1,C6}$ (~13 Hz) may be a better choice for pyrimidine nucleosides. The H1',C1',N9,8H experiment can be applied to deoxyguanosine residues and to purine nucleosides in RNA oligomers as well because the relevant coupling constants in the magnetization transfer steps are almost identical for ribo- and 2'-deoxyribo purine nucleosides.

The HCCH-COSY, HCCH-RELAY, or HCCH-TOCSY methods for assigning proton signals for intraresidue sugar have also been applied to labeled RNA oligomers (Nikonowicz and Pardi, 1993; Pardi and Nikonowicz, 1992). In addition, we used the HCCH-TOCSY spectrum for assignment of sugar resonances in a DNA dodecamer, as is discussed below.

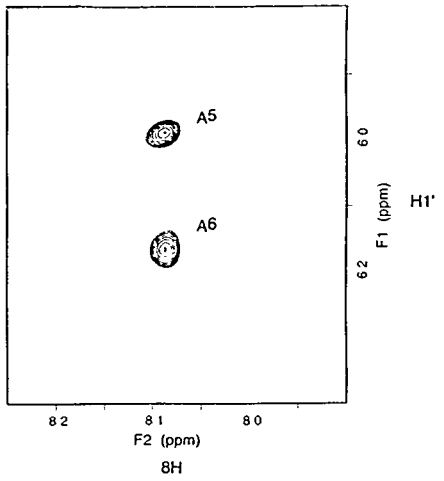


Fig. 7. The 2D H1',C1',N9,8H spectrum of d(CGCGAATTCGCG)$_2$ at 35°C (Tate* et al., *1994a).

Determination of Sugar Conformation

Determination of sugar conformations is one of the important factors that elucidate nucleic acid conformations (Wüthrich, 1986). This method, by which sugar conformation is estimated from J coupling constants of protons in a sugar, has been used to determine sugar conformations by NMR (Rinkel and Altona, 1987). However, because of the severe overlap of sugar proton signals for DNA/RNA oligomers, it has been difficult to determine sugar conformations in relatively large oligomers. The resolution of NMR experiments can be dramatically improved by spreading the spectral data into an increased number of dimensions through multidimensional heteronuclear NMR techniques that employ labeled samples. Furthermore, by using residue-specific labels, it is possible to determine

sugar conformations without any signal overlaps or time-consuming signal assignment processes. One of the advantages of the chemical method is that DNA/RNA oligomers labeled at specific residue(s) can be easily synthesized. The following paragraphs describe an experiment in which sugar conformation was determined using DNA dodecamer labeled at a specific residue: 5′-d(CGCGAA**T**TCGCG)-3′ (**T** = [1′,2′,3′,4′,5′-$^{13}C_5$]thymidine).

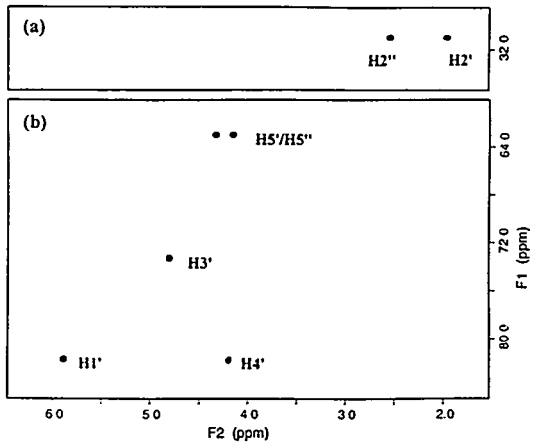


Fig. 8. The CT ^{1}H-^{13}C HSQC spectrum of a D_2O solution of 1.4 mM d(CGCGAA**T**TCGCG)$_2$ containing 0.1 M NaCl, 0.01 M sodium phosphate, and 0.1 mM EDTA. Cross peaks for C2′-H2′, -H2″ are shown in (a) and cross peaks for the others are shown in (b) (Ono* et al., *1994a).

Figure 8 shows a constant-time ^{1}H-^{13}C HSQC (Vuister and Bax, 1992) spectrum with assignments. Because of the ^{13}C labels, it is possible to observe all proton signals without any overlaps. The spectrum was measured with 1.4 mM solution of the dodecamer over 4.8 hr, but a spectrum with a sufficient signal-to-noise ratio could be measured within a shorter time. Because this signal-to-noise ratio can be obtained quickly, many recently developed multidimensional heteronuclear NMR techniques can be used; for example, a HCCH-TOCSY (Kay *et al.*, 1990) spectrum, as is shown in Fig. 9. Cross peaks corresponding to proton signals correlated indirectly by ^{13}C-^{13}C scalar spin couplings are observed in the spectrum. A similar experiment was done with a labeled RNA oligomer (Nikonowicz and Pardi, 1993; Pardi and Nikonowicz, 1992). A NOESY-HMQC (Marion *et al.*, 1989) experiment was also useful for assignment, as is shown in Fig. 10. The intensity of the cross peak of H1′ and H2″ is stronger than that of the cross peak of H1′ and H2′. Similarly, the intensity of the cross peak of H2′ and H3′ is stronger than that observed between H2″ and H3′. These results are useful for the prochiral assignment of H2′/H2″ (Hare *et al.*, 1983).

Figure 11 shows a 2D HCCH-E.COSY (Griesinger and Eggenberger, 1992) spectrum of the dodecamer. All of the vicinal coupling constants listed in Table 2 are determined from the spectrum. The ^{3}J coupling constants are obtained from the cross peaks of methine carbon; for example, $^{3}J_{H1',H2'}$ and $^{3}J_{H1',H2''}$ are measured from the H2′-C1′ and H2″-C1′ cross peaks, respectively. Information obtained from the cross peaks of methylene carbon is different from those of methine carbon. For instance, in the case of C2′carbon, the two values

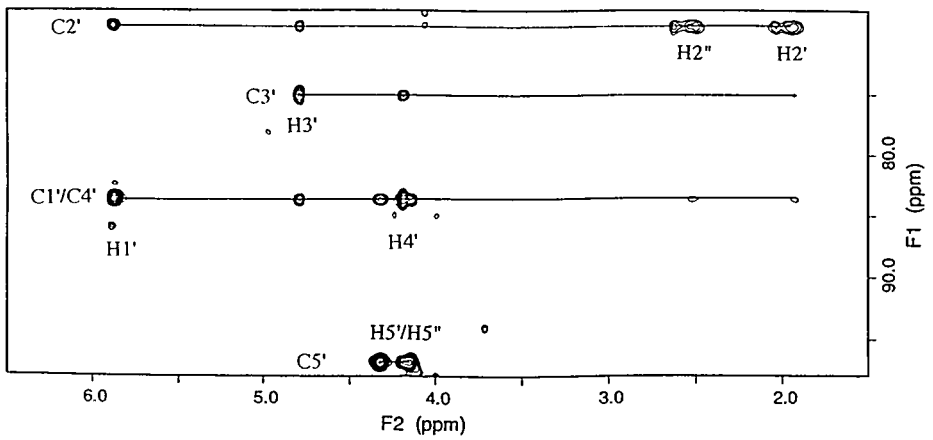


Fig. 9. The 2D HCCH-TOCSY spectrum of d(CGCGAATTCGCG)$_2$. The ^{13}C axis is folded over one time.

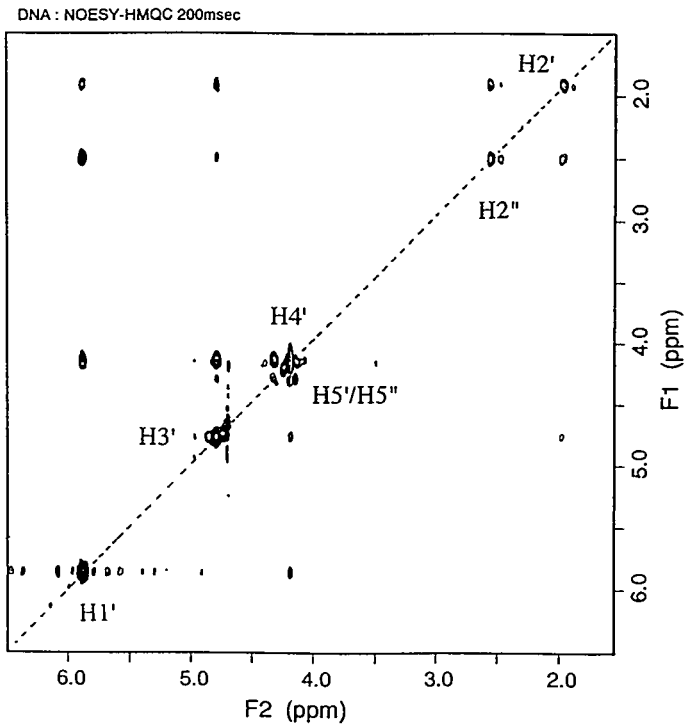


Fig. 10. The 2D NOESY-HMQC spectrum of d(CGCGAATTCGCG)$_2$. The ^{13}C frequency is not encoded. τ_m = 200 ms.

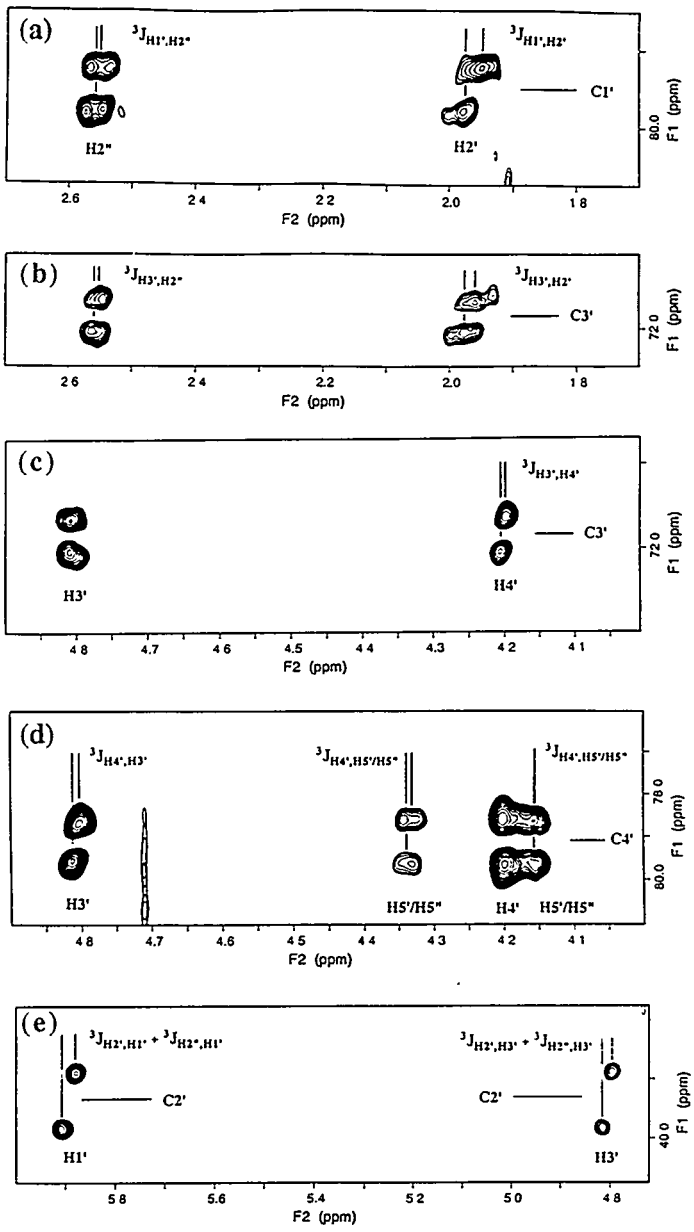


Fig. 11. The 2D HCCH-E.COSY spectrum of a D_2O solution of 1.4 mM d(CGCGAATTCGCG)$_2$ (Ono et al.*, 1994a) Selected parts of the obtained spectrum are shown: (a) C1'-H2'/H2"; (b) C3'-H2'/H2"; (c) C3'-H3',H4'; (d) C4'-H5'/H5" and C4'-H3'; and (e) C2'-H1',H3'.*

($^3J_{H1',H2'} + {}^3J_{H1',H2'}$) and ($^3J_{H3',H2'} + {}^3J_{H3',H2''}$) are obtained from the cross peaks C2'-H1' and C2'-H3', respectively. The accuracy of the observed coupling constants can be confirmed by comparing the coupling constants observed for the C2' and C5' methylene groups with those calculated by adding the values measured independently for the correlation peaks of the relevant methine carbon.

Table 2. Vicinal ^{1}H-^{1}H Spin Coupling Constants of T^7 Residue in d(CGCGAATTCGCG)$_2$, as Determined by the 2D HCCH-E.COSY Spectrum

$^3J_{Hi,Hj}$	Observed Values (Hz)[a]	Cross Peaks[b]
$J_{1'2'}$	11.3	C1'-H2'
$J_{1'2''}$	4.4	C1'-H2"
$[J_{1'2'} + J_{1'2''}]$	15.2	C2'-H1'
$[J_{1'2'} + J_{1'2''}]$	15.7[c]	
$J_{2'2''}$	-14.5	C2'-H2'(H2")
$J_{3'2'}$	7.3	C3'-H2'
$J_{3'2''}$	2.4	C3'-H2"
$[J_{3'2'} + J_{3'2''}]$	9.2	C2'-H3'
$[J_{3'2'} + J_{3'2''}]$	9.7 [c]	
$J_{3'4'}$	4.2	C3'-H4'
$J_{4'5'(5'')}$	3.0	C4'-H5'(H5")
$J_{4'5''(5')}$	<0.6	C4'-H5"(H5')
$[J_{4'5'} + J_{4'5''}]$	3.4	C5'-H4'
$[J_{4'5'} + J_{4'5''}]$	<3.6[c]	

[a] Coupling constants were measured directly with the cursor in the FELIX window and contain a deviation of ± 0.6 Hz.

[b] Cross peaks used to determine the vicinal ^{1}H-^{1}H coupling constants shown to the immediate left.

[c] Values designated were calculated from spin coupling constants observed for the cross peaks between the methine carbon and the methylene protons. Those values were identical (± 0.6Hz) to the observed spin coupling constants for the cross peaks between the methylene carbon and the methine proton.

The preferred conformation of the sugar moiety of the T^7 residue in the dodecamer can be estimated by using the relevant vicinal ^{1}H-^{1}H coupling constants, as described by Rinkel and Altona (1987). The population of the S-conformer is ~95%, with a pseudorotational phase angle of ~120°, which indicates that the labeled deoxyribose ring adopts a nearly C1'-exo conformation. The estimated sugar conformation for the labeled residue is slightly different from the C2'-endo conformation that is typical for B-DNA, but it is similar to the corresponding sugar conformation determined by x-ray analysis of the dodecamer (Dickerson and Drew, 1981a,b): O4'-endo and C1'-exo for T^7 and T^{19} (in the opposite strand), respectively.

Recent advances in computer-assisted analysis of homonuclear ^{1}H 2D NMR experiments make it possible to estimate most of the vicinal coupling constants in the sugar rings of small DNA/RNA oligomers with higher precision than before (Wüthrich, 1986; Schmitz *et al.*, 1992). However, proton signal assignment becomes difficult for larger samples. When specific labels are used, signal assignment becomes more accurate. All vicinal coupling constants of a sugar residue in the DNA dodecamer are determined in the present experiment. The $^3J_{H3',H4'}$ and $^3J_{H4',H5'/5''}$ coupling contstants, which were difficult to obtain by previous ^{1}H-^{1}H homonuclear experiments, can be used not only for conformational analysis of sugar residues but also for conformational analysis of the sugar-phosphate backbone.

Conclusion

In this report, we focused on the chemical synthesis of DNA oligomers labeled at specific residue(s). RNA oligomers with specific labels, which are difficult to synthesize by the enzymatic method, can be synthesized by the chemical method. The specific labels are useful for conformational analysis of larger molecules such as protein-nucleic acid complexes.

Acknowledgments

These studies were supported in part by the special coordination fund of the Science and Technology Agency, the Human Frontier Science Program Organization for a Research Grant (Strasbourg, France), and the Ministry of Education for a Grant-in-Aid for Specially Promoted Research (No. 05101004). We thank the Central Research Laboratories of Ajinomoto Co., Inc. for their assistance in preparing the isotope-labeled adenosine. We also thank Mr. Hiroyuki Kyomori of Kobayashi Perfumery, Ltd., for his generous help in preparing the ^{13}C-labeled thymidine.

References

Atrinson, T. and Smith, M. (1985) *Oligonucleotide Synthesis: A Practical Approach;*, Gait, M.J., ed., IRL Press, Inc.

Batey, R.T., Inada, M., Kujawinski, E., Puglisi, J.D., and Williamson, J.R. (1992) *Nucleic Acids Res. 20*, 4515-4523.

Battiste, J.L., Tan, R., Frankel, A.D., and Williamson, J.R. (1994) *Biochemistry 33*, 2741-2747.

Beaucage, S.L. and Caruthers, M.H. (1981) *Tetrahedron Lett. 22*, 1859-1862.

Bornet, O., Lancelot, G., Chanteloup, L., Thuong, N.T., and Beau, J.-M. (1994) *J. Biomol. NMR, 4*, 575-580.

Chary, K.V.R., Pastogi, V.K., and Govil, G. (1993) *J. Mag. Res. B.110*, 81-83.

Clore, G.M. and Gronenborn, A.M. (1991) *Science 252*, 1390-1399.

Dickerson, R.E. and Drew, H.R. (1981) *J. Mol. Biol. 149*, 761-786.

Dickerson, R.E. and Drew, H.R. (1981) *Proc. Natl. Acad. Sci. USA 78*, 2179-2183.

Farmer II, B.T., Muller, L., Nikonowicz, E.P., and Pardi, A. (1993) *J. Am. Chem. Soc. 115*, 11040-11041.

Farmer II, B.T., Muller, L., Nikonowicz, E.P., and Pardi, A. (1994) *J. Biomolecular NMR 4*, 129-133.

Feigon, J., Leupin, W., Denny, W.A., and Kearns, D.R. (1983) *Biochemistry 22*, 5943-5951.

Fera, B., Singrun, B., Kupferschmitt, G., Svhmidt, J., Buck, F., and Ruterjans, H. (1987) *Nucleosides Nucleotides 6*, 477-481.

Gaffney, B.L., Kung, P.-P., and Jones, R.A. (1990) *J. Am. Chem. Soc. 112*, 6748-6749.

Gaffney, B.L., Wang, C., and Jones, R.A. (1992) *J. Am. Chem. Soc. 114*, 4047-4050.

Gaffney, B.L., Goswami, B., and Jones, R.A. (1993) *J. Am. Chem. Soc. 115*, 12607-12608.

Gao, X. and Jones, R.A. (1987) *J. Am. Chem. Soc. 109*, 1275-1278.

Goswami, B. and Jones, R.A. (1991) *J. Am. Chem. Soc. 113*, 644-647.

Goswami, B., Gaffney, B.L., and Jones, R.A. (1993) *J. Am. Chem. Soc. 115*, 3832-3833.

Griesinger, C. and Eggenberger, U. (1992) *J. Magn. Reson. 97*, 426.

Hardegger, E. (1962) *Methods Carbohydr. Chem. 1*, 177-179.

Hare, D.R., Wemmer, D.E., Chou, S.-H., and Drobny, G. (1983) *J. Mol. Biol. 171*, 319-336.

Heus, H.A., Wijmenga, S.S., van de Ven, F.J.M., and Hilbers, C.W. (1994) *J. Am. Chem. Soc. 116*, 4983-4984.

Hines, J.V., Landry, S.M., Varani, G., and Tinoco, Jr., I. (1994) *J. Am. Chem. Soc. 116*, 5823-5831.

Hoffer, M. (1960) *Chem. Ber. 93*, 2777-2781.

Hubbard, A.J., Jones, A.S., and Walker, R.T. (1984) *Nucleic Acids Res. 12*, 6827-6837.

Ikura, M., Kay, L.E., and Bax, A. (1990) *Biochemistry 29*, 4659-4667.

Japanese Patent No. 4-52118 (1992).

Kawakami, H., Matsushita, H., Shibagaki, M., Naoi, Y., Itho, K., and Yoshikoshi, H. (1989) *Chem. Lett.* 1365-1368.

Kay, L. E., Ikura, M., and Bax, A. (1990) *J. Am. Chem. Soc. 112*, 888-889.

Kazimierczuk, Z., Cottam, H.B., Revankar, G.R., and Robins, R.K. (1984) *J. Am. Chem. Soc. 106*, 6379-6382.

Kellenbach, E.R., von den Elst, H., van der Marel, G.A., van Boom, J.H., and Kaptein, R. (1991) *Recl. Trav. Chim. Pays-Bas 110*, 387-388.

Kellenback, E.R., Remerowski, M.L., Eib, D., Boelens, R., van der Marel, G.A., van den Elst, H., van Boom, J.H., and Kaptein, R. (1992) *Nucleic Acids Res. 20*, 653-657.

Kellogg, G, W. and Schweitzer, B.I. (1993) *J. Biomol. NMR 3*, 577-595.

Kieper, I., Schmidt, T., Fera, B., and Kupferschmitt, G., Schmidt, J., Schmidt, Th., Fera, B., and Ruterjans, H. (1987) *Nucleic Acids Res. 15*, 6225-6241.

Lancelot, G., Chanteloup, L., Beau, J.-M., and Thung, N.T. (1993) *J. Am. Chem. Soc. 115*, 1599-1600.

Legault, P., Farmer II, B.T., Mueller, L., and Pardi, A. (1994) *J. Am. Chem. Soc. 116*, 2203-2204.

Live, D.H., Radhakrishnan, I., Misra, V., and Patel, D.J. (1991) *J. Am. Chem. Soc. 113*, 4687-4688.

Manoharan, M., Gerlt, J.A., Wilde, J.A., Withka, J.M., and Bolton, P.H. (1987) *J. Am. Chem. Soc. 109*, 7217-7220.

Manoharan, M., Ransom, S.C., Mazumder, A., Gerlt, J.A., Wilde, J.A., Withka, J.M., and Bolton, P.H. (1988) *J. Am. Chem. Soc. 110*, 1620-1622.

Marino, J.P., Prestigard, J.H., and Crother, D.M. (1994) *J. Am. Chem. Soc. 116*, 2205-2206.

Marino, J.P., Schwalbe, H., Anklin, C., Bermel, W., Crother, D.M., and Griesinger, C. (1994) *J. Am. Chem. Soc. 116*, 6472-6473.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M., and Clore, G.M. (1989) *Biochemistry 87*, 6150-6156.

Massefski, Jr. W., Redfield, A., Sarma, U.D., Bannerji, A., and Roy, S. (1990) *J. Am. Chem. Soc. 112*, 5350-5351.

Michnicka, M.J., Harper, J.W., and King, G.C. (1993) *Biochemistry 32*, 395-400.

Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987) *Nucleic Acids Res. 15*, 8783-8789.

Morris, G.A. and Freeman, R. (1979) *J. Am. Chem. Soc. 101*, 760-762.

Niconowicz, E.P. and Pardi, A. (1992) *J. Am. Chem. Soc. 114*, 1082-1083.

Niconowicz, E.P. and Pardi, A. (1992) *Nature 355*, 184-186.

Nikonowicz, E.P. and Pardi, A. (1993) *J. Mol. Biol. 232*, 1141-1156.

Niconowicz, E.P., Sirr, A., Legault, P., Jucker, F.M., Baer, L.M., and Pardi, A. (1992) *Nucleic Acids Res. 20*, 4507-4513.

Ono, A., Tate, S., and Kainosho, M. (1994a) *J. Biomol. NMR 4*, 581-586.

Ono, A., Kurita, J., Inoue, T., Pae, I.-H., Tate, S., Ishido, Y., and Kainosho, M. (1994b), *Nucleic Acids Res. Symposium Series No 30*, in press.

Pardi, A. and Nikonowicz, E.P. (1992) *J. Am. Chem. Soc. 114*, 9202-9203.

Pardi, A., Walker, R., Rapoport, H., Wider, G., and Wüthrich, K. (1983) *J. Am. Chem. Soc. 105*, 1652-1653.

Reese, C.B. and Skone, P.A (1984) *J. Chem. Soc. Perkin Trans. I*, 1263-1271.

Rhee, Y., Wang, C., Gaffney, B.L., and Jones, R.A. (1993) *J. Am. Chem. Soc. 115*, 8742-8746.

Rinkel, L.J. and Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621-649.

Robins, M.J., Wilson, J.S., and Hansske, F. (1983) *J. Am. Chem. Soc. 105*, 4059-4065.

Ruterjans, H. (1988) *Nucleosides Nucleotides, 7*, 821-825.

Schmitz, U., Sethson, I., Egan, W.M., and James, T.L. (1992) *J. Mol. Biol. 227*, 510-531.

Sinha, N.D., Biernat, J., and Koster, H. (1983) *Tetrahedron Lett. 24*, 5843-5846.

Sklenar, V., Miyoshi, H., Zon, G., Miles, H.T., and Bax, A. (1986) *FEBS Lett. 208*, 94-98.

Sklenar, V., Peterson, R.D., Rejante, M.R., Wang, E., and Feigon, J. (1993) *J. Am. Chem.Soc. 115*, 12181-12182.

Sklenar, V., Peterson, R.D., Rejante, M.R., and Feigon, J. (1993) *J. Biomol. NMR 3*, 721-727.

Sklenar, V., Peterson, R.D., Rejante, M.R., and Feigon, J. (1994) *J. Biomol. NMR 4*, 117-122.

Tate, S., Ono, A., and Kainosho, M. (1994a) *J. Am. Chem. Soc. 116*, 5977-5978.

Tate, S., Ono, A., and Kainosho, M. (1994b), submitted for publication.

Vuister, G.W. and Bax, A. (1992) *J. Magn. Reson. 98*, 428-435.

Wang, C., Gao, X., and Jones, R.A. (1991) *J. Am. Chem. Soc. 113*, 1448-1450.

Williamson, J.R. and Boxer, S.G. (1988) *Nucleic Acids Res. 16*, 1529-1540.

Williamson, J.R. and Boxer, S.G (1989) *Biochemistry 28*, 2819-2831.

Williamson, J.R. and Boxer, S.G (1989) *Biochemistry 28*, 2831-2836.

Wu, J. and Serianni, A.S. (1994, in press) *Biopolymers. 34*, 1175-1186.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, Inc.

^{13}C Relaxation in an RNA Hairpin

Garry C. King,[1,2] Zhijian Xi,[2]
Malgorzata J. Michnicka,[2] and Connie Akratos[1]

[1]School of Biochemistry and Molecular Genetics
University of New South Wales
P.O. Box 1,
Kensington NSW 2033, Australia

[2]Department of Biochemistry and Cell Biology
Rice University
P.O. Box 1892
Houston TX 77251, USA

Molecular motions play important roles in many, and probably most, protein-nucleic acid interactions. Because biomolecules possess large numbers of rotatable bonds, complexation is generally associated with some degree of structural rearrangement within each partner—going as far as wholesale binding-driven folding events for some proteins (see Spolar and Record, 1994). Relatively large-scale conformational changes have also been observed for several nucleic acid targets: the helical axis of dsDNA can be bent quite severely for example (Schultz *et al.*, 1991), or new tertiary interactions such as base triples can be generated by complexation of RNAs (Puglisi *et al.*, 1993). The existence of these binding-induced conformational rearrangements invites two questions for any given nucleic acid. (a) Is the bound conformation newly created by complex formation or is an existing minor conformer selected? The Gorenstein laboratory has suggested that the strongest *lac* DNA operators possess the most flexible phosphodiester backbones (Botuyan *et al.*, 1993), which may be consistent with selection of an existing conformer. (b) How significant are the various energetic contributions to recognition? Entropic components must be important, but are relatively difficult to assign to individual mechanisms. In this context, experimental methods that

are able to monitor entropic properties of either biomolecular partner and/or solvent molecules have the potential to make important contributions to the development and validation of theoretical descriptions.

Although it is even more fundamental than the role of dynamics in protein-nucleic acid recognition, our understanding of free nucleic acid dynamics remains relatively poor. Most of the detailed information in the literature derives from molecular dynamics (MD) simulations (see McCammon and Harvey, 1987), which have attained a general dominance over the relatively small quantity of available experimental data in the form of x-ray crystallographic temperature factors, and NMR, ESR, fluorescence anisotropy decay or depolarized dynamic light-scattering measurements. As an example, local mobilities within nucleotide units have long been thought to decrease in the order phosphodiester > (deoxy)ribose > base; however, recent measurements of $^{1}H^{1}H$ correlation times in DNA oligomers (Reid *et al.*, 1989) and solid state NMR data (Alam *et al.*, 1991) have begun to cast doubt on this notion. MD and molecular mechanics simulations are largely responsible for a longstanding view of sugar dynamics which posits rapid repuckering between 2'-endo and 3'-endo forms (McCammon and Harvey, 1987), particularly in RNAs. In contrast, Manoharan *et al.* (1987) have observed slow exchange behavior in a selectively ^{13}C-labeled oligonucleotide duplex, which might be assigned to repuckering at a very much longer timescale, and solid state NMR data on DNA oligomers (Huang *et al.*, 1990) provide no evidence for repuckering events. Clearly, there is need for a substantial body of experimental data that can be used to test the validity of MD simulations of nucleic acids, and perhaps help to improve them if they should prove deficient.

^{13}C relaxation studies should provide a useful experimental avenue into the examination of rapid motions in nucleic acids and their complexes. In general, these studies report on the motional behavior of ^{13}C-^{1}H bond vectors, providing information on the frequency and amplitude of motions in the ps-ns timescale (see, for example, Lipari and Szabo, 1982a,b). With the recent introduction of efficient methods for isotope labeling of synthetic RNAs in NMR quantities (Nikonowicz and Pardi, 1992; Batey *et al.*, 1992; Michnicka *et al.*, 1993), and ongoing progress in the production of ^{13}C-labeled DNA (Kainosho, this volume; Zimmer and Crothers, this volume; Michnicka *et al.*, unpublished), the way is now open to more fully probe the utility of ^{13}C relaxation measurements to describe the motional properties of nucleic acids in solution. As a first step in this process, we have surveyed the ^{13}C relaxation behavior of a large number of signals from a 29mer RNA. The results obtained and the experience gained here can be used to guide future studies of the relaxation properties of both RNAs and DNAs.

The HIV-1 TAR RNA Element

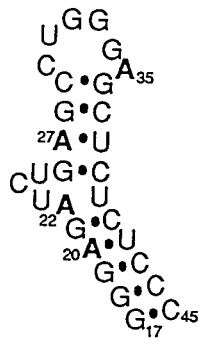
Our initial experimental subject has been an RNA fragment derived from the *trans*-activator response (TAR) element of HIV-1. A schematic structure of this species, known as ΔTAR, is shown in Fig. 1. The ΔTAR RNA element was selected for a number of reasons.

- It possesses three different kinds of structural features—duplex stem, internal bulge, and loop regions respectively, which might be expected to display different kinds of dynamic behavior.
- TAR has an important biological function in which conformational dynamics are known to be important (Puglisi *et al.*, 1993), and at least some of its cognate proteins have been identified and characterized, making examination of free and complexed RNA possible.
- ΔTAR contains a small number of any one nucleotide type (4 adenosines), so selective RNA labeling can be usefully employed for complete resolution of 2D $^{1}H,^{13}C$ spectra—a property that is essential for detailed dynamics studies but not always readily obtainable for any but the smallest RNAs (Michnicka *et al.*, 1993).

Methodology

The overall methodology for these studies consists of (1) preparation of ^{13}C-labeled RNA, (2) NMR data acquisition, and (3) motional modeling. These steps will be considered in turn.

Several different labeling strategies are available for ^{13}C relaxation analysis. The simplest is not to label at all, working at the ^{13}C natural abundance of 1.1 %, where relaxation analysis of small proteins has been quite successfully conducted (see, for example, Palmer *et al.*, 1991). However, given a typical working concentration of <2 mM for most RNAs, inverse-detected relaxation experiments at an effective ^{13}C concentration of 20 μM will be difficult and time-consuming, though not impossible. At the other end of the scale, uniform enrichment at 95 to 99% abundance can be employed. Uniform nucleotide enrichment offers the



HIV-1 ΔTAR RNA
29mer M_r = 9.4 kd

Fig. 1. Schematic structure of ΔTAR RNA, numbered for consistency with the wild-type TAR sequence. Watson-Crick basepairs with slowly exchanging imino proton resonances are indicated by (●).

potential for full coverage of all carbon sites in one set of experiments, but suffers from the presence of ^{13}C-^{13}C coupling, particularly in the ribose spin system. The peak multiplicities caused by ^{13}C-^{13}C coupling can be removed by the insertion of the minimal constant-time evolution period $T_c = 1/J_{CC}$ into pulse sequences (Santoro and King, 1992), producing a final signal intensity that is proportional to $2n \cdot e^{-T_c/T_2}$, where n is the number of carbons coupled to the site of interest. Our initial studies on ΔTAR RNA have employed uniform enrichment to make the resulting labeled molecule most useful for other kinds of NMR experiments, and constant-time methods to minimize the homonuclear coupling problem. A number of alternative labeling patterns is also available. Random fractional enrichment at all carbon sites can reduce the problems associated with ^{13}C homonuclear coupling, since signals without any homonuclear coupling appear at amplitudes approximately given by $A_S = F_R (1 - F_R)^n$, where F_R is the fractional level of random enrichment. At low enrichment levels, peaks attributable to the fraction of molecules that contain geminal ^{13}C sites appear at much lower levels than the singlet resonances. It appears that ~15% random enrichment may be optimal for relaxation analysis of moderate-sized RNAs. As with uniform labeling, random fractional enrichment permits all resolvable ^{13}C sites in a molecule to be examined simultaneously. Selective labeling of one carbon per nucleotide (or a small set of carbons that are not spin coupled) is another alternative that has the advantage of no ^{13}C-^{13}C coupling, but the disadvantage that a relatively small number of carbon sites can be examined using a single RNA sample. Nevertheless, it appears that species in which the ribose C1′ and base C6/C8 carbons are labeled will be very useful for the analysis of relative nucleotide mobilities (see below). Finally, ensemble labeling, which involves the preparation of a mixture of selectively labeled species, has the same qualities as selective labeling, with the disadvantage that several different labeled RNAs must be synthesized. It is expected that all available ^{13}C labeling strategies will find application in nucleic acid relaxation experiments in the near future.

Several two-dimensional pulse sequences have been described for the acquisition of heteronuclear relaxation data in protein systems (see, for example, Kay *et al.*, 1989; Palmer *et al.*, 1991; Stone *et al.*, 1992; Nicholson *et al.*, 1992; Peng and Wagner, 1992). We have employed constant-time variants of the experiments of Nicholson *et al.* (1992) for our studies of uniformly labeled RNAs. This family of sequences was selected for its simplicity: the experiments begin by generating single-quantum ^{13}C magnetization with one pulse at a time when ^1H magnetization is zero due to presaturation. The constant-time modification for the T_1 and NOE experiments involves the addition of a single 180° ^{13}C pulse. The T_2 experiment used here differs from its nonconstant-time progenitor in that no effective spin-locking field is employed. Free precession during the constant-time period allows the

relaxation decay to be fit to an oscillating function that is modulated by the homonuclear ^{13}C-^{13}C coupling. The effect of cross relaxation between adjacent dipolar interactions and between dipolar and chemical shift anisotropy (CSA) relaxation terms is minimized with a series of hard ^{1}H pulses or WALTZ-16 decoupling during the relaxation period.

Experimental Relaxation Times for the ΔTAR Element

Relaxation data have been collected for two forms of ΔTAR. Uniformly-labeled RNA was used to examine the behavior of the 27 resolved ribose C1′ and 28 base C6/C8 signals, giving almost complete coverage of the 29 nucleotide residues of ΔTAR. Signal overlap in 2D ^{1}H-^{13}C spectra of this species makes it very difficult to monitor many of the other carbon sites. Experimental T_1 values for the uniformly labeled RNA range from 298 to 1121 ms, with a 446 ms mean. T_2 values range from 15.3 to 51.7 ms (corresponding to linewidths of 6 to 21 Hz), with a mean of 33.6 ms (9.5 Hz). $\{^{1}H\}^{13}C$ NOEs vary from 0.10 to 0.62, with a mean of 0.23. All sets of experiments were repeated at least three times, yielding average experiment-to-experiment deviations for the T_1, T_2, and NOE measurements of 9%, 7%, and 30%, respectively. The poor average deviations of the NOE data largely result from the fact that they were conducted at a time when exchange of purine 8H signals with solvent deuterium had become significant. Similar NOE experiments on selectively labeled RNA yield standard deviations of 10 to 15%. The experimental data are summarized in a normalized form in Fig. 2, where their variation as a function of nucleotide sequence is readily apparent.

Selective labeling of only the adenosine residues of ΔTAR provides sufficient spectral resolution for all 28 proton-bound carbon sites of the 4 adenosine residues to be analyzed. In this case, the range of nT_1 values is 352 to 540 ms, with a 437-ms mean; T_2 values range from 20.7 to 77.2 ms, with a 40.1-ms (7.9-Hz) mean; and NOEs ranged from 0.02 to 0.62, with a mean of 0.25. For several repetitions of these experiments, average standard deviations of measured nT_1, T_2, and NOE values were 9, 7, and 15%, respectively.

Motional Modeling for ΔTAR

Motional analysis can follow several courses—the most common being the simultaneous fit of T_1, T_2, and NOE data to the familiar relaxation expressions (Lipari and Szabo, 1982a,b). A wide range of spectral density models can be employed. Our initial work has employed the widely used order parameter formalism of Lipari and Szabo (1982a,b), which generates values for the overall molecular rotational correlation time τ_m, and a second correlation time for fast internal motions τ_e, with its associated order parameter S^2, which describes the spatial restriction of motion. The least-squares fit to determine the optimum values of τ_m, τ_e, and S^2 is performed by means of a minimization procedure that samples

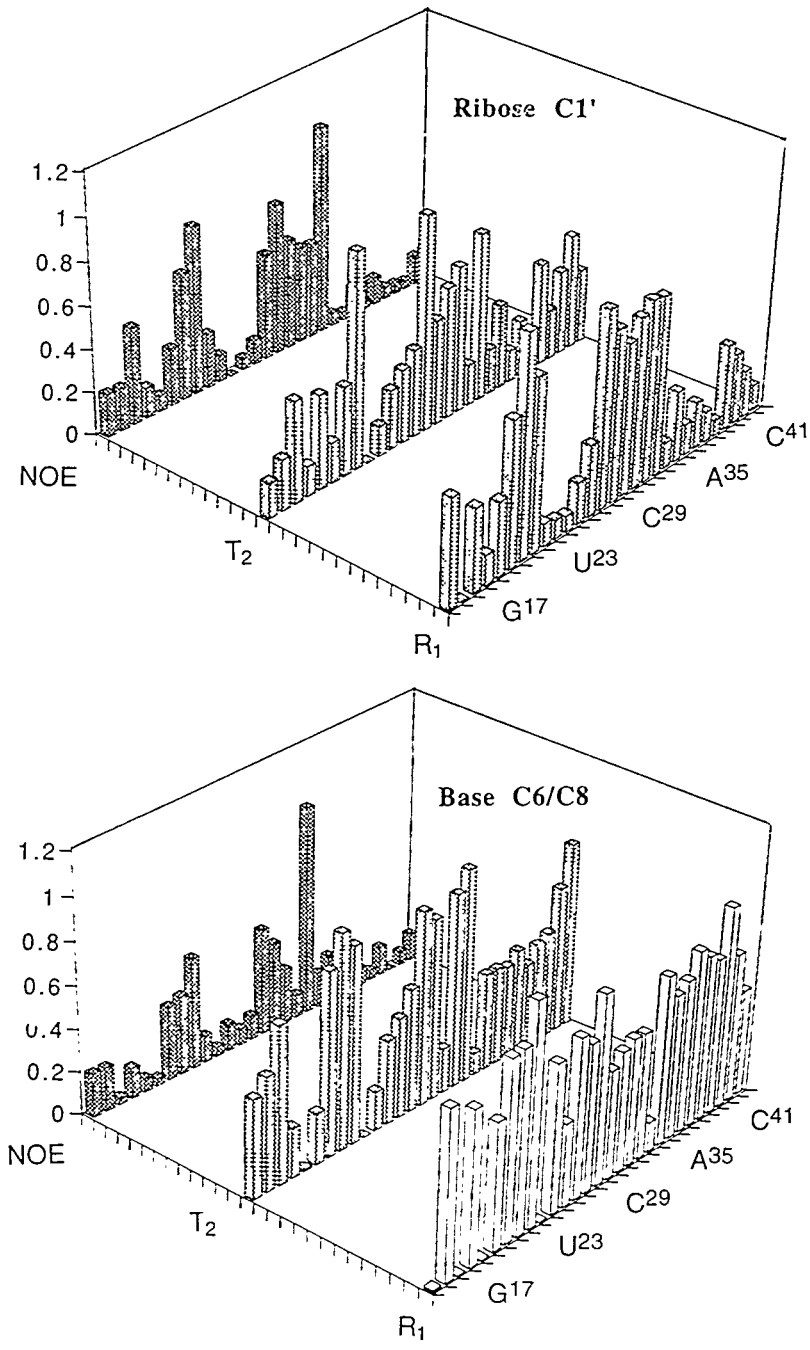


Fig. 2. Relaxation data for the ribose C1' and base C6/C8 signals of uniformly labeled ΔTAR, normalized to occupy values between 0 and 1.0. The R_1 ($1/T_1$), T_2, and NOE data have been plotted to enable direct comparison of different relaxation times.

the surface of a χ^2 error function describing the difference between experimental and calculated relaxation parameters. An example of the shape of the χ^2 error surface for the C4′ resonance of A^{35} is shown in Fig. 3. In addition to the original order parameter model, we have also employed a modification in which the internal correlation time is decomposed into fast and slow components, τ_f and τ_s with associated order parameters S_f^2 and S_s^2, where $S^2 = S_f^2 \bullet S_s^2$ (Clore *et al.*, 1990). As expected, addition of an extra parameter generally results in an improved fit, although order parameters determined with the two models do not differ by large amounts.

Depending on the RNA sample, the spectral density model, and the mode of analysis employed, best-fit overall correlation times τ_m range from 3.5 to 6.1 ns. As might be expected for an RNA that contains several different structural features, a wide distribution of order parameters is obtained, ranging from 0.35 to 1.0. Order parameters that represent an average value obtained for the two major models are plotted in three-dimensional histogram form in Fig. 4.

Behavior of Nucleotide Moieties

Information on the motional behavior of nucleotide moieties in RNAs emerges at several levels. On the question of the

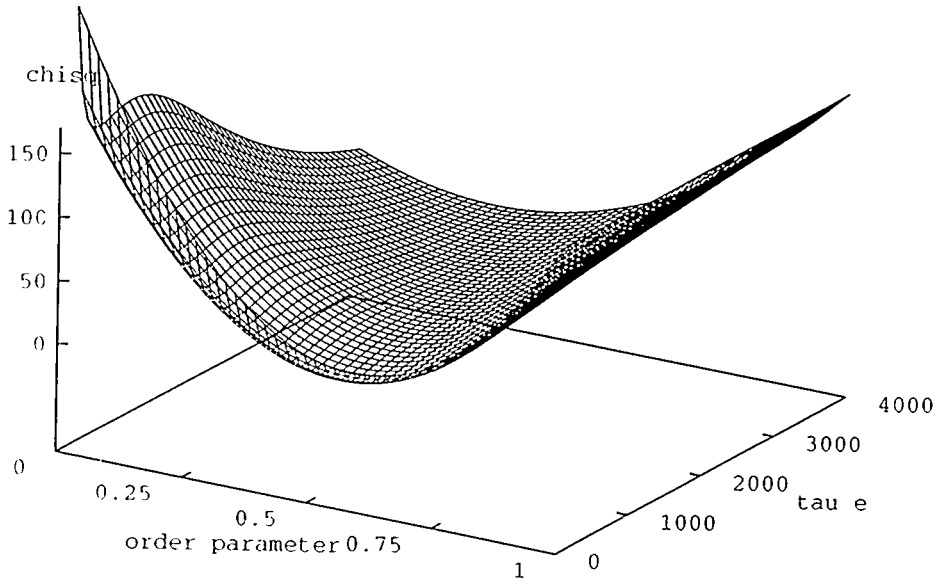


Fig. 3. The χ^2 error surface for a fit of the Lipari-Szabo model to experimental relaxation data for the C4′ signal of ΔTAR residue A^{35}.

relative mobilities of the ribose and base components, the average relaxation times determined for uniformly labeled ΔTAR are instructive. The average ribose and (base) relaxation times are T_1: 461(427) +/- 28(54) ms; T_2: 35.8(31.5) +/- 2.3(2.3) ms; NOE: 0.24(0.21) +/- 0.08(0.07), which are very nearly same within experimental error; this suggests that ribose and base mobilities are not substantially different. This proposal is supported by the order parameters determined for both uniformly and selectively labeled ΔTAR, which have similar global averages for the ribose and base moieties respectively. At a greater level of detail, order parameters determined for adenosine-labeled ΔTAR (Fig. 4) reveal a generally consistent trend for each nucleotide, where mobilities group into three classes: C3′, C1′, C2, C8 < C5′ < C2′, C4′. The consistently greater mobility of the ribose C2′ and C4′ atoms is striking. An immediate explanation for this result is not apparent, but may reflect the fact that C2′ and C4′ are the only ribose carbons not further tethered to a phosphodiester or base moiety, permitting the C-H bond vectors to wobble more freely. Our relaxation data provide no evidence for sugar repuckering events. From coupling constant analysis, it is known that the riboses of A^{20} and A^{27} have close-to-pure N-type pucker, whereas residues A^{22} and A^{35} are ~30 and 70% S-type, respectively. Repuckering on the appropriate timescale might be

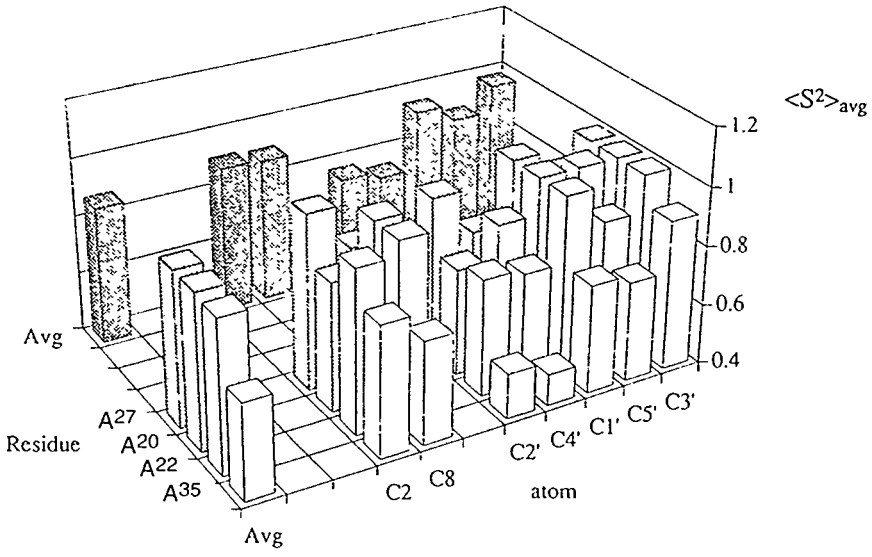


Fig. 4. Average order parameters determined for each ^{13}C atom of the four adenosine residues from selectively labeled ΔTAR. Average values for each residue and atom class are shown on the margins.

expected to manifest itself through mobilities that are sugar pucker-dependent. There are no obvious trends in Fig. 4.

Distribution of TAR Residue Mobilities

As the relaxation data of Fig. 2 and the order parameters of Fig. 4 suggest, the nucleotide residues of ΔTAR display a relatively wide range of mobilities. The least mobile residues occupy the duplex stem regions; however, the nucleotides of the pyrimidine-rich strand exhibit significantly greater mobility than those of the purine-rich strand. Residue A^{22}, which lies at the 5′ boundary of the bulge, shows a small but significant increase in mobility over A^{20} and A^{27}. This is consistent with the environment of this residue, which is stacked into the duplex but does not form a long-lived Watson-Crick basepair with U^{40}. The most mobile nucleotides occur in the bulge and loop regions, as might be expected. C^{24}, A^{35}, and C^{30} possess the lowest nucleotide S^2 values. It is interesting to note that none of these residues is essential for recognition by cognate proteins, hinting at the possibility that nucleotides with higher conformational entropy make less favorable binding partners and are therefore selected against. C^{30} and A^{35} appear to have higher mobilities than the nucleotides to either side—suggesting that these residues might act as a buffer between a somewhat more structured UGGG tetraloop and the duplex stem.

Conclusions

This initial survey of ^{13}C relaxation in the ΔTAR RNA element has generated a number of interesting results that should prove generally useful for future studies. The most readily comparable study in the literature is that of Schmidt *et al.* (1987), who monitored ^{13}C relaxation of the methyl groups from unusual bases in $tRNA^{Phe}$. Their study, which used T_1 and NOE data only, reported order parameters for the methyl group axis that ranged between 0.51 and 0.97—a range similar to that observed here. However, they reported a breakdown of the standard order parameter analysis at higher (118-MHz ^{13}C) frequencies, which should serve to emphasize the need for a thorough exploration of suitable motional models.

Future Prospects

NMR relaxation analysis of nucleic acids, already conducted for a number of years, is now poised for a period of major expansion and refinement, largely as a result of ongoing developments in isotope labeling, multidimensional pulse sequences, and to a lesser extent, theoretical descriptions. The next few years should see many studies directed towards a range of questions, including the relationships between local structure and dynamics, reconciling MD simulations with experimental data, correlating motions with measures of conformational entropy, and determining the significance of dynamics in various recognition processes. ^{13}C data will play a central role, although we expect that a great deal of useful information will be extracted from other nuclei, particularly ^{31}P.

Acknowledgements

The NMR studies described here were supported by the US National Institutes of Health (GM-42913) and the Welch Foundation (C-1166). The Rice 500-MHz NMR facility was established with the support of NIH grant RR-05759 and the W.M. Keck Foundation.

References

Alam, T.M., Orban, J., and Drobny, G.P. (1991) *Biochemistry 30*, 9229-9237.

Batey, R.T., 150, M., Kujawinski, E., Puglisi, J.D., and Williamson, J.R. (1992) *Nucleic Acids Res. 20*, 4515-4523.

Botuyan, M.V., Keire, D., Kroen, C., and Gorenstein, D.G. (1993) *Biochemistry 32*, 6863-6874.

Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C., and Gronenborn, A.M. (1990) *J. Am. Chem. Soc. 112*, 4989-4991.

Huang, W.-C., Orban, J., Kintanar, A., Reid, B.R., and Drobny, G.P. (1990) *J. Am. Chem. Soc. 112*, 9059-9068.

Kay, L.E., Torchia, D.A., and Bax, A. (1989) *Biochemistry 28*, 8972-8979.

Lipari, G. and Szabo, A. (1982a) *J. Am. Chem. Soc. 104*, 4546-4559.

Lipari, G. and Szabo, A. (1982b) *J. Am. Chem. Soc. 104*, 4559-4570.

Manoharan, S., Gerlt, J.A., Wilde, J.A., Withka, J.M., and Bolton, P.H. (1987) *J. Am. Chem. Soc. 109*, 7217-7219.

McCammon, J.A. and Harvey, S. (1987) *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press.

Michnicka, M.J., Harper, J.W., and King, G.C. (1993) *Biochemistry 32*, 395-400.

Nicholson, L.K., Kay, L.E., Baldisseri, D.M., Arango, J., Young, P.E., Bax, A., and Torchia, D.A. (1992) *Biochemistry 31*, 5253-5263.

Nikonowicz, E.P., Sirr, A., Legault, P., Jucker, F.M., Baer, L.M., and Pardi, A. (1992) *Nucleic Acids Res. 20*, 4507-4513.

Palmer, A.G., Rance, M., and Wright, P.E. (1991) *J. Am. Chem. Soc. 113*, 4371-4380.

Peng, J.W. and Wagner, G. (1992) *J. Magn. Reson. 98*, 308-332.

Puglisi, J.D., Chen, L., Frankel, A.D., and Williamson, J.R. (1993) *Proc. Natl. Acad. Sci. USA 90*, 3680-3684.

Reid, B.R., Banks, K., Flynn, P., and Nerdal, W. (1989) *Biochemistry 28*, 10001-10007.

Santoro, J. and King, G.C. (1992) *J. Magn. Reson. 97*, 202-207.

Schmidt, P.G., Sierzputowska-Gracz, H., and Agris, P.F. (1987) *Biochemistry 26*, 8529-8534.

Schultz, S.C., Shields, G.C., and Steitz, T.A. (1991) *Science 253*, 1001-1007.

Spolar, R.S. and Record, M.T. (1994) *Science 263*, 777-784.

Stone, M.J., Fairbrother, W.J., Palmer, A.G., Reizer, J., Saier, M.H., and Wright, P.E. (1992) *Biochemistry 31*, 4394-4406.

Selective ^{2}H and ^{13}C Labeling in NMR Analysis of Solution Protein Structure and Dynamics

David M. LeMaster

Northwestern University
2-100 Hogan Hall
2153 Sheridan Road
Evanston, IL 60208

Preparation of samples bearing combined isotope enrichment patterns has played a central role in the recent advances in NMR analysis of proteins in solution. In particular, uniform ^{13}C, ^{15}N enrichment has made it possible to apply heteronuclear multidimensional correlation experiments for the mainchain assignments of proteins larger than 30 kDa. In contrast, selective labeling approaches can offer advantages in terms of the directedness of the information provided, such as chirality and residue type assignments, as well as through enhancements in resolution and sensitivity that result from editing the spectral complexity, the relaxation pathways and the scalar coupling networks. In addition, the combination of selective ^{13}C and ^{2}H enrichment can greatly facilitate the determination of heteronuclear relaxation behavior.

Chiral β Deuteration of Amino Acids

Protein sidechain conformational analysis has provided an important stimulus for developing chiral isotopic labeling techniques for prochiral methyl and methylene groups. Biosynthetic incorporation of fractional ^{13}C enrichment yields differential one-bond ^{13}C-^{13}C scalar couplings, which provide chirality determinations for the gem-dimethyl groups of valine and leucine (Senn *et al.*, 1989; Neri *et al.*, 1989). In addition, chemical syntheses of the chirally deuterated methyl groups for both valine (Wilde *et al.*, 1988) and leucine (Ostler *et al.*, 1993) have been incorporated into proteins for assignment studies.

In a similar vein, although the vicinal $\alpha^{1}H$-$\beta^{1}H$ spin coupling constants provide the primary conformational information, the ambiguity in the prochiral identity

of the β methylene protons commonly precludes dihedral angle determination. Chiral β deuterated amino acids have been used to resolve stereochemical assignments in the determination of sidechain dihedral angles in free amino acids (Kainosho and Ajisaka, 1975; Kobayashi and Nagai, 1978), peptides (Fischman *et al.*, 1980; Kessler *et al.*, 1987), and proteins (LeMaster, 1987). The generality of this approach has been limited by inadequate availability of the chirally deuterated amino acid types. Comparatively straightforward procedures have been described for chiral labeling of the aromatic amino acids (Strange *et al.*, 1972; Kirby *et al.*, 1973, 1974; Battersby *et al.*, 1980; Matteson and Beedle, 1988), serine (Slieker and Benkovic, 1982; Aberhart and Russell, 1984; Ramer *et al.*, 1986), and aspartic acid (Field and Young, 1979; Rohm and Etten, 1985). Chiral β deuteration of the other common amino acids has been more problematic.

When probing enzymatic approaches to the stereoselective introduction of deuterium at the β position of the common amino acids, it was noted that only a small number of pyridoxal phosphate-dependent enzymes carry out reactions that activate at the γ carbon of amino acids through removal of a β hydrogen. In the case of cystathionine γ-synthase, it has been demonstrated that homoserine—closely related to the normal substrate O-succinyl homoserine—was chirally exchanged at the β position with a selectivity of approximately 100-fold (Posner and Flavin, 1972). As the *E. coli* cystathionine γ-synthase has recently been subcloned and over-expressed (Holbrook *et al.*, 1990), we chose to investigate use of this enzyme for preparative synthesis of chirally deuterated amino acids (Homer *et al.*, 1993).

Chiral exchange at the α hydrogen position can be carried out on the gram-scale in ^2H_2O for all of the standard amino acids. The α hydrogen exchange of glycine is highly stereoselective. Amino acids fall into four classes with respect to the selectivity of exchange at the β position: β′/β rate ~50 (arginine, glutamine, histidine, homoserine and lysine), β′/β rate ~8 (asparagine, glutamate, methionine, ornithine, and S-methyl cysteine), β′/β ~1 (leucine, aspartate, and alanine), and those seemingly too slow for large-scale exchange (isoleucine, phenylalanine, O-methyl tyrosine, serine, threonine, tryptophan, and valine). The stereochemistry of selective β exchange is the same for all cases studied.

Synthesis of ^{15}N, α–, β–Deuterated Amino Acids

The most widely used general procedure for synthesis of ^{15}N-labeled amino acids is based on the reaction of an ammonia equivalent with an α-halo ester. Ammonia itself is not generally used because of its tendency toward multiple alkylations. Instead, the potassium phthalimide salt is more commonly used, necessitating the preliminary conversion of ^{15}N ammonia to the phthalimide form (Ott, 1981). Two

procedures that use ^{15}N ammonia directly in the synthesis of amino acids are the Strecker reaction and reductive amination of the conjugate α-keto acid (Borch *et al.*, 1971). In both cases, it is generally impractical to obtain complete conversion in the presence of a single equivalent of the enriched ammonia.

Each of the procedures mentioned above suffers from the obvious limitation that they generally produce a racemic mixture, whereas, most commonly, it is only the L-form that is desired. Enzymatic catalysis offers an obvious alternative for chiral synthesis of standard biological metabolites. Greenaway and Whatley (1975) demonstrated earlier that when using a glucose 6-phosphate dehydrogenase system to regenerate the required redox cofactor NADPH, ^{15}N-labeled L-glutamate can be produced in gram amounts by employing glutamate dehydrogenase. Unfortunately, glutamate dehydrogenase is highly selective for its α-ketoglutarate substrate. The specific activity of its reaction with other cognate α-keto acids is generally 1% or less of that seen for the normal substrate (Struck and Sizer, 1960).

Seven of the standard amino acids have as the last step of their biosynthesis a glutamate-dependent transamination of the corresponding α-keto acid. Genetic studies in *E. coli* (Gelfand and Steinberg, 1977; Berg *et al.*, 1988) have demonstrated that one of these glutamate-dependent aminotransferases (the branched chain amino acid aminotransferase with genetic locus ilvE) is capable of catalyzing the biologically required amount of synthesis for most of these amino acids. Using an earlier published characterization of this gene (Lopes and Lawther, 1986), we subcloned ilvE into the T7 RNA polymerase expression vector pGEM2 and purified the induced aminotransferase protein to examine its utility in a coupled system with the glutamate dehydrogenase.

Gram amounts of various ^{15}N-enriched L-amino acids have been synthesized by using a coupled enzymatic system. Catalytic amounts of ^{15}N-labeled L-glutamate are generated using $(^{15}NH_4)_2SO_4$, α-ketoglutarate, and glutamate dehydrogenase. The labeled glutamate, in turn, serves as an amine donor to an appropriate α-keto acid using the *E. coli* branched chain amino acid aminotransferase. Because the aminotransferase catalyzes exchange of the α-hydrogen, carrying out this reaction in $^{2}H_2O$ gives rise to [2-^{2}H, 2-^{15}N]amino acids. Deuteration can be readily extended to the β position as well by pre-exchanging the α-keto acids in basic $^{2}H_2O$. The isotopically labeled amino acids are recovered in yields of 70 to 80%.

Enantiomeric Conversion of Racemic Amino Acid Mixtures

Stereoselective procedures for synthesizing amino acids have been developed that are suitable for the chiral introduction of each of the four groups attached to the α carbon. These include rhodium-chiral phosphine catalyzed hydrogenation of α,β-dehydroamino acid derivatives (Knowles *et al.*, 1975), alkylation of

carbanions formed from glycine-chiral Shiff base complexes (Seebach *et al.*, 1983), Strecker reactions on Schiff bases formed between aldehydes and chiral amines (Harada and Okawara, 1973), and electrophilic amination of chiral enolates (Evans *et al.*, 1986; Oppolzer and Tamura, 1990). For many applications, these chiral syntheses offer a clear advantage over racemic procedures only when the enantiomeric purity obtained is sufficiently high to obviate the need for an often-laborious resolution procedure. On the other hand, racemic syntheses generally use simpler and more readily available starting materials. This is particularly relevant for isotopic labeling syntheses because the initial form of the isotope often strongly circumscribes the practical synthetic pathways. Development of a straightforward means to convert the undesired (usually D-form) to the desired (usually L-form) enantiomer would serve to overcome the dominant limitation of racemic amino acid synthetic procedures. We discuss a one-pot enantiomeric conversion system in which the intermediate α-oxo acid, generated by means of D-amino acid oxidase, is converted *in situ* to the corresponding L-form amino acid.

Several studies have generated L-amino acids from the corresponding α-oxo acids either directly or by a coupled system using NAD(P)H-dependent glutamate Greenaway and Watley, 1975, 1977), malate (Baldwin *et al.*, 1987), or leucine plus formate (Nakajima *et al.*, 1990) dehydrogenases. These redox-coupled systems exhibit a limited substrate range. More importantly, they are restricted by the chemical instability of NAD(P)H, particularly in the presence of oxygen. With the exception of a 100 μmol enantiomeric conversion of D,L methionine using the leucine dehydrogenase (Nakajima *et al.*, 1990), these redox enzyme procedures have made use of the independently generated α-oxo acids.

We have examined the comparatively oxygen insensitive *E. coli* branched-chain amino acid aminotransferase as a means of converting the intermediate α-oxo acids to the L-form amino acids. The initial biochemical studies (Inoue *et al.*, 1988; Chanatry *et al.*, 1993; LeMaster, personal observation) suggest the specificity of the aminotransferase is nearly as broad as that of the widely studied oxidase (Greenstein *et al.*, 1953).

Having circumvented the dehydrogenase-driven reaction, we required an alternate means to shift the equilibrium toward the desired L-amino acid. The simple expedient of adding an excess of the physiological amine donor L-glutamate was found effective. Monosodium glutamate is soluble to above 0.5 M. Negligible inhibition of either the oxidase or transaminase reaction was observed under these conditions. The negative charge on the sidechain makes the separation of glutamate from the neutral amino acids straightforward through ion exchange chromatography.

Enantiomeric excesses of ~99% were obtained for several of the standard amino acids. The amount of oxidase can

be increased when higher enantiomeric purity is required. With the exception of methionine, the recovered yields for the three day protocol are ~90%. In this case, instability of the intermediate α-oxo acid probably contributed to the reduced methionine yield. We conducted a 1-day methionine experiment using a threefold increase in the amount of oxidase and aminotransferase and obtained a 90% yield.

This coupled oxidase-aminotransferase system is quite robust in the consistency of its high product yields. The enantiomeric conversion system is amenable to substantially larger scale reactions than have been reported here. Porcine D-amino acid oxidase is already widely used commercially, and the amount of *E. coli* branched-chain amino acid aminotransferase used in the 10 millimole valine reaction represents <0.05% of that obtained from a standard 10-ℓ fermenter culture. It should be noted that only one equivalent of L-glutamate is consumed for each equivalent of D-amino acid converted. The chromatographic isolation procedure allows for direct recycling of the remaining L-glutamate, as warranted.

One useful application of the enzymatic enantiomeric conversion procedure is in combination with the wide range of racemic syntheses developed for production of ^{13}C-labeled amino acids. As was illustrated, a Strecker synthesis of isobutyraldehyde, $K^{13}CN$, and ammonia was used to produce D,L-[1-^{13}C]valine in 83% yield. This racemic mixture was then converted to the L-form through a D-amino acid oxidase-branched amino acid aminotransferase system for an overall yield of 75%, based on $K^{13}CN$ (Polach *et al.*, 1993). Given the generality of the Strecker synthesis as well as the broad substrate specificity of the oxidase and aminotransferase enzymes, it is anticipated that this procedure will be comparably efficient for most standard aliphatic and aromatic amino acids.

Deuterium Enrichment in the Enhancement of Resolution and Sensitivity in Heteronuclear Correlation

To date, the vast majority of heteronuclear correlation experiments have made use of single or multiquantum free precession techniques. Several detailed comparisons of the HSQC and HMQC experiments (Norwood *et al.*, 1989, 1990; Bax *et al.*, 1990) have discussed effects such as differential relaxation rates for the single and multiquantum coherences, ^{1}H-^{1}H scalar relaxation and, in the case of the HMQC experiment, ^{1}H-^{1}H scalar modulation of the heteronuclear frequency. The geminal ^{1}H-^{1}H interactions present in methylene positions give rise to decidedly inferior performance compared to that of the IS systems considered in earlier studies.

By eliminating the geminal ^{1}H-^{1}H scalar and dipolar interactions, deuteration offers an exceedingly effective means of enhancing both the resolution and sensitivity of the methylene resonances for these correlation experiments, as illustrated in Fig. 1 (Kushlan and LeMaster, 1993a). Panel A presents an expansion of

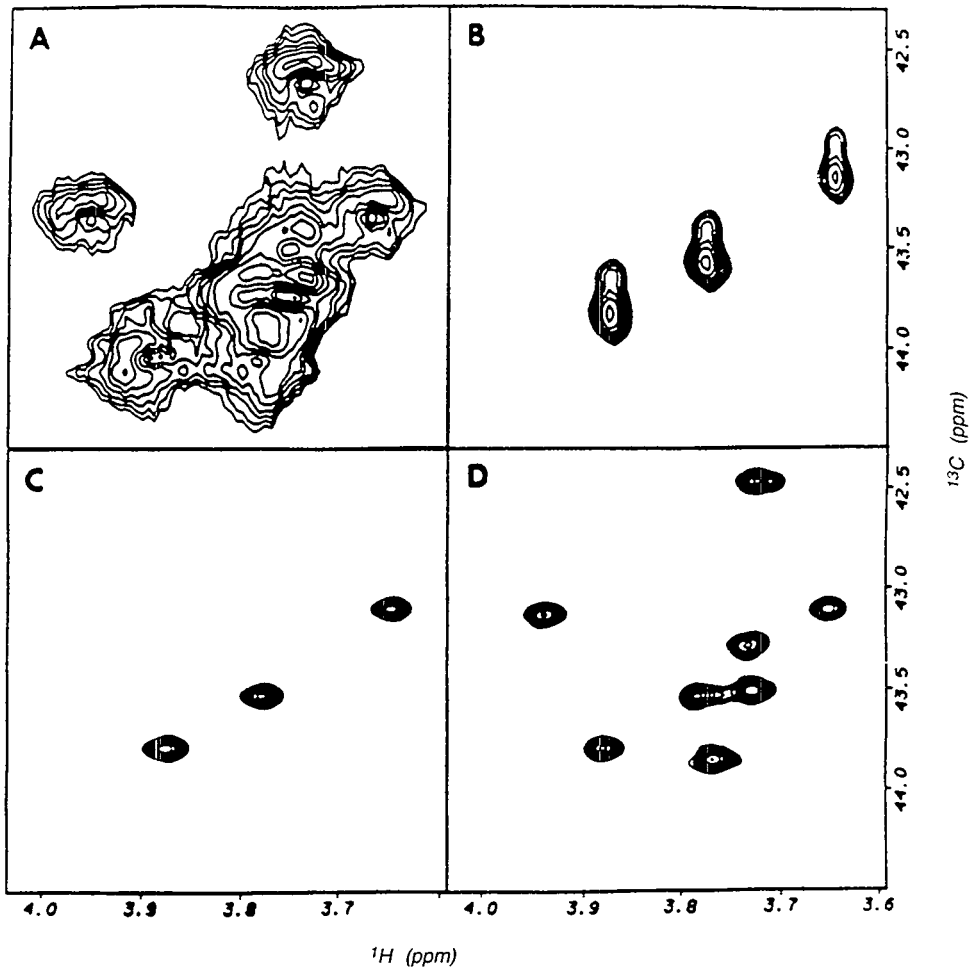


Fig. 1. The ^{1}H-^{13}C HMQC spectra as a function of deuteration, ^{2}H decoupling, and multiplet editing.
(a) Expansion of the HMQC spectrum of [2-^{13}C]glycine-labeled E. coli *thioredoxin.*
(b) Corresponding HMQC spectrum for the [2-$^{2}H_{R}$,2-^{13}C]glycine-labeled protein. [2-^{13}C]Glycine (Cambridge Isotope Laboratory) chirally deuterated by cystathionine γ-synthase catalyzed exchange (Homer et al., *1993).*
(c) Collection as for (b) with ^{2}H decoupling applied from initial ^{1}H pulse to start of acquisition. Gated ^{2}H lock system from Bruker Instruments, Inc.
(d) DEPT-HMQC spectrum of [50% U-^{2}H,2-^{13}C]glycine-labeled E. coli *thioredoxin with ^{2}H decoupling. Edit pulse angle set to 45°.*
[Reprinted with permission from D. M. Kushlan and D. M. LeMaster (1993) J. Biomolec. NMR 3, *701-708; Copyright 1993 by ESCOM Science Publishers B.V.]*

the most overlapped portion of the HMQC spectrum of [2-^{13}C]glycine-labeled *E. coli* thioredoxin (11.7 kDa). Panel B contains the corresponding HMQC spectrum of [2-$^{2}H_R$,2-^{13}C]glycine-labeled *E. coli* thioredoxin. Elimination of the ^{1}H-^{1}H geminal scalar and dipolar interactions yield a significant improvement in resolution for the ^{1}H dimension.

In contrast, the resolution in the ^{13}C dimension has not been significantly improved because of the presence of ^{2}H-^{13}C scalar mediated interactions. Recently, Bax and coworkers (Grzesiek *et al.*, 1993) described a high-power ^{2}H gated decoupling system and applied it to multidimensional heteronuclear experiments on the random fractionally deuterated, uniform ^{13}C-enriched calcineurin B protein. In this experiment, the $^{13}C\alpha$ positions of interest were effectively perdeuterated. Hence, although the ^{13}C T_2 times can be significantly increased, it is at the sacrifice of a potential ^{1}H spectral dimension. The HMQC spectrum in panel C illustrates that by applying high-power ^{2}H decoupling for the [2-$^{2}H_R$,2-^{13}C]glycine-labeled sample, it is possible to gain significant improvement in the resolution of the ^{13}C dimension without sacrificing ^{1}H detection.

Because stereoselective ^{2}H labeling is often difficult to achieve, it is desirable to apply these experiments to the more readily accessible random fractionally deuterated protein samples (Grzesiek *et al.*, 1993; Kalbitzer *et al.*, 1985; LeMaster and Richards, 1988; Wang *et al.*, 1990). However, such an experiment requires suppression of the signals arising from the diprotio components, which can be achieved by spectral editing through a DEPT-HMQC sequence (Kessler *et al.*, 1989). Panel D shows the ^{2}H decoupled DEPT-HMQC spectrum of [2-^{2}H,2-^{13}C]-glycine-labeled *E. coli* thioredoxin by using an editing pulse angle of 45° to eliminate the I_2S component.

Comparison of panels A and C indicates an approximately threefold gain in resolution for both the ^{1}H and ^{13}C dimensions of this 11.7-kDa protein. This enhanced resolution results in a corresponding ninefold increase in sensitivity. The analogous ^{2}H decoupled DEPT-HMQC spectrum of the [2-^{2}H,2-^{13}C]glycine sample in panel D exhibits a twofold greater sensitivity than that of the [2-^{13}C]glycine sample in panel A, even though for the [2-^{2}H,2-^{13}C]glycine sample each crosspeak arises from only one-quarter of the total number of molecules.

This ^{2}H enrichment and decoupling approach can be applied to the HSQC experiment as well. As a result of the superior ^{13}C resolution offered by the HSQC experiment, compared to the HMQC experiment, the corresponding enhancements in sensitivity and resolution provided by deuteration combined with ^{2}H decoupling are reduced twofold.

Differential Deuterium Isotope Shifts in the Conformational Analysis of Protein Glycine Residues

Using a mixture of chirally deuterated [2-^{13}C]glycine labeled samples for both *E. coli* thioredoxin and staphylococcal nuclease, we found that the ^{13}C chemical shifts for the pro-*R* and pro-*S* resonances of the same residue vary by as much as 70 ppb (11 Hz at 14.1 T) (LeMaster *et al.*, 1994). A similar variation in the $^1J_{\alpha H,C\alpha}$ values in peptides had been previously observed (Egli and Philipsborn, 1981). More recently, Bax and coworkers (Vuister *et al.*, 1992, 1993) and Kessler and coworkers (Mierke *et al.*, 1992) have used isotopically enriched protein samples to demonstrate the predicted trigonometric dependence of $^1J_{\alpha H,C\alpha}$ on the mainchain dihedral angles. Because of the lack of stereochemical assignments, these studies did not include data from the glycine residues.

Using chirally deuterated glycine samples, we have verified that both the pro-*R* and pro-*S* glycine crosspeaks fit a trigonometric dependence quite similar to that reported by Bax and coworkers (Vuister *et al.*, 1992). Because both the pro-*R* and pro-*S* protons fit the same parameters, all of the structurally useful information on (ϕ,ψ) is contained in the differential $^1J_{\alpha H,C\alpha}$ values. An analogous trigonometric parameterization was used to fit the differential deuterium isotope shift data. The contour plot of this trigonometric fit function is given in Fig. 2. A regression analysis of the experimental *vs* predicted differential isotope shifts yields an r^2 value of 0.85.

The differential isotope shift and coupling constant values both exhibit opposite signs for the $\cos^2\phi$ and $\cos^2\psi$ dependencies as well as a threefold dominance of the $\cos^2\psi$ term. On the other hand, the $^1J_{\alpha H,C\alpha}$ values vary over a range of only 10% of the average value (that is, 140 Hz), whereas the relative variation in the differential isotope shifts is nearly 30% of the average value (that is, 250 ppb).

As illustrated in Fig. 2, judging from 37 of the highest resolution x-ray structures (Bernstein *et al.*, 1977), almost 60% of all glycine residues have dihedral angles lying in the α-helix-to-bridge regions of the (ϕ,ψ) conformational map. A linear least squares analysis to these elongated distributions in the left-handed and right-handed regions yields an r.m.s.d. fit of 11.3°. This fit line lies approximately perpendicular to the gradient of the trigonometric fit functions for both the differential isotope shifts and the $^1J_{\alpha H,C\alpha}$ values (Vuister *et al.*, 1992). Hence, if independent structural information can discriminate between a glycine conformation in the extended regions *vs* the helix-to-bridge regions, the combination of differential isotope shifts and ^1J values should provide a useful estimate of the corresponding (ϕ,ψ) value along this best fit line. Indeed, a least squares parameterization on the differential isotope shifts and coupling constants for the 11 glycines of *E. coli* thioredoxin and staphylococcal nuclease in this conformational region yields an r.m.s.d. of 6.5° between the best fit line and the observed x-ray (ϕ,ψ) values.

The analysis given above presupposes the known stereochemical assignment of the deuterium labeling. This logic can be reversed for the 16 glycines with non-degenerate methylenes in these two proteins. If the sign of ϕ is independently known, the chirality can be determined from the following protocol:

- If $|\Delta^1 J_{\alpha H,C\alpha}|$ is > 6 Hz, the pro-*R* hydrogen has the larger coupling for $\phi < 0$ and the smaller coupling for $\phi > 0$.
- Otherwise, the pro-*R* hydrogen is the most upfield-shifted resonance for $\phi < 0$ and most downfield shifted for $\phi > 0$.

Although the glycine residues of these two proteins represent only a modest subset of the observed glycine conformations given in Fig. 2, these results suggest that reliable stereochemical assignments can be obtained from the isotope shift and scalar coupling data with only fairly limited additional structural information.

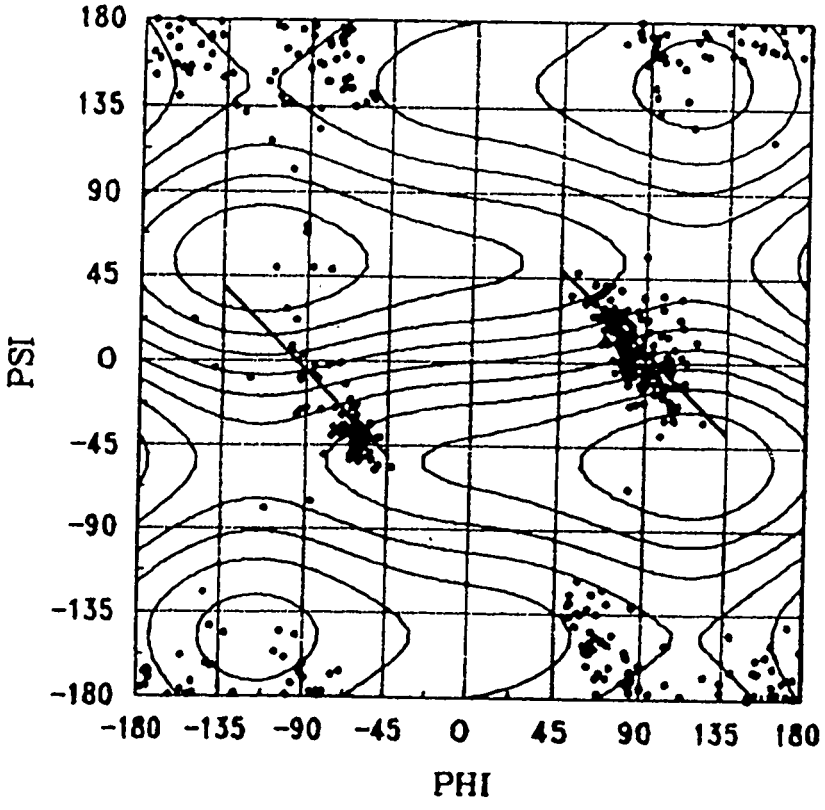


Fig. 2. Contour plot illustrating the predicted dihedral angle dependence of the differential one-bond deuterium isotope shift for the glycine Cα resonances. The contour interval is 20 ppb with the zero level contour passing through the origin. Also plotted are the (ϕ,ψ) values for 498 glycine residues observed in 37 high-resolution x-ray structures. Diagonal lines illustrate the best fit to the 281 glycine residues found in the helix-to-bridge region of the conformational plane (that is, -90° < ψ < 90°). [Reprinted with permission from D. M. LeMaster et al. *(1994)* J. Biomolec. NMR 4, *(Copyright 1994 by ESCOM Science Publishers, B.V.)]*

^{1}H-Detected NMR Relaxation of Methylene Carbons Through Stereoselective and Random Fractional Deuteration

With the availability of two-dimensional ^{1}H-detected ^{13}C relaxation experiments for methine (Nirmala and Wagner, 1988; Kay *et al.*, 1989) and methyl (Palmer *et al.*, 1991a; Kay *et al.*, 1992a) spin systems, the lack of a corresponding technique for measuring the dynamics of methylene carbons represents one of the major impediments to carrying out complete relaxation analyses of macromolecules. The cross-correlation effects between ^{1}H-^{13}C dipolar interactions (Werbelow and Grant, 1977; Vold and Vold, 1978) can not be as readily suppressed for methylene resonances as they can for the symmetric I_3S methyl system (Palmer *et al.*, 1991a; Kay *et al.*, 1992a). We have studied the use of deuterium enrichment as a means of converting a methylene spin system into a pair of independent ^{1}H-^{13}C and ^{2}H-^{13}C dipolar systems for relaxation analysis.

For the AIS system of monodeuterated methylene resonances, ^{2}H decoupling suppresses cross-correlation between ^{1}H-^{13}C and ^{2}H-^{13}C dipoles. As a result, for both longitudinal (Werbelow and Grant, 1977; Fagerness *et al.*, 1975) and transverse (Vold and Vold, 1976, 1978) ^{13}C relaxation, the ^{1}H-^{13}C dipole can be treated as an IS system that has a small ^{13}C{^{2}H} NOE perturbation of the steady state magnetization. Relaxation measurements were carried out on both [^2H_R, 2-^{13}C]glycine- and [50%U-^{2}H, 2-^{13}C]-glycine-labeled *E. coli* thioredoxin (Kushlan and LeMaster, 1993b). The previously published pulse sequence for measuring T_2 (Kay *et al.*, 1992b) required slight modifications to achieve efficient suppression of the diprotio signal in the latter sample.

Because chemical shift anisotropy contributions are negligible at this field strength (Palmer *et al.*, 1991b) with the decoupling field suppression of the ^{2}H scalar interaction, relaxation is expected to be purely dipolar. To interpret the relaxation rates ($R_1 = 1/T_1$ and $R_2 = 1/T_2$), it is necessary to partition the individual contributions of the ^{1}H-^{13}C and ^{2}H-^{13}C dipoles. In the extreme narrowing limit, both $R_1(^2H\text{-}^{13}C)$ / $R_1(^1H\text{-}^{13}C)$ and $R_2(^2H\text{-}^{13}C)$ / $R_2(^1H\text{-}^{13}C)$ are 0.063, as determined by the $S(S+1)\gamma^2$ factors in the relaxation formula for a single exponential correlation function.

$$R_1(^2H\text{-}^{13}C) / R_1(^1H\text{-}^{13}C) = \frac{1(2)\gamma_{^2H}^2 [J(\omega_{^2H}-\omega_{^{13}C}) + 3J(\omega_{^{13}C}) + 6J(\omega_{^2H}+\omega_{^{13}C})]}{1/2(3/2)\gamma_{^1H}^2 [J(\omega_{^1H}-\omega_{^{13}C}) + 3J(\omega_{^{13}C}) + 6J(\omega_{^1H}+\omega_{^{13}C}}$$

The expression for the R_2 ratio is analogous with the additional spectral density functions $4J(0) + 6J(\omega_H)$. The R_1 ratio increases for correlation times near the various Larmor frequencies and approaches 0.223 in the slow tumbling limit. In contrast, the R_2 ratio reaches a maximum of 0.127 at ~1-ns correlation time and then decreases again to 0.063 in

the slow tumbling limit. It should be noted that for correlation times >2 ns, the estimated fraction of R_1 and R_2 attributable to the ^{1}H-^{13}C dipole varies less than 5%.

In conclusion, accurate ^{13}C relaxation data can be obtained for methylene positions through ^{1}H detection by using suitable deuteration. The practical utility of applying this approach to protein ^{13}C relaxation analysis will generally be dependent on combination with a suitable carbon enrichment pattern. Homonuclear magnetization exchange and dipolar relaxation during the ^{13}C relaxation delay precludes the use of high levels of uniform ^{13}C enrichment with the relaxation pulse sequences described to date. However, a highly selective alternating carbon enrichment pattern can be obtained for each of the constituent amino acids in appropriate *E. coli* strains (LeMaster and Cronan, 1982). With the combination of these ^{2}H and ^{13}C labeling approaches, it should prove feasible to determine the relaxation behavior at virtually every site of a biological macromolecule.

References

Aberhart, D.J. and Russell, D.J. (1984) *J. Am. Chem. Soc. 106*, 4902-4906.

Baldwin, J.E., Dyer, R.L., Ng, S.C., Pratt, A.J., and Russell, M.A. (1987) *Tet. Lett. 28*, 3745-3746.

Battersby, A.R., Nicoletti, M., Staunton, J., and Vleggaar, R. (1980) *J. Chem. Soc. Perkin Trans. 1*, 43-51.

Bax, A., Ikura, M., Kay, L.E., Torchia, D.A., and Tschudin, R. (1990) *J. Magn. Reson. 86*, 304-318.

Berg, C.M., Wang, M.D., Vartak, N.B., and Liu, L. (1988) *Gene 65*, 195-202.

Bernstein, F. C., Koetzle, T.F., Williams, G.J.B., Meyer, J., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535-542.

Borch, R.F., Bernstein, M.D., and Durst, H.D. (1971) *J. Am. Chem. Soc. 93*, 2897-2904.

Chanatry, J.A., Schafer, P.H., Kim, M.S., and LeMaster, D.M. (1993) *Anal. Biochem. 213*, 147-151.

Egli, H. and Philipsborn, W.v. (1981) *Helv. Chim. Acta 64*, 976-988.

Evans, D.A., Britton, T.C., Dorow, R.L., and Dellaria, J.F. (1986) *J. Am. Chem. Soc. 108*, 6395-6397.

Fagerness, P.E., Grant, D.M., Kuhlmann, K.F., Mayne, C.L., and Parry, R.B. (1975) *J. Chem. Phys. 63*, 2524-2532.

Field, S.J. and Young, D.W. (1979) *J. Chem. Soc. Chem. Commun.* 1163-1165.

Fischman, A.J., Live, D.H., Wyssbrod, H.R., Agosta, W.C., and Cowburn, D. (1980) *J. Am. Chem. Soc. 102*, 2533-2539.

Gelfand, D.H. and Steinberg, R.A. (1977) *J. Bacteriol. 130*, 429-440.

Greenaway, W. and Whatley, F.R. (1975) *J. Labelled Cmpds. Radiopharm. 11*, 395-400.

Greenaway, W. and Whatley, F.R. (1977) *FEBS Lett. 75*, 41-43.

Greenstein, J.P., Birnbaum, S.M., and Otey, M.C. (1953) *J. Biol. Chem. 204*, 307-321.

Grzesiek, S., Anglister, J., Ren, H., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 4369-4370.

Harada, K. and Okawara, T. (1973) *J. Org. Chem. 38*, 707-710.

Holbrook, E. L., Greene, R. C., and Krueger, J.H. (1990) *Biochemistry 29*, 435-442.

Homer, R.J., Kim, M.S., and LeMaster, D.M. (1993) *Anal. Bioch. 215*, 211-215.

Inoue, K., Kuramitsu, S., Aki, K., Watanabe, Y., Takagi, T., Nishigai, M., Ikai, A., and Kagamiyama, H. (1988) *J. Biochem. 104*, 777-784.

Kainosho, M. and Ajisaka, K. (1975) *J. Am. Chem. Soc. 97*, 5630-5631.

Kalbitzer, H.R., Leberman, R., and Wittinghofer, A. (1985) *FEBS Lett. 180*, 40-42.

Kay, L.E., Torchia, D.A., and Bax, A .(1989) *Biochemistry 28*, 8972-8979.

Kay, L.E., Bull, T.E., Nicholson, L.K., Greisinger, C., Schwalbe, H., Bax, A., and Torchia, D.A. (1992a) *J. Magn. Reson. 100*, 538-558.

Kay, L.E., Nicholson, L.K., Delaglio, F., Bax, A., and Torchia, D.A. (1992b) *J. Magn. Reson. 97*, 359-375.

Kessler, H., Griesinger, C., and Wagner, K. (1987) *J. Am. Chem. Soc. 109*, 6927-6933.

Kessler, H., Schmieder, P., and Kurz, M. (1989) *J. Magn. Reson. 85*, 400-405.

Kirby, G.W. and Michael, J. (1973) *J. Chem. Soc., Perkin Trans. I* 115-120.

Kirby, G.W. and Varley, M.J. (1974) *J. Chem. Soc. Chem. Commun.* 833-834.

Knowles, W.S., Sabacky, M.J., Vineyard, B.D., and Weinkauff, D.J. (1975) *J. Am. Chem. Soc. 97*, 2567-2568.

Kobayashi, J. and Nagai, U. (1978) *Biopolymers 17*, 2265-2277.

Kushlan, D.M. and LeMaster, D.M. (1993a) *J. Biomolec. NMR 3*, 701-708.

Kushlan, D.M. and LeMaster, D.M. (1993b) *J. Am. Chem. Soc. 115*, 11026-11027.

LeMaster, D.M. and Cronan, J.E. (1982) *J. Biol. Chem. 257*, 1224-1230.

LeMaster, D.M. (1987) *FEBS Lett. 223*, 191-196.

LeMaster, D.M. and Richards, F.M. (1988) *Biochemistry 27*, 142–150.

LeMaster, D.M., LaIuppa, J.C., and Kushlan, D.M. (1994) *J. Biomolec. NMR. 4*, 863-870.

Lopes, J.M. and Lawther, R.P. (1986) *Nucleic Acids Res. 14*, 2779-2798.

Matteson, D.S. and Beedle, E.C. (1988) *J. Labelled Compd. Radiopharm. 25*, 675-683.

Mierke, D.F., Grdadolnik, S.G., and Kessler, H. (1992) *J. Am. Chem. Soc. 114*, 8283-8284.

Nakajima, N., Esaki, N., and Soda, K. (1990) *J. Chem. Soc. Chem. Commun.* 947-948.

Neri, D., Szyperski, T., Otting, G., Senn, H., and Wüthrich, K. (1989) *Biochemistry 28*, 7510-7516.

Nirmala, N.R. and Wagner, G. (1988) *J. Am. Chem. Soc. 110*, 7557-7558.

Norwood, T.J., Boyd, J., and Campbell, I.D. (1989) *FEBS Lett. 255*, 369-371.

Norwood, T.J., Boyd, J., Heritage, J.E., Soffe, N., and Campbell, I.D. (1990) *J. Magn. Reson. 87*, 488-501.

Oppolzer, W. and Tamura, O. (1990) *Tetrahedron Lett. 31*, 991-994.

Ostler, G., Soteriou, A., Moody, C.M., Khan, J.A., Birdsall, B., Carr, M.D., Young, D.W., and Feeney, J. (1993) *FEBS Lett. 318*, 177-180.

Ott, D.G. (1981) *Synthesis with Stable Isotopes*, Wiley, New York, pp. 111.

Palmer, A.G., Wright, P.E., and Rance, M. (1991a) *Chem. Phys. Lett. 185*, 41-46.

Palmer, A.G., Rance, M., and Wright, P.E. (1991b) *J. Am. Chem. Soc. 113*, 4371-4380.

Polach, K.J., Shah, S.A., LaIuppa, J.C., and LeMaster, D.M. (1993) *J. Labeled Compd. Radiopharm. 33*, 809-816.

Posner, B.I. and Flavin, M. (1972) *J. Biol. Chem. 247*, 6402-6411.

Ramer, S.E., Moore, R.N., and Vederas, J.C. (1986) *Can. J. Chem. 64*, 706-713.

Rohm, K.H. and Etten, R.L.V. (1985) *J. Labelld Compd. Radiopharm. 22*, 909-915.

Seebach, D., Boes, M., Naef, R., and Schweizer, W.B. (1983) *J. Am. Chem. Soc. 105*, 5390-5398.

Senn, H., Werner, B., Messerle, B.A., Weber, C., Traber, R., and Wüthrich, K. (1989) *FEBS Lett. 249*, 133-118.

Slieker, L. and Benkovic, S.J. (1982) *J. Labeled Compd. Radiopharm. 9*, 647-657.

Strange, P.G., Staunton, J., Wiltshire, H.R., Battersby, A.R., Hanson, K.R., and Havir, E.A. (1972) *J. Chem. Soc., Perkin Trans I*, 2364-2372.

Struck, J. and Sizer, I.W. (1960) *Arch. Biochem. Biophys. 86*, 260-266.

Vold, R.R. and Vold, R.L. (1976) *J. Chem. Phys. 64*, 320-332.

Vold, R.L. and Vold, R.R. (1978) *Prog. NMR Spectrosc. 12*, 79-133.

Vuister, G.W., Delaglio, F., and Bax, A. (1992) *J. Am. Chem. Soc. 115*, 9674-9675.

Vuister, G.W., Delaglio, F., and Bax, A. (1993) *J. Biomolec. NMR 3*, 67-80.

Wang, J., LeMaster, D.M., and Markley, J.L. (1990) *Biochemistry 29*, 88-101.

Werbelow, L.G. and Grant, D.M. (1977) *Adv. Magn. Reson. 9*, 189-299.

Wilde, J.A., Bolton, P.H., Dell'Acqua, M., Hibler, D.W., Pourmotabbed, T., and Gerlt, J.A. (1988) *Biochemistry. 27*, 4127-4132.

Magnetic Resonance Studies of Isotopically Labeled Paramagnetic Proteins: [2Fe-2S] Ferredoxins

Hong Cheng,[1] Bin Xia, Young Kee Chae, William M. Westler, and John L. Markley

Department of Biochemistry
College of Agricultural and Life Sciences
University of Wisconsin-Madison
420 Henry Mall
Madison, WI 53706

[1]Present address: Department of Pharmacology
Mayo Clinic and Research Foundation
7 Guggenheim
Rochester, MN 55905

Introduction

Recent developments in NMR spectroscopy, especially multidimensional, multinuclear NMR techniques, have made NMR the most versatile tool available for studying protein structure and function in solution. Unlike diamagnetic proteins, paramagnetic proteins contain centers with unpaired electrons. These unpaired electrons interact with magnetic nuclei either through *chemical bonds* by a contact mechanism or through *space* by a pseudocontact mechanism. Such interactions make the acquisition and analysis of NMR spectra of paramagnetic proteins more challenging than those of diamagnetic proteins. Some NMR signals from paramagnetic proteins are shifted outside the chemical shift region characteristic of diamagnetic proteins; these "hyperfine-shifted" resonances originate from nuclei that interact with unpaired electrons from the paramagnetic center. The large chemical shift dispersion in spectra of paramagnetic proteins makes it difficult to excite the entire spectral window and leads to distortions in the baseline. Interactions with paramagnetic centers shorten T_1 and T_2 relaxation times of nuclei; the consequences are line broadening and

lower spectral sensitivity. Scalar (through bond) and dipolar (through space) interactions between pairs of nuclei are what give rise to crosspeak signals in multidimensional NMR spectra of small diamagnetic proteins. When such interactions involve a nucleus that is strongly relaxed by interaction with a paramagnetic center, specialized methods may be needed for its detection or it may be completely undetectable by present nD NMR methods (Oh and Markley, 1990a, 1990b; Oh *et al.*, 1990, Skjeldal *et al.*, 1991a,b).

Success in overcoming difficulties in detecting and assigning NMR signals from nuclei that experience paramagnetic effects can be rewarding, however, because such resonances often contain important information about the active centers of the protein. The chemical shifts and relaxation times of these signals are sensitive to the delocalization of unpaired electron density from the metals onto nuclei. Because most paramagnetic proteins are involved in electron transfer, NMR data can provide information that is vital for understanding mechanisms of inter- and intramolecular electron transfer, such as the rates and equilibria of electron exchange reactions.

The early NMR studies of hyperfine-shifted resonances of iron-sulfur proteins focused on their detection and temperature-dependence (Poe *et al.*, 1971; Chan and Markley, 1983b; Oh and Markley, 1990b) and on characterizing the patterns of the hyperfine-shifted resonances from proteins containing different paramagnetic centers (Skjeldal *et al.*, 1991a). Interpretation of these early NMR studies was limited by the lack of assignments of the hyperfine-shifted resonances to specific atoms in individual amino acid residues.

Recently, two new strategies have been applied in NMR investigations of iron-sulfur proteins. The first is the detection of cross relaxation between pairs of nuclei where one or both experience hyperfine interactions; this is accomplished either by one-dimensional (1D) NOE or by two-dimensional (2D) NOESY recorded with a very short mixing time (Dugad *et al.*, 1990; Skjeldal *et al.*, 1991b; Bertini *et al.*, 1991; Cheng *et al.*, 1992). The other approach is to apply multinuclear NMR methods to paramagnetic proteins that have been enriched with NMR-active stable isotopes (Markley and Kainosho, 1993). Part of the rationale for isotopic labeling is to make use of the additional coupling pathways that are created. For example, single-bond *J*-couplings between ^{13}C and ^{1}H ($^{1}J_{CH}$ = ~120 Hz) and between ^{15}N and ^{1}H ($^{1}J_{NH}$ = ~90 Hz) are much larger than three-bond ^{1}H-^{1}H coupling ($^{3}J_{HH}$ = ~7 Hz); the former interactions are much better resolved in regions of a protein affected by paramagnetism. Another use of isotopic labeling is to simplify spectra so that peaks of interest can be resolved and assigned more readily. Finally, because hyperfine interactions are proportional to the square of the magnetogyric ratio, signals from nuclei such as ^{2}H, ^{13}C, and ^{15}N, which

have magnetogyric ratios (γ) that are smaller than that of ^1H, are less affected by paramagnetic centers. In this review, we present recent results from our laboratory that illustrate the use of isotopic labeling in NMR investigations of structure-function relationships in three recombinant [2Fe-2S] ferredoxins: the ferredoxin from the vegetative form of *Anabaena* 7120 (which functions in photosynthesis), the ferredoxin from the heterocyst form of *Anabaena* 7120 (which functions in nitrogen fixation), and human mitochondrial ferredoxin (which participates in a variety of redox reactions including interactions with the cytochrome P450 system). We discuss the preparation of labeled samples, various stable-isotope assisted NMR strategies for studying paramagnetic proteins, and results obtained from these studies. Thomas Pochapsky and coworkers (personal communication) are applying the multinuclear NMR approach to putidaredoxin, a bacterial [2Fe-2S] ferredoxin that is closely related to human ferredoxin and other vertebrate respiratory ferredoxins.

Preparation of Isotopically Labeled [2Fe-2S] Ferredoxins

We have used two overexpression systems in preparing stable-isotope labeled ferredoxins. One system is used primarily for making uniformly labeled protein in high yield with $^{15}NH_4Cl$ as the nitrogen source and/or D-[$^{13}C_6$]glucose as the carbon source. The other system is used when incorporating labeled amino acids to achieve selective labeling. Both systems use the T7 promoter and the T7 polymerase (Studier *et al.*, 1990), but they differ in the way the T7 RNA polymerase is produced. Another plasmid, pLysS, ensures tight control of protein induction. The expression of ferredoxin is induced at mid-log phase (o.d. 1.0). This results in efficient protein production (typically 20–30 mg ferredoxin per liter of culture).

DNA coding for the three ferredoxins, *Anabaena* 7120 vegetative ferredoxin (Alam *et al.*, 1986), *Anabaena* 7120 heterocyst ferredoxin (Böhme and Haselkorn, 1989), and human [2Fe-2S] ferredoxin (Coghlan and Vickery, 1989), have been subcloned into expression vectors, pET3a or pET9a, between *Nde I* (*Nhe I* for heterocyst ferredoxin) and *BamH I* sites (Fig. 1). The vectors containing the ferredoxin coding sequences were transformed into various expression hosts: with pLysS into the nonauxotrophic *E. coli* strain BL21(DE3) or the histidine auxotrophic strain AW608Thr+T7 (both with the λ(DE3) lysogen in their genomes) or with pGp1-2 into other auxotrophic strains of *E. coli*.

Ferredoxins were labeled uniformly with ^{13}C and/or ^{15}N by using the BL21(DE3)/pLysS protein expression system and M9 medium. The appropriate labels were introduced by adding 1 g $^{15}NH_4Cl$ and/or 2 g D-[^{13}C]glucose per liter of medium. The use of plasmid pGp1-2, which contains the temperature-inducible gene for T7 RNA polymerase, makes it easy to move protein expression

into different auxotrophic strains of *E. coli*, as needed for selective isotopic labeling. This approach avoids the necessity of introducing the IPTG inducible T7 RNA polymerase gene into the genomic DNA of each host *E. coli* strain; for example, by P1 transduction, which requires auxotrophic strains with proper λ(DE3) insertion sites.

The temperature-induced expression system has been used to label *Anabaena* 7120 vegetative ferredoxin selectively by incorporating [U-^{15}N]arginine, 26% enriched [U-^{13}C]arginine [α,β–$^{2}H_3$]-arginine, [α–^{2}H]cysteine, [β–$^{2}H_2$]cysteine, [^{15}N]cysteine, and [2-^{13}C]cysteine (Cheng *et al.*, 1995). With this system, it was possible to double the yield of protein for a given quantity of labeled amino acid by recycling the bacterial growth medium. For example, with the arginine and cysteine auxotrophic strains (PA200 and JM15, respectively), after harvesting the first batch of bacterial cells by centrifugation, autoclaved concentrated M9 salts were added to the supernatant, which still contained labeled amino acid, and cell growth was re-initiated. With this procedure, the yield of purified ferredoxin was 40 to 50 mg from either 60 mg of labeled cysteine or 125 mg of labeled arginine.

A sample of the *Anabaena* vegetative ferredoxin was selectively labeled with ^{13}C and ^{15}N for the purpose of assigning one of the cysteine nitrogens. For this, a cysteine auxotroph (JM15) was used as the host strain, and the M9 medium was

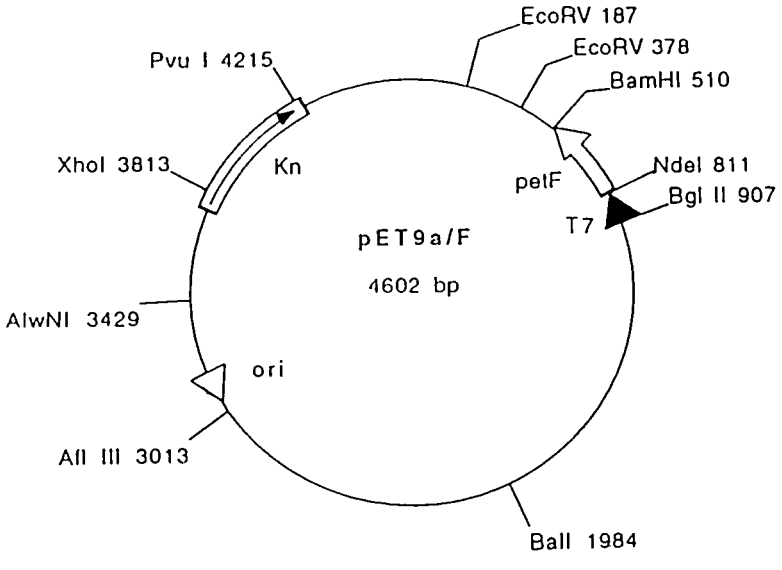


Fig. 1. Restriction map of the pET9a/F vector used to over-express* Anabaena *7120 vegetative ferredoxin. The strong bacteriophage T7 transcription and translation signals were derived from the T7 expression vector, pET9a (Studier* et al., *1990).

supplemented with a mixture of amino acids, all of which were unlabeled except for alanine ([1-^{13}C]alanine) and cysteine ([^{15}N]cysteine). No migration of the label was detected by either ^{13}C or ^{15}N NMR spectroscopy (Cheng *et al.*, 1995).

The T7 promoter/RNA polymerase expression system constructed for production of the *Anabaena* heterocyst ferredoxin [BL(21)/pLysS/pKID] was used for the preparation of uniformly labeled (^{15}N) and double labeled (^{13}C plus ^{15}N) samples. For unknown reasons, attempts to produce *Anabaena* 7120 heterocyst ferredoxin by using the temperature-induced expression system were unsuccessful. Instead, selective labeling of the heterocyst ferredoxin was achieved by incorporating ^{15}N-labeled amino acids according to the recipes of Muchmore *et al.* (1989); a single *E. coli* strain BL21(DE3)/pLysS was used in place of the several auxotrophic strains. In most cases (that is, when incorporating [^{15}N]glycine, [^{15}N]leucine, [^{15}N]valine, [α-^{15}N]lysine, [^{15}N]alanine, or [^{15}N]cysteine), the nitrogen of the labeled amino acid did not migrate to other amino acid types (Chae *et al.*, 1994).

The 26% enriched [U-^{13}C]histidine-labeled human placental ferredoxin was prepared (Xia*et al.*, 1994) by using the histidine auxotrophic *E. coli* strain AW608Thr^{+}T7, whose genomic DNA contains λ(DE3) lysogen as the host strain. The bacteria were grown on an M9 medium plus a mixture of amino acids, all unlabeled except for [U-^{13}C]histidine (26%).

The yields of ferredoxins reported above were achieved by solubilizing the protein and reconstituting the [2Fe-2S] cluster. Results showed that ~75% of the ferredoxin chain produced in the cells was apoferredoxin (lacking the iron-sulfur cluster). It was possible to purify the holoferredoxin and apoferredoxin components separately and to reconstitute the cluster in the apoferredoxin by adding $FeCl_3$ and Na_2S. However, higher yields were obtained by reducing and denaturing the entire ferredoxin-containing fraction before reconstitution, as described by Coghlan and Vickery (1991). Details of the reconstitution and purification procedures used have been described by Chae and Markley (1994), Cheng *et al.*, (1995) and Xia *et al.* (1994).

Multinuclear NMR Studies

Sequence-Specific Assignments of Diamagnetic Resonances

The paramagnetism of the iron-sulfur cluster leads to rapid T_1 and T_2 relaxation of nuclei in its vicinity and spectral overlap in both the hyperfine and diamagnetic regions. Consequent losses in resolution and sensitivity make it difficult to assign the spectrum by the conventional methods of 2D ^{1}H NMR (Wüthrich, 1986). Without isotopic labeling, only a few sequence-specific assignments had been obtained for these ferredoxins (Chan and Markley, 1983a; Chan *et al.*, 1983a). Early ^{13}C labeling studies of *Anabaena variabilis* ferredoxin demonstrated the feasibility of heteronuclear 2D NMR (Chan and Markley, 1982) and the advantages of

^{13}C-labeling for studies of interactions between a ferredoxin and one of its redox partners, ferredoxin NADP$^+$ oxidoreductase (Chan *et al.*, 1983b). Oh and coworkers prepared samples of *Anabaena* 7120 vegetative ferredoxin labeled uniformly with ^{15}N (98%) and/or ^{13}C (26%) and extended the multinuclear NMR studies to the full assignment of the diamagnetic spectrum and an analysis of the secondary structure of the protein in its oxidized state. The ^1H, ^{13}C, and ^{15}N spin systems of individual amino acid residues were identified by comparing the results of several 2D NMR experiments: single- and multiple-bond $^1H\{^{13}C\}$ correlation (Oh *et al.*, 1989), $^1H\{^{15}N\}$ correlation (Oh *et al.*, 1989), $^{13}C\{^{13}C\}$ correlation (Oh *et al.*, 1988), and $^{13}C\{^{15}N\}$ correlation (Mooberry *et al.*, 1989). Sequential connectivities were deduced from standard NOESY spectra and from an isotope-filtered 2D NOESY spectrum (Oh and Markley, 1990a).

More recent developments in multidimensional, multinuclear NMR spectroscopy have led to improved strategies for backbone assignments in proteins labeled with ^{13}C and ^{15}N (Edison *et al.*, 1994). These newer approaches were used in analyzing the NMR spectrum of oxidized *Anabaena* 7120 heterocyst ferredoxin (Chae *et al.*, 1994). Most of the sequential assignments could be derived from the detection of ^{15}N-edited interresidue α^1H^i / N^1H^{i+1} NOE connectivities detected in a 3D NOESY-HMQC (Kay *et al.*, 1989) spectrum (Fig. 2). Intraresidue connectivities were identified from TOCSY-HMQC (Marion *et al.*, 1989) data from the U-^{15}N-sample and from HMQC data from samples labeled selectively with ^{15}N. These assignments were confirmed and extended (Chae *et al.*, 1994) by results from three triple-resonance experiments, HNCA, HNCO, and HN(CO)CA (Grzesiek and Bax, 1992; Bax and Grzesiek, 1993), which led to assignments of three additional residues near the iron-sulfur cluster. Starting points for sequential assignments were provided by 2D HMQC data from several selectively labeled samples (Fig. 3). Samples selectively labeled with [^{15}N]glycine, [^{15}N]leucine, [^{15}N]valine, [^{15}N]alanine, or [α-^{15}N]lysine proved to be most useful for this purpose.

The multinuclear approach led to assignment signals for 80 of the 98 residues in *Anabaena* 7120 heterocyst ferredoxin (Chae *et al.*, 1994). Unassigned peaks come from residues that are close (^1H atoms within ~8 Å) to the iron-sulfur cluster in the x-ray structure of the same ferredoxin (Jacobson *et al.*, 1993). Thus, other approaches had to be developed to obtain information about residues in the vicinity of the cluster, including the four cysteines that are covalently linked to the iron atoms and residues thought to donate hydrogen-bonds to sulfur atoms of the cluster (Rypniewski *et al.*, 1991; Jacobson *et al.*, 1993; Holden *et al.*, 1994).

Assignment of Signals from the Cysteines Ligated to the [2Fe-2S] Cluster

Unlike many other kinds of paramagnetic protein—cytochromes c, for example, which have both diamagnetic and

paramagnetic redox states—the [2Fe-2S] ferredoxins are paramagnetic in both their oxidized and reduced states. Although hyperfine-shifted ^{1}H (Poe *et al.*, 1971), ^{13}C (Chan and Markley, 1983c), and ^{15}N (Oh and Markley, 1990b) signals of [2Fe-2S] ferredoxins were observed much earlier, their sequence-specific assignments have been determined only recently (Dugad *et al.*, 1990; Skjeldal *et al.*, 1991b; Cheng *et al.*, 1995).

The first sequence specific cysteinyl assignments were for the ^{1}H NMR signals. The strategy used involved comparisons of 1D (Dugad *et al.*, 1990) and 2D (Skjeldal *et al.*, 1991b) NOE spectra (obtained with very short mixing times) with interproton distances derived from x-ray structures (Tsukihara *et al.*, 1981; Rypniewski *et al.*, 1991). The optimal mixing times for 2D NOESY were approximately those for the T_1 values of the hyperfine-shifted proton resonances (1 to 10 ms). The results showed that Cys^{41} and Cys^{46} are ligated to Fe(II), and that Cys^{49} and Cys^{79} are ligated to Fe(III) in reduced ferredoxins. Analysis of the NOEs provided stereospecific assignments to the cysteinyl β-methylenes of Cys^{41} and Cys^{46} (Skjeldal *et al.*, 1991b). It proved possible to extend the assignments to several resonances of the oxidized ferredoxin by reference to a 2D exchange spectrum of the half reduced protein (Skjeldal *et al.*, 1991b). These assignments to cysteinyl α- and β-protons in the two oxidation states have been confirmed by selective labeling with deuterium. Figure 4 shows ^{2}H NMR spectra of oxidized and reduced ferredoxin samples labeled with [α–^{2}H]cysteine or [β–$^{2}H_2$]cysteine. These results suggest that ^{2}H NMR spectroscopy in conjunction with selective deuteration may be a fruitful approach to studying hydrogens that interact with the cluster. The lower γ of the ^{2}H nucleus may result in relatively sharp hyperfine peaks, even with the quadrupolar relaxation mechanism.

The cysteinyl ^{15}N and ^{13}C signals of oxidized and reduced *Anabaena* 7120 vegetative ferredoxin were first distinguished from other hyperfine-shifted resonances by reference to spectra of protein samples labeled selectively with [2–^{13}C]cysteine and [^{15}N]cysteine (Figs. 5 and 6). All expected signals were detected in 1D ^{13}C and ^{15}N NMR spectra. The four cysteinyl $^{13}C\beta$ resonances in the reduced state were assigned to individual residues on the basis of correlations with their (previously assigned) β-protons (Cheng *et al.*, 1994b).

The addition of methyl viologen to a solution of partially reduced protein speeds up the electronic self-exchange reaction (Skjeldal *et al.* 1990). As shown in Fig. 5, the addition of methyl viologen increased the electronic self exchange reaction rate so that it was possible to follow the titration of $^{13}C\beta$ peak 3′ of the reduced protein upon progressive oxidation to its position at peak 3 in the fully oxidized ferredoxin. Thus peak 3 is assigned to Cys^{46}. The other three $^{13}C\beta$ peaks, which experience larger chemical shift changes on oxidation/reduction remained in slow exchange on the NMR chemical shift time scale (Fig. 5).

(A)

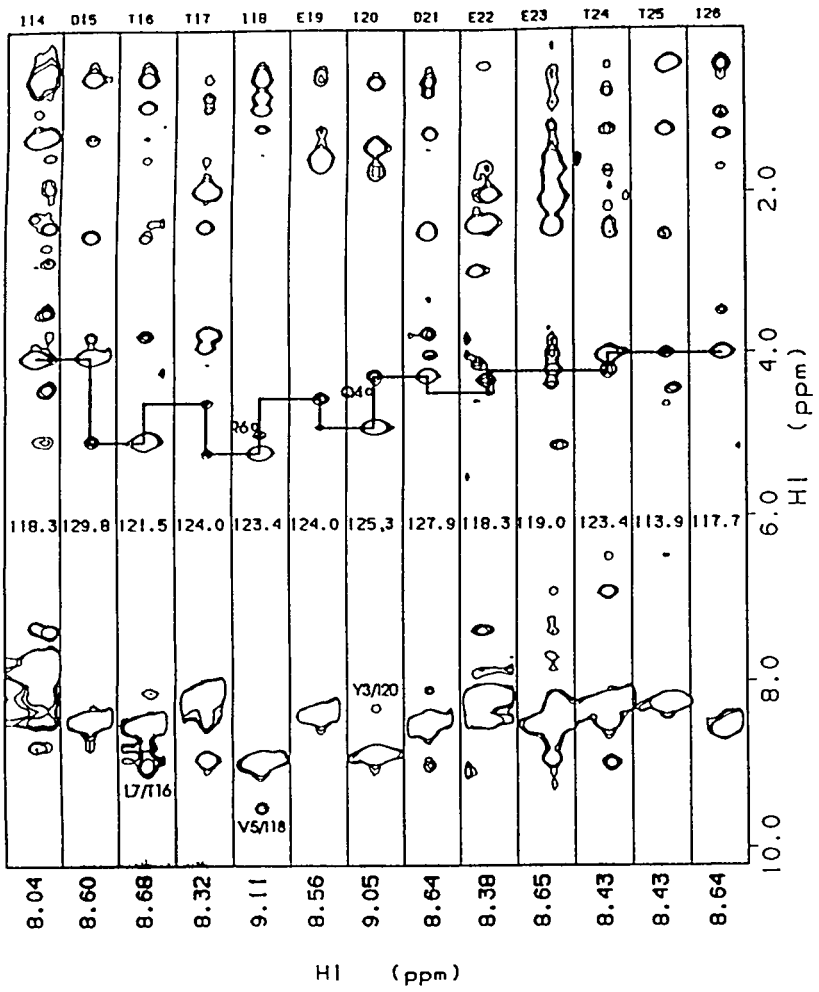
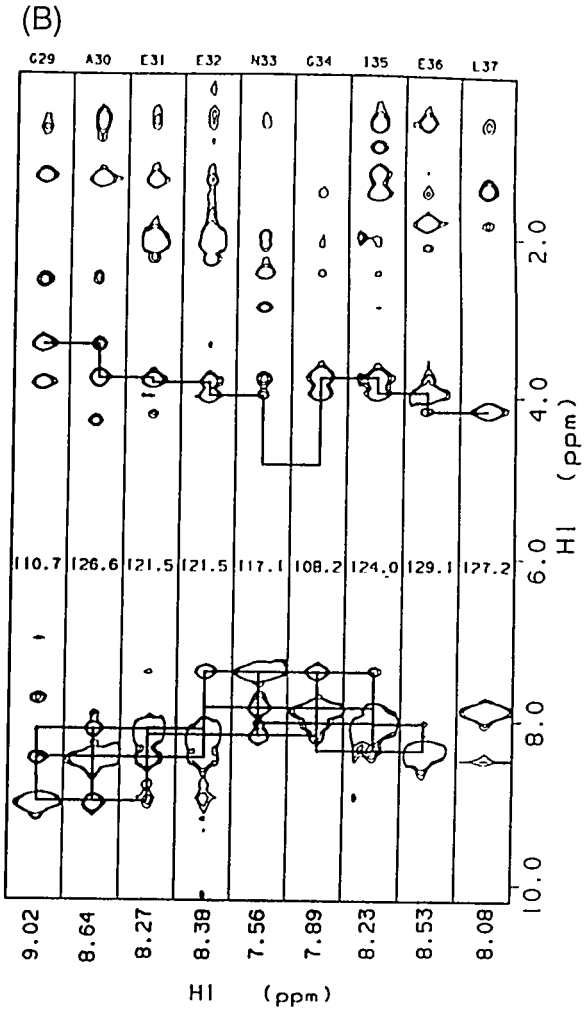


Fig. 2. Series of strips selected from 2D ^{1}H-^{1}H of the 3D NOESY-HMQC spectrum of uniformly ^{15}N labeled Anabaena 7120 heterocyst ferredoxin (adapted from Chae et al., *1994). The assignment of each strip is given at its top. The nitrogen frequencies are given in the middle of the strips. (A) Sequential walk from Ile^{14} to Ile^{26}. These residues form the second β-strand. (B) Sequential walk from Gly^{29} to Leu^{37}. In addition to the $\alpha H^{i}/NH^{i+1}$-type NOE connectivities, NH^{i}/NH^{i+n}-type connectivities are shown. This panel contains data from residues that form the first α–helix. (Amino acid designations are as noted for Wright's Fig. 3, page 6.)*

The ^{15}N NMR signals from the cysteines proved difficult to assign. The cysteinyl $N^{1}H$ signals were not detected (possibly because of paramagnetic effects and exchange with water protons), and multiple bond ^{1}H-^{15}N couplings did not show up in 1D or 2D spectra.



Selective $^{13}C^{i}$-$^{15}N^{i+1}$ double labeling (Kainosho and Tsuji, 1982) was used to identify the unique alanine-cysteine linkage and assign the ^{15}N of Cys^{46}. The one-bond coupling was detected by single-frequency ^{15}N decoupling of the 1D ^{13}C NMR spectrum (Cheng *et al.*, 1995).

The very different spin-lattice relaxation times of the $^{13}C\beta$ of the cysteines ligated to Fe(II) (Cys^{41} = 19 ms; Cys^{46} = 29 ms) and Fe(III) (Cys^{49} = 2.9 ms; Cys^{79} = 2.4 ms) of the reduced ferredoxin (Cheng *et al.*, to be published) indicates the antiferromagnetic coupling between the two irons (J = 100 cm^{-1}) does not average their electronic relaxation times.

Studies of temperature effects on the chemical shifts of ^{1}H, ^{13}C, and ^{15}N resonances from the cysteines that ligate the cluster of *Anabaena* 7120 vegetative ferredoxin (Cheng *et al.*, 1995) showed that signals from a given residue can have either the same or opposite temperature-dependence. For example, in reduced *Anabaena* 7120 vegetative ferredoxin, $\Delta\delta/T^{-1}$ is positive for Cys^{49} $\beta^{1}H$ but is negative for Cys^{49} $^{13}C\beta$. Moreover, although $\Delta\delta/T^{-1}$ is positive for the $\beta^{1}H$ signals of $Cys^{49}$and Cys^{79} and negative for those of $Cys^{41}$and Cys^{46}, $\Delta\delta/T^{-1}$ is positive for the nitrogens of all four cysteines. These results are difficult to accommodate within the current theory used to describe the temperature-dependence of hyperfine shifts in magnetically coupled systems (Banci *et al.*, 1990). These results call into question the common practice of deducing the oxidation state of a given iron in a reduced [2Fe-2S] ferredoxin from the temperature-dependence of NMR peaks from its ligated residues—at least for signals other than ^{1}H.

Analysis of Hydrogen Bonding to Sulfur Atoms in the [2Fe-2S] Cluster

The x-ray structures of [2Fe-2S] ferredoxins suggest that hydrogen-bonds to the cluster may be important in

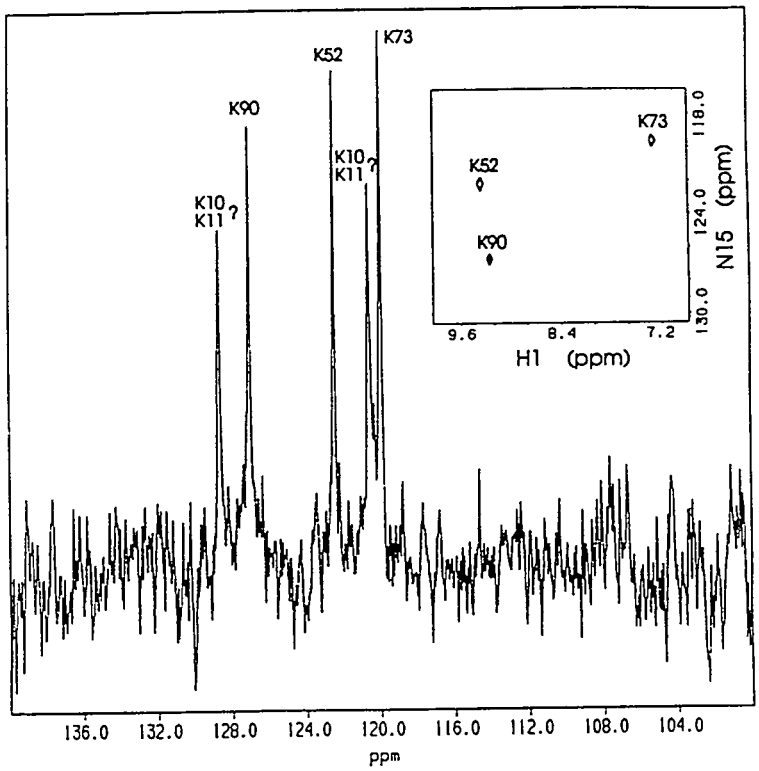


Fig. 3. A 1D ^{15}N spectrum of [α-^{15}N]lysine-labeled Anabaena *7120 heterocyst ferredoxin. Proton decoupling was used during the direct detection of ^{15}N. The inset shows the 2D $^{1}H\{^{15}N\}$ HMQC spectrum of the same sample (adapted from Chae* et al., *1994). The Lysine peaks are labeled as K, followed by their sequence number.*

stabilizing the protein and fine-tuning the redox potential. One such interaction in the *Anabaena* 7120 vegetative ferredoxin appears to involve a hydrogen bond from the backbone amide proton of the residue at position 42 to one of the sulfides of the cluster (Rypniewski *et al.*, 1991). Residue 42 is arginine in nearly all ferredoxins involved in photosynthesis; it is histidine in all ferredoxins known to be involved in nitrogen fixation; and the analogous residue is glutamate in all respiratory ferredoxins. Analysis of the NOESY spectrum of reduced *Anabaena* 7120 vegetative ferredoxin in light of the x-ray structure suggested that hyperfine-shifted peak "K" (12.4 ppm) corresponds to the α-proton of this residue (Skjeldal *et al.*, 1991b). This interpretation, however, proved false when the full ^{1}H-^{13}C spin system of Arg^{42} was determined from NMR studies of a ferredoxin sample selectively labeled with [U-^{13}C]arginine (26% enrichment) (Cheng *et al.*, 1995); peak "K" is not part of the Arg^{42} spin system. Nevertheless, evidence for interaction between Arg^{42} and the cluster comes from ^{15}N NMR data (Fig. 6). Comparison of the 1D ^{15}N NMR spectra of oxidized *Anabaena* 7120 vegetative ferredoxin labeled uniformly with ^{15}N with spectra of a sample labeled selectively with [U-^{15}N]arginine identified a broad, hyperfine-shifted signal at 201.6 ppm as

that from the backbone amide nitrogen of Arg^{42}. The signal is not seen in spectra of reduced ferredoxin; presumably it is still broader and/or shifted much farther downfield. This may indicate that unpaired electron density is delocalized onto Arg^{42} and that the hydrogen bond between the Arg^{42} backbone nitrogen and a sulfide of the cluster, which was inferred from a close contact in the x-ray structure of oxidized ferredoxin (Rypniewski *et al.*, 1991), is present in solution in both the oxidized and reduced protein.

Additional hyperfine-shifted nitrogen signals were seen in spectra of the sample labeled uniformly with ^{15}N (Fig. 6). Recent studies of *Anabaena* 7120 heterocyst ferredoxin have shown interesting correlations between the distances of the nitrogens from the cluster and a dipolar analysis of ^{15}N T_1 relaxation of these resonances (Chae and Markley, 1994). Relaxation studies of these backbone amide nitrogens indicate that the T_1 values are predominantly dipolar. Remarkably close correlations have been found between distances calculated from

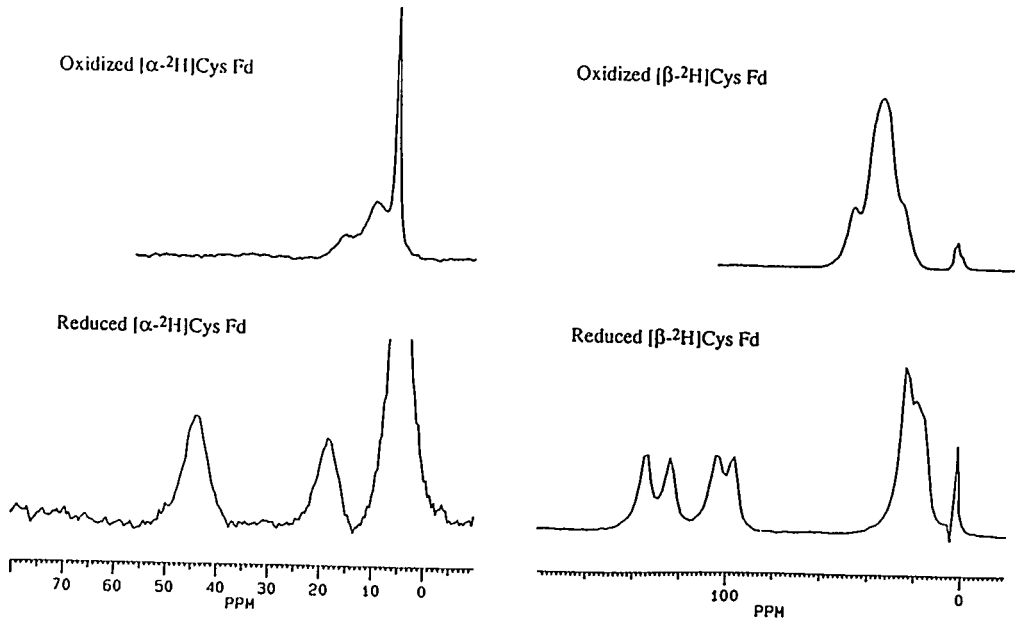


Fig. 4. ***^2H NMR spectra (61.425 MHz) of*** Anabaena ***7120 vegetative ferredoxin labeled selectively with deuterated cysteine (A) with incorporated [α-^2H]cysteine, oxidized; (B) with incorporated [α-^2H]cysteine, reduced; (C) with incorporated [β-$^2H_2$]cysteine, oxidized; (D) with incorporated [β-$^2H_2$]cysteine, reduced. Samples contained 3–5 mM protein in 100 mM sodium pyrophosphate buffer and 100 mM sodium chloride; pH was 8.5. The solvent was deuterium-depleted H_2O (CIL, Woburn, Massachusetts). The spectra were recorded at 283 K (adapted from Cheng*** et al., ***1995).***

^{15}N T_1 values and those derived from the refined x-ray structure for nitrogens that are 4.5 Å or more distant from one or both iron atoms and up to about 8.5 Å from both irons (Chae and Markley, 1994). Lack of agreement for nitrogens at shorter distances to the cluster (including all the cysteine nitrogens) may result either from effects of electron delocalization from the iron atoms to sulfur atoms or from breakdown of the point-dipole approximation.

Properties of Histidine-56, which is Conserved in all Respiratory and Bacterial [2Fe-2S] Ferredoxins

A final example of stable isotope labeling comes from an investigation of His^{56} of human ferredoxin. His^{56} is adjacent to Cys^{55}, one of the cysteines ligated to the iron-sulfur cluster. His^{56} is conserved in the sequences of all known vertebrate [2Fe-2S] ferredoxins; in plant-type ferredoxins, the corresponding residue is alanine, threonine, or leucine (Matusbara and Hase, 1983). It was proposed that His^{56} hydrogen-bonds to a sulfide in the iron-sulfur cluster and helps to transfer electrons to the iron-sulfur center (Lambeth *et al.*, 1982). His^{56} has also been proposed as the residue responsible for the pH-dependence of the reduction potential (pH_{mid} ~7.2) observed in this class of ferredoxin (Cooper *et al.*, 1973). The two previous ^1H NMR studies of the histidines of bovine ferredoxin reached different assignments for His^{56} (Greenfield *et al.*, 1989; Miura *et al.*, 1991). Because overlaps made it difficult to interpret the histidyl region in 1D and 2D ^1H NMR spectra of human

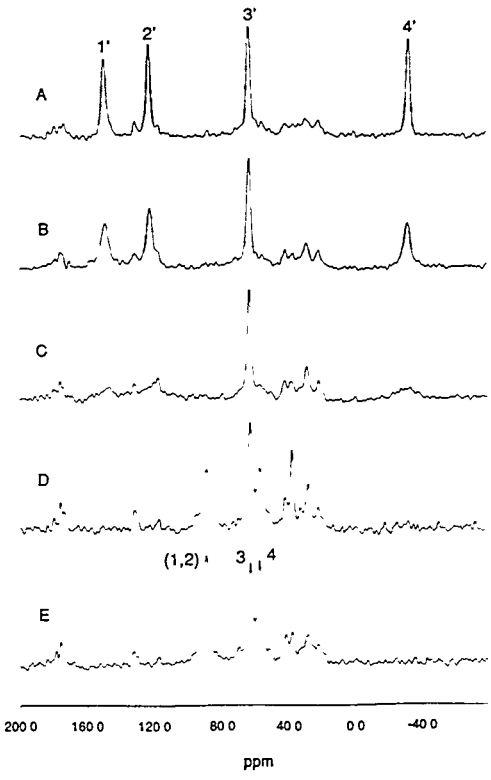


Fig. 5. *^{13}C NMR spectra (150.924 MHz) of progressively oxidized [$^{13}C\beta$]cysteine* Anabaena *7120 vegetative ferredoxin at 298 K (adapted from Cheng* et al., *1995). The sample contained ~4 mM protein in 100 mM sodium pyrophosphate buffer and 100 mM sodium chloride; pH was 8.5; the solvent was 90% ^2H_2O/10% ^1H_2O. The sample was first reduced completely by adding 1 mg of sodium dithionite; then 10 µℓ of 0.29 M methyl viologen was introduced under argon: (A) fully reduced ferredoxin; (B) ~25% air oxidized; (C) ~50% air oxidized; (D) ~75% air oxidized; (E) fully oxidized ferredoxin.*

ferredoxin, the protein was labeled uniformly with ^{15}N and selectively labeled with [U-^{13}C]histidine (26%). NMR analysis of these samples (Xia

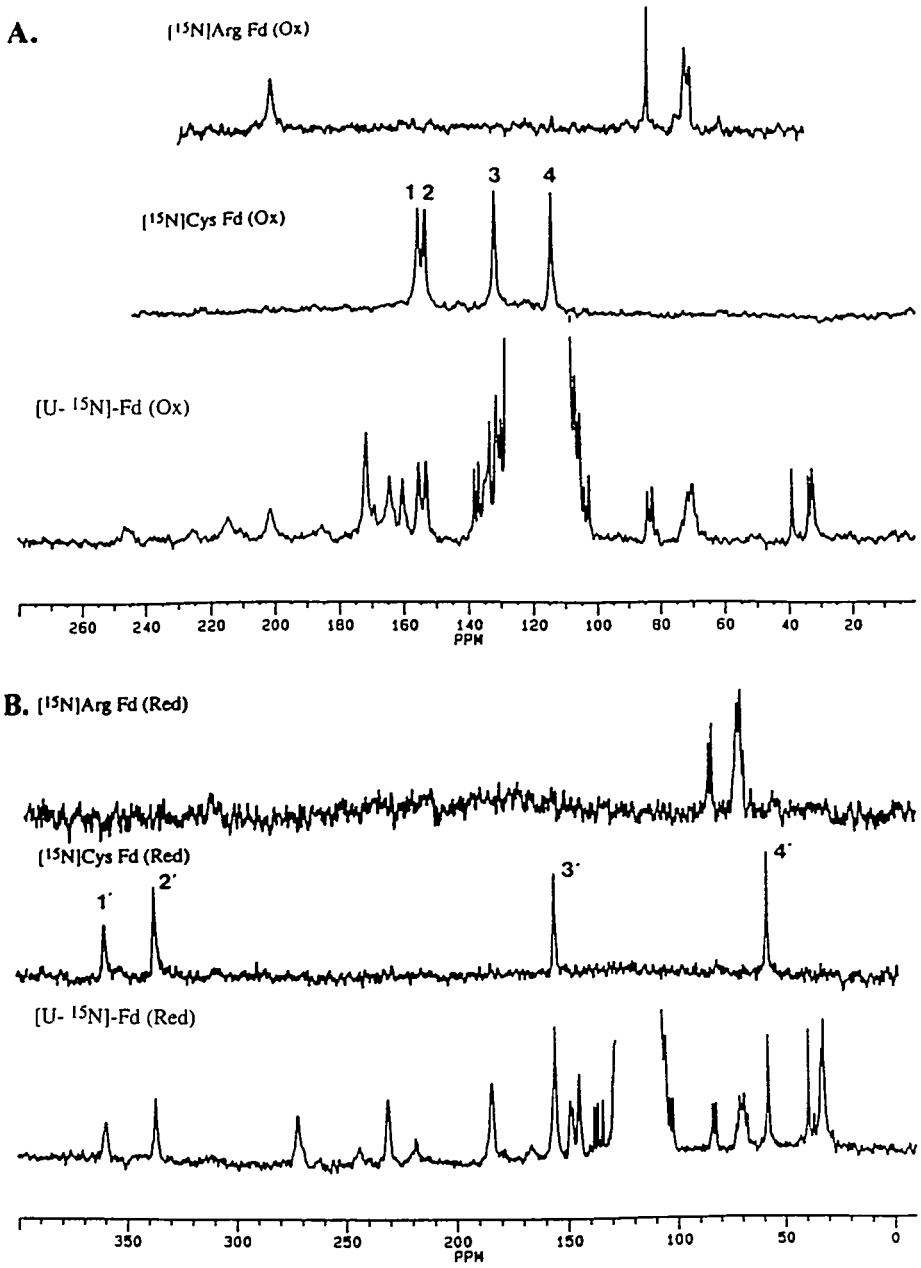


Fig. 6. ^{15}N NMR spectra (60.802 MHz) at 298 K of [U-^{15}N] Anabaena *7120 vegetative ferredoxin and ferredoxins labeled selectively with [^{15}N]cysteine and [^{15}N]arginine (adapted from Cheng* et al., *1995). The samples contained ~4 mM protein. The solvent was 90% $^{2}H_2O$/10% $^{1}H_2O$. (A) Oxidized samples in 50 mM phosphate buffer; pH was 7.1. (B) Reduced samples in 100 mM sodium pyrophosphate buffer and 100 mM sodium chloride; pH was 8.5.*

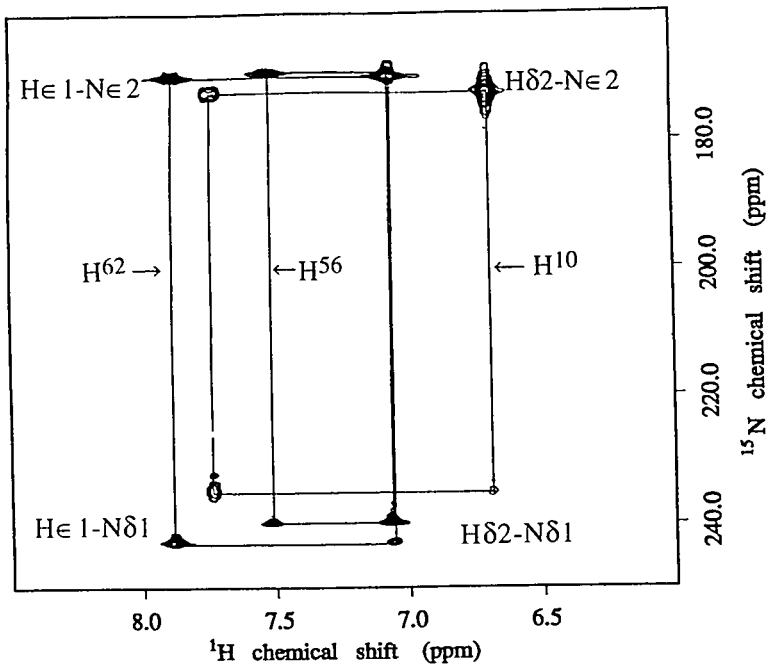


Fig. 7. Histidine region of the 600-MHz 2D $^1H\{^{15}N\}$MBC spectrum of oxidized human ferredoxin labeled uniformly with ^{15}N. The sample was dissolved in ^2H_2O/10% ^1H_2O; pH value was 7.4. Rectangular connectivities link crosspeaks from multibond ^1H-^{15}N couplings in the imidazole ring of each of the three histidines (adapted from Xia et al., *1994.*

et al., 1994) made it possible to assign nearly all the sidechain ^1H, ^{13}C, and ^{15}N resonances from the three histidines of human ferredoxin.

Figure 7 shows the $^1H\{^{15}N\}$MBC spectrum of oxidized human ferredoxin labeled uniformly with ^{15}N, which identifies the spin systems of the sidechains of the three histidines. Figure 8 shows ^{13}C-edited ^1H NMR spectra of the sample with [U-^{13}C]histidine (26%) collected at three pH values. Note that the signals assigned to His^{56} (c1 = ε1H and c2 = δ2H) have pH-independent chemical shifts between pH 6.0 and 8.6. These data were analyzed in conjunction with the pH-dependence of normal 1D ^1H NMR spectra of the protein at natural abundance to yield the pK_a values of the histidines. The assignments to His^{56} of human ferredoxin were consistent with those of Miura and coworkers (Miura and Ichikawa, 1991; Miura *et al.*, 1991) for bovine ferredoxin. Also, in agreement with the results of Miura *et al.* (1991) for oxidized bovine ferredoxin, His^{56} of human ferredoxin was found to have an abnormal pK_a value. Results from NMR analysis of the ^{15}N- and ^{13}C-labeled samples showed that His^{56} of human ferredoxin has $pK_a < 5$. Miura *et al.* (1991)

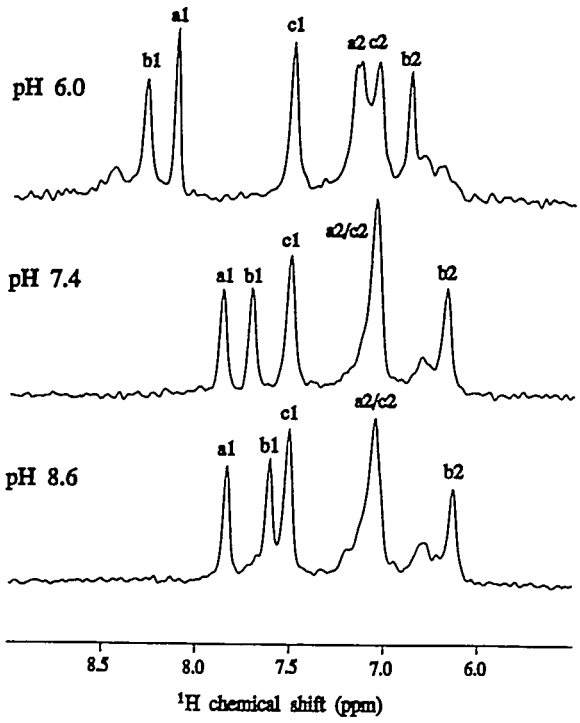


Fig. 8: Aromatic region of 1D ^{13}C-edited ^{1}H NMR spectra of oxidized human ferredoxin with incorporated [26% U-^{13}C]-histidine at 25°C and at three pH values (adapted from Xia et al., 1994*). Peak assignments: a1, His^{62} ε1H; a2, His^{62} δ2H; b1, His^{10} ε1H; b2, His^{10} δ2H; c1, His^{56} ε1H; c2, His^{62} δ2H.*

have attributed this to an interaction with Ser^{88}. Because the pK_a values of the two titratable histidines (His^{10} = 6.2; His^{62} = 5.8) differ from 7.2, none of the three histidines can be responsible for the observed pH-dependence of the reduction potential of vertebrate ferredoxins (Xia *et al.*, 1994).

Summary

Stable-isotope-assisted NMR spectroscopy has proven to be a profitable approach in our structure-function studies of three [2Fe-2S] ferredoxins: *Anabaena* 7120 vegetative ferredoxin, *Anabaena* 7120 heterocyst ferredoxin, and human ferredoxin. These investigations have revealed new information about antiferromagnetic coupling between the two irons in the reduced cluster, properties of the cysteines that ligate the cluster, and characteristics of other amino acid residues in the vicinity of the cluster—some of which participate in hydrogen bonding to the cluster. Multinuclear NMR studies have led to solution structures of the *Anabaena* 7120 vegetative ferredoxin (Chae, 1994), and to comparisons of the secondary structure of the heterocyst ferredoxin in solution and in the crystal (Chae *et al.*, 1994). Some of the problems with NMR spectroscopy of paramagnetic proteins can be solved by selective labeling experiments that simplify spectra and facilitate assignments. Labeling with nuclei of lower magnetogyric ratio than protons (^{2}H, ^{13}C, and ^{15}N) allows detection of signals from nuclei that are close to the most interesting part of the protein: the paramagnetic center. Stable-isotope labeling of recombinant proteins can be relatively easy and economical. Efficient overexpression of proteins (such as these ferredoxins) that can be toxic to host cells requires the use of a promoter with very tight control of induction. For selective labeling by incorporation of labeled amino acids, it is convenient to have all

of the expression machinery on plasmids that can easily be transformed into appropriate auxotrophic host strains of *E. coli*.

Acknowledgements

This work was supported by NSF grant MCB-9215142 and USDA grant CSRS 92-37206-7699; we made use of the National Magnetic Resonance Facility at Madison, which is supported in part by NIH grant RR023021 from the Biomedical Research Technology Program, Division of Research Resources. Equipment in the facility was purchased with funds from the University of Wisconsin, the NSF Biological Biomedical Research Technology Program (grant PR02781), and the US Department of Agriculture.

References

Alam, J., Whitaker, R.A., Krogmann, D.W., and Curtis, S.E. (1986) *J. Bacteriol. 168*, 1265-1271.

Banci, L., Bertini, I., and Luchinat, C. (1990) *Struct. Bonding 72*, 113-136.

Bax, A. and Grzesiek, S. (1993) *Accts. Chem. Res. 26*, 131-138.

Bertini, I., Briganti, F., Luchinat, C., Messori, L., Monnanni, R., Scozzafava, A., and Vallini, G. (1991) *FEBS Lett. 28*, 253-2556.

Bhme, H. and Haselkorn, R. (1989) *Plant Mol. Biol. 12*, 667-672.

Chae, Y.K. (1994) Ph.D. thesis, University of Wisconsin, Madison.

Chae, Y.K., Abildgaard, F., Mooberry, E.S., and Markley, J.L. (1994) *Biochemistry 33*, 3287-3295.

Chae, Y.K. and Markley, J.L. (1994) *Biochemistry*, in press.

Chan, T.-M. and Markley, J.L. (1982) *J. Am. Chem. Soc. 104*, 4010-4011.

Chan, T.-M. and Markley, J.L. (1983a) *Biochemistry 22*, 5982-5987.

Chan, T.-M. and Markley, J.L. (1983b) *Biochemistry 22*, 5996-6002.

Chan, T.-M. and Markley, J.L. (1983c) *Biochemistry 22*, 6008-6010.

Chan, T.-M., Hermodson, M.A., Ulrich, E.L., and Markley, J.L. (1983a) *Biochemistry 22*, 5988-5995.

Chan, T.-M., Ulrich, E.L., and Markley, J.L. (1983b) *Biochemistry 22*, 6002-6007.

Cheng, H., Grohmann, K., and Sweeney, W. (1992) *J. Biol. Chem. 26,* 8073-8080.

Cheng, H., Westler, W.M., Xia, B., Oh, B-H., and Markley, J.L. (1995) *Arch. Biochem. Biophys.*, in press.

Cheng, H., Xia, B., Reed, G.H., and Markley, J.L (1994) *Biochemistry 33,* 3155-3164.

Coghlan, V.M. and Vickery, L.E. (1989) *Proc. Natl. Acad. Sci. USA 86,* 835-839.

Coghlan, V.M. and Vickery, L.E. (1991) *J. Biol. Chem. 266,* 18606-18612.

Cooper, D.Y., Schleyer, H., Levin, S.S., and Rosenthal, O. (1973) *Ann. NY Acad. Sci. 212,* 227-247.

Dugad, L.B., La Mar, G.N., Banci, L., and Bertini, I. (1990) *Biochemistry 29,*2263-2271.

Edison, A.S. Abildgaard, F., Westler, W.M., Mooberry, E.S., and Markley, J.L. (1994) *Methods Enzymol.* 239, 3-79.

Greenfield, N.J., Wu., X., and Jordan, F. (1989) *Biochim. Biophys. Acta. 995,* 246-254.

Holden, H.M., Jacobson, B.L., Hurley, J.K., Tollin, G., Oh. B.-H., Skjeldal, L., Chae, Y.K., Cheng, H., Xia, B., and Markley, J.L. (1994) *J. Bieng. Biomemb. 26,* 67-88.

Jacobson, B.L., Chae, Y.K., Markley, J.L., Rayment, I., and Holden, H.M. (1993) *Biochemistry 32,* 6788-6793.

Kay, L.E., Marion, D., and Bax, A. (1989) *J. Magn Reson. 84,* 72-84.

Kainosho, M. and Tsuji, T. (1982) *Biochemistry 21,* 6273-6279.

Lambeth, J.D., Seybeth, D.W., Lancaster, J.R., Jr., Salerno, J.C., and Kamin, H. (1982) *Mol. Cell. Biochem. 45,* 13-31.

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A., and Bax, A. (1989) *J. Am. Chem. Soc. 111,* 1515-1517.

Markley, J.L. and Kainosho, M., (1993) in *NMR of Biological Macromolecules: A Practical Approach,* Roberts, G.C.K, ed., Oxford University Press, Oxford, pp. 101-152.

Matsubara, H. and Hase, T. (1983) in *Proteins and Nucleic Acids in Plant Systematics,* Jensen, U., and Fairbrothers, D.E., eds., Springer-Verlag, Berlin, pp. 168-181.

Miura, S. and Ichikawa, Y. (1991) *J. Biol. Chem. 266,* 6252-6258.

Miura, S., Tamita, S., and Ichikawa, Y. (1991) *J. Biol. Chem. 266,* 19212-19216.

Mooberry, E.S., Oh, B.-H., and Markley, J.L. (1989) *J. Magn. Reson. 85,* 147-149.

Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E., and Dahlquist, F.W. (1989) *Methods Enzymol. 177,* 44-73.

Oh, B.-H., Westler, W.M., Darba, P., and Markley, J.L. (1988) *Science 240,* 908-911.

Oh, B.-H., Westler, W.M., Darba, P., and Markley, J.L. (1989) *J. Am. Chem. Soc. 111,* 3083-3085.

Oh, B.-H. and Markley, J.L. (1990a) *Biochemistry 29,* 3994-4004.

Oh, B.-H. and Markley, J.L. (1990b) *Biochemistry 29,* 4012-4017.

Oh, B.-H.,Mooberry, E.S., and Markley, J.L. (1990) *Biochemistry 29*, 4004-4011.

Poe, M., Phillips, W.D., Glickson, J.D., and San Pietro, A. (1971) *Proc. Natl. Acad. Sci. USA 68*, 68-71.

Rypniewski, W.R., Breiter, D.R., Benning, M.M., Wesenberg, G., Oh, B.-H., Markley, J.L., Rayment, I., and Holden, H.M. (1991) *Biochemistry 30*, 4126-4131.

Skjeldal, L., Westler, W.M., and Markley, J.L. (1990) *Arch. Biochem. Biophys. 278*, 482-485.

Skjeldal, L., Coghlan, V.M., Vickery, L., and Markley, J.L. (1991a) *Biochemistry 30*, 9078-9083.

Skjeldal, L., Westler, W.M., Oh, B.-H., Krezel, A.M.,. Holden, H.M., Jacobson, B.L., Rayment, I., and Markley, J.L. (1991b) *Biochemistry 30*, 7363-7368.

Studier, F.W., Rosenburg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) *Methods Enzymol. 185*, 60-89.

Tsukihara, T., Fukuyama, K., Nakamura, M., Katsube, T., Tanaka, N., Kakudo, M., Wada, K., Hase, T., and Matsubara, H. (1981) *J. Biochemistry 90*, 1763-1773.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.

Xia, B, Cheng, H., Skjeldal, L., Choghlan, V.M., Vickery, L.E., and Markley, J.L. (1994) *Biochemistry*, in press.

NMR Studies of Isotopically Labeled RNA

Arthur Pardi

Department of Chemistry and Biochemistry
Campus Box 215
University of Colorado, Boulder
Boulder, CO 80309-0215

RNA has long been known to play a critical role in translation where mRNA, tRNA, and rRNA are all absolutely required for this essential biological function. However, in the last decade it has become clear that the biological role for RNA is much more diverse than was initially thought. The fact that RNAs not only can carry genetic information in RNA viruses, but also can perform catalysis in RNA splicing and cleavage reactions has led to renewed interest in the functional role of RNAs in biology (Cech, 1987; Cech, 1989; Symons, 1992). It is now known that RNAs are required for proper function of a growing number of biological processes such as RNA splicing, RNA editing, and telomere formation. There is even recent evidence indicating that the ribosomal RNA itself can carry out a peptidyl-transferase reaction (Noller *et al.*, 1992). However, this increased understanding of RNA's role in cellular functions has not been matched by a complementary increase of information on the structure of RNA. In contrast to the protein field, where there are hundreds of high-resolution 3D structures of functional proteins, there are currently only a handful of 3D structures of functional RNAs. One reason for the lack of 3D structural data originally was the difficulty in synthesizing milligram quantities of defined RNA sequences. This problem has been overcome in the last 5 years by improved methods for RNA synthesis by both *in vitro* transcription (Milligan and Uhlenbeck, 1989) and chemical synthesis (Gait *et al.*, 1991). In addition, x-ray studies of RNA have suffered from difficulties in obtaining highly diffracting single crystals of RNAs; tRNA is still the only high-resolution x-ray structure of an RNA with tertiary structure (Saenger, 1984). Thus, most recent structural data on RNA have been obtained from solution NMR studies

(Varani and Tinoco, 1991). Unfortunately, these NMR studies have generally been limited to relatively small systems because of the high degree of overlap in the proton spectra of RNAs (Varani and Tinoco, 1991). Without methods for isotopic labeling of RNA, it was not clear that high-resolution structures of larger RNAs (more than 15 to 20 residues) could be obtained from solution NMR studies.

Advances in multidimensional heteronuclear NMR spectroscopy have made it possible to determine 3D structures of larger biomolecules in solution (Fesik and Zuiderweg, 1990; Clore and Gronenborn, 1991). These heteronuclear experiments require milligram quantities of isotopically (^{13}C and/or ^{15}N)-labeled molecules and therefore, until recently, they had only been used in protein structure determinations. The application of 2D, 3D, and even 4D heteronuclear NMR experiments has revolutionized the resonance assignment and structure determination of isotopically labeled proteins, and these techniques are starting to prove equally powerful in structure determinations of isotopically labeled nucleic acids. In the past 2 years, efficient methods for generating NMR quantities of ^{15}N- and/or ^{13}C-labeled RNA oligomers of defined sequence have become available (Nikonowicz *et al.*, 1992; Batey *et al.*, 1992). This ability to generate isotopically labeled molecules opened the RNA structure field for application of heteronuclear multidimensional NMR techniques (Nikonowicz and Pardi, 1992a; Nikonowicz and Pardi, 1992b; Pardi and Nikonowicz, 1992; Nikonowicz and Pardi, 1993). This manuscript outlines recent progress in methods for synthesis of isotopically labeled RNAs and briefly describes some recent applications of heteronuclear multidimensional NMR to the structure determination of isotopically labeled RNAs.

Efficient and economical synthesis of RNA oligomers has proven much more challenging than synthesis of DNA oligomers. Difficulties in chemical synthesis of RNA led to development of alternate methods for producing RNA oligomers, namely enzymatic synthesis by *in vitro* transcription (Milligan and Uhlenbeck, 1989; Milligan *et al.*, 1987). In this procedure, T7, T3, or SP6 RNA polymerase is combined with a DNA template and ribonucleotide triphosphates (NTPs) to generate RNA oligomers of defined sequence. It is possible to generate milligram quantities of almost any size or sequence of RNA by *in vitro* transcription, but two limitations must be considered. The first is that very short sequences (<10 nucleotides) cannot be efficiently transcribed, and the second is that the RNA polymerases have specific nucleotide requirements at the 5′ end of the transcript (Milligan and Uhlenbeck, 1989; Milligan *et al.*, 1987). For example, the most commonly used polymerase, T7 RNA polymerase, requires a G residue at the 5′ end of the transcript, and the yield of transcription varies with the sequence of the first 3 to 5 nucleotides in the RNA (Milligan *et al.*, 1987). Other factors, such as runs of U residues or self-structure in the DNA template, can also influence the

yields of *in vitro* transcription. This sequence-dependent variation of the yields means that the transcription conditions must be carefully optimized for each individual RNA sequence.

Although there are limitations in the *in vitro* transcription procedure, a major advantage over chemical synthesis is that the precursors for the RNA oligomer simply are NTPs. Thus if one needs to introduce isotope labels into an RNA oligomer generated by *in vitro* transcription, all that is needed is a source of isotopically labeled NTPs. Although there are presently no commercial sources of ^{13}C- or ^{15}N-labeled NTPs, it is possible to efficiently prepare ^{13}C- and/or ^{15}N-labeled NTPs by biosynthetic methods. We and others have recently reported efficient procedures for generating 99% ^{13}C- and/or ^{15}N-labeled NTPs that can be used to synthesize isotopically labeled RNAs by *in vitro* transcription (Nikonowicz *et al.*, 1992; Batey *et al.*, 1992). Figure 1 shows a schematic flow chart of our procedure for generating ^{13}C- and/or ^{15}N-labeled RNA oligomers of defined sequence. Polymeric RNA is isolated from *E. coli* grown on minimal medium where 99% ^{13}C glucose and 99% ^{15}N ammonium sulfate are the only sources of carbon and nitrogen. The polymeric RNA is enzymatically degraded to 5′ nucleotide monophosphates by nuclease P1. If all four residue types are to be labeled in the RNA, then this NMP mixture is enzymatically converted to NTPs, which are then used in the *in vitro* transcription reaction. If only a subset of residue types are to be labeled, then the NMP mixture is chromatographically separated, and the individual NMPs are enzymatically converted to NTPs. The individual NTPs are then combined with unlabeled NTPs to generate RNA

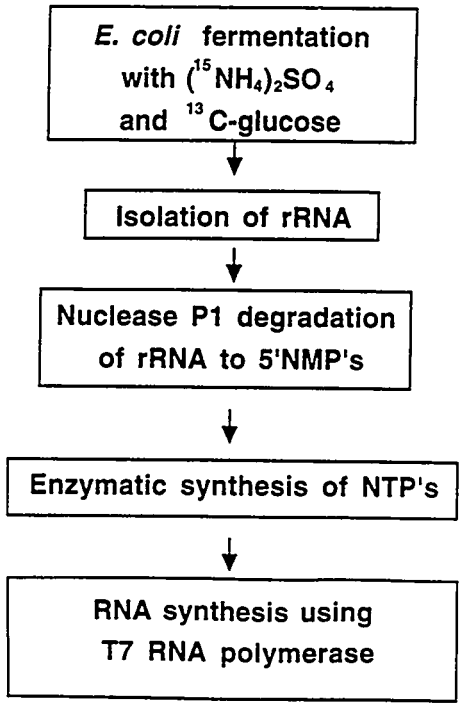


Fig. 1. This schematic outlines the protocol for preparing isotopically labeled RNA (Nikonowicz et al., *1992). Because the ^{13}C-labeled glucose and ^{15}N-labeled ammonium sulfate are the only carbon and nitrogen sources in the* E. coli *growth medium, all the RNA in the cell is fully isotopically labeled. After the ribosomes are isolated, the ribosomal RNA is extracted and then degraded to 5′ NMPs. The NMPs can be directly converted to NTPs, as shown here, or separated and then individually converted to NTPs (Nikonowicz* et al., *1992). These NTPs are then used in the* in vitro *transcription reactions.*

oligomers that are specifically labeled by residue type. The cost of generating a ^{13}C- and/or ^{15}N-labeled RNA oligomer by this procedure depends primarily on the amount of ^{13}C glucose needed—which varies substantially, primarily because of sequence-dependent variation in the yields of the transcription reaction. For RNAs that give very high transcription yield, it can cost as little as $300 in ^{13}C glucose (at $475/g) to produce a 1-mM NMR sample of a 99% ^{13}C-labeled RNA; however, the more typical cost for producing a ^{13}C-labeled RNA is $1500 to $2000 in ^{13}C-labeled glucose. This cost can be reduced further by growing methylotropic bacteria on ^{13}C-labeled methanol (at $80/g) as the sole carbon source (Batey *et al.*, 1992), and by recycling the unused labeled NTPs from the transcription reactions (P. Legault and A. Pardi, unpublished results).

The biosynthetic production of isotopically labeled NTPs for use in *in vitro* transcription reactions makes it possible to apply the powerful multidimensional heteronuclear NMR techniques that are revolutionizing the solution structure determination of proteins to nucleic acids as well. The following section summarizes how these heteronuclear techniques can be used to simplify the resonance assignment and structure determination of isotopically labeled RNA oligomers.

One of the main difficulties in NMR studies of RNA arises from the poor dispersion of the proton spectrum. Figure 2 shows a 1D proton spectrum of

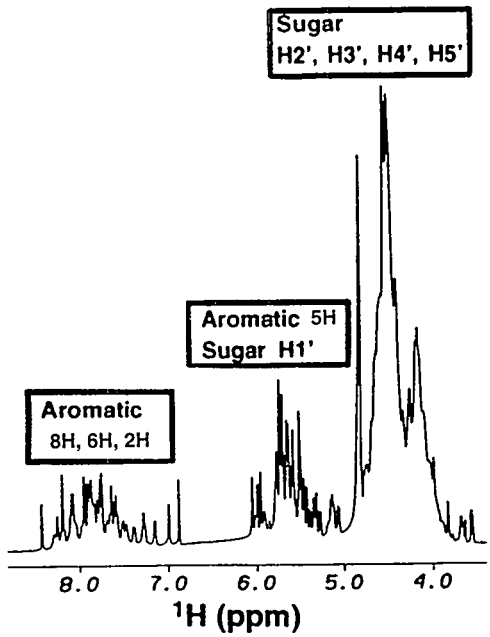


Fig. 2. The 1D proton spectrum of a 30-nucleotide RNA, a lead-dependent ribozyme (Pan and Uhlenbeck, 1992).

a 30-nucleotide lead-dependent ribozyme called the leadzyme (Pan and Uhlenbeck, 1992). This illustration shows the high degree of overlap in the spectra, especially in the H2′, H3′, H4′, H5′, and H5″ sugar region between 4 and 5 ppm. Heteronuclear NMR can help alleviate this overlap, as is illustrated in the constant-time HSQC spectrum (Santoro and King, 1992) of the leadzyme shown in Fig. 3, where the H2′ through H5″ resonances can now be resolved by the differences in chemical shifts of the C2′ through C5′ resonances. This much greater dispersion

of the carbon resonances in the ribose sugar provides the basis for applying a wide range of powerful heteronuclear experiments, such as 3D HCCH-COSY, HCCH-RELAY, HCCH-TOCSY, 3D HMQC-NOESY, and 4D HMQC-NOESY-HMQC experiments, to RNA (Nikonowicz and Pardi, 1992a; Nikonowicz and Pardi, 1992b; Pardi and Nikonowicz, 1992; Nikonowicz and Pardi, 1993). We have recently shown that these 3D and 4D experiments enormously simplify the resonance assignment of 99% ^{13}C-labeled RNAs (Nikonowicz and Pardi, 1992a; Nikonowicz and Pardi, 1992b; Pardi and Nikonowicz, 1992; Nikonowicz and Pardi, 1993).

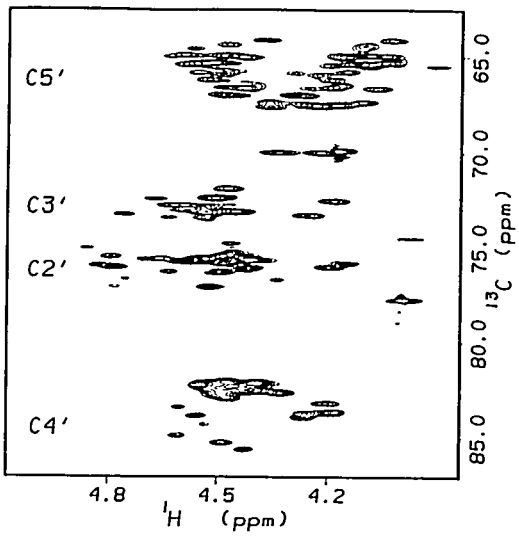


Fig. 3. The 2' to 5' sugar region of a 2D ^{1}H–^{13}C constant-time HSQC (Santoro and King, 1992) experiment on the leadzyme (Pan and Uhlenbeck, 1992) This spectrum illustrates that even though the H2' to H5'/H5" sugar resonances are overlapped in the proton spectrum, the C2' to C5' resonances are resolved from one another in the carbon spectrum.

The first step in any solution NMR structure determination of a biomolecule is assignment of the proton spectrum (Wüthrich, 1986). Proton resonance assignments in proteins and nucleic acids were originally made by a combination of through-bond *J* coupling connectivities and through-space nuclear Overhauser effect connectivities. A very important advance in resonance assignment in proteins has been the application of 3D NMR techniques for complete through-bond assignment of the backbone proton, carbon, and nitrogen resonances in isotopically labeled molecules (Kay *et al.*, 1990; Olejniczak *et al.*, 1992). Identification of all the protons that belong to the same sugar ring has been a major problem in resonance assignment of RNAs because of the small H1'–H2' coupling constant for sugar residues in A-form RNA (Varani and Tinoco, 1991). We have shown that application of 2D and 3D HCCH-COSY, HCCH-RELAY, and HCCH-TOCSY experiments on 99% ^{13}C-labeled RNAs leads to facile and unambiguous assignment of the ribose proton and carbon resonances (Pardi and Nikonowicz, 1992; Nikonowicz and Pardi, 1993). In these experiments, magnetization is transferred between protons and carbons in the ribose ring through large one-bond heteronuclear

^{1}H–^{13}C (150 – 165 Hz) and homonuclear ^{13}C–^{13}C (38 – 42 Hz) coupling constants. The large coupling constants allow much more efficient transfer of magnetization than can be obtained through the much smaller three-bond ^{1}H–^{1}H coupling constants.

After homonuclear or heteronuclear through-bond experiments are used to identify sugar and base spin systems in a DNA or RNA oligomer (Nikonowicz and Pardi, 1993), the classic sequential resonance assignment procedure is generally employed for assigning the sugar and base protons to specific residues in the molecule (Wüthrich, 1986). This sequential resonance assignment involves identification of sugar proton to base proton NOEs on neighboring residues. We have recently described a systematic procedure for sequential resonance assignment of the exchangeable and nonexchangeable protons in uniformly labeled RNAs by application of a variety of 3D ^{13}C and ^{15}N NOESY-HMQC or HMQC-NOESY experiments (Nikonowicz and Pardi, 1993). Although these techniques are critical for the assignment of larger RNA, one limitation of this NOE based procedure is that it is not possible to differentiate between intra- and inter-residue NOEs. Therefore, it is desirable to have procedures for linking sugar protons and base protons in the same residue by applying through-bond scalar coupling. Recently, we and others independently developed a set of triple-resonance 3D NMR experiments that can be used to make through-bond connectivities between sugar and base protons and carbons in uniformly ^{13}C- and/or ^{15}N-labeled RNAs (Farmer *et al.*, 1993; Farmer *et al.*, 1994; Sklenar *et al.*, 1993a; and Sklenar *et al.*, 1993b). These experiments provide important complementary information to the 3D NOESY-HMQC experiments because they represent conformation-independent methods for generating intraresidue resonance assignments.

In summary, the ability to generate NMR quantities of ^{15}N- and ^{13}C-labeled RNAs has led to the development of heteronuclear multi-dimensional NMR techniques for simplifying the resonance assignment and structure determination of RNAs. These methods for synthesizing isotopically labeled RNAs are only several years old, and thus there are still relatively few applications of heteronuclear multi-dimensional NMR techniques to RNA. However, given the critical role that RNAs play in cellular function, one can expect to see an increasing number of NMR structural studies of biologically active RNAs.

Acknowledgments

I thank Pascale Legault and Kathy Morden for a critical reading of the manuscript. This work was supported in part by NIH grant AI30726 and NIH grant AI33098 and NIH Research Career Development Award AI01051.

References

Batey, R.T., Inada, M., Kujawinski, E., Puglisi, J.D., and Williamson, J.R. (1992) *Nucleic Acids Res. 20*, 4515-4523.

Cech, T.R. (1987) *Science 236*, 1532-9.

Cech, T.R., ed. (1989) *Molecular Biology of RNA, Vol. 94*, A.R. Liss, New York.

Clore, G.M. and Gronenborn, A.M. (1991) *Science 252*, 1390-1399.

Farmer, B.T., II, Mueller, L., Nikonowicz, E.P., and Pardi, A. (1993) *J. Am. Chem. Soc. 115*, 11040-11041.

Farmer, B.T., II, Mueller, L., Nikonowicz, E.P., and Pardi, A. (1994) *J. Biomol. NMR 4*, 129-133.

Fesik, S.W. and Zuiderweg, E.R.P. (1990) *Q. Rev. Biophys. 23*, 97-131.

Gait, M.J., Pritchard, C., and Slim, G. (1991) in *Oligonucleotides and Analogues*, F. Eckstein, ed., Oxford University Press, New York, pp. 25-48.

Kay, L.E., Ikura, M., Tschudin, R., and Bax, A. (1990) *J. Magn. Reson. 89*, 496-514.

Milligan, J.F., Broebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987) *Nucleic Acids Res. 15*, 8783-8789.

Milligan, J.F. and Uhlenbeck, O.C. (1989) *Methods Enzymol 180*, 51-62.

Nikonowicz, E.P. and Pardi, A. (1992a) *Nature 355*, 184-186.

Nikonowicz, E.P. and Pardi, A. (1992b) *J. Am. Chem. Soc. 114*, 1082-1083.

Nikonowicz, E.P., Sirr, A., Legault, P., Jucker, F.M., Baer, L.M., and Pardi, A. (1992) *Nucleic Acids Res. 20*, 4507-4513.

Nikonowicz, E.P. and Pardi, A. (1993) *J. Mol. Biol. 232*, 1141-1156.

Noller, H.F., Hoffarth, V., and Zimniak, L. (1992) *Science 256*, 1416-9.

Olejniczak, E.T., Xu, R.X., Petros, A.M., and Fesik, S.W. (1992) *J. Magn. Reson. 100*, 444-450.

Pan, T. and Uhlenbeck, O.C. (1992) *Nature 358*, 560-563.

Pardi, A. and Nikonowicz, E.P. (1992) *J. Am. Chem. Soc. 114*, 9301-9302.

Saenger, W. (1984) *Principles of Nucleic Acid Structure* , Springer-Verlag, New York.

Santoro, J. and King, G.C. (1992) *J. Magn. Reson. 97*, 202-207.

Sklenar, V., Peterson, R.D., Rejante, M.R., and Feigon, J.J. (1993a) *Biomol. NMR 3*, 721-727.

Sklenar, V., Peterson, R.D., Rejante, M.R., Wang, E., and Feigon, J. (1993b) *J. Am. Chem. Soc. 115*, 12181-12182.

Symons, R.H. (1992) *Ann. Rev. Biochem. 61*, 641-671.

Varani, G. and Tinoco, I., Jr. (1991) *Q. Rev. Biophys. 24*, 479-532.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.

REDOR NMR
OF
STABLE-ISOTOPE-LABELED PROTEIN BINDING SITES

JACOB SCHAEFER

Department of Chemistry
Washington University
St. Louis, MO 63130

REDOR and TEDOR

Rotational-echo, double-resonance (REDOR) NMR, a new analytical spectroscopic technique for solids spinning at the magic angle, has been developed over the last 5 years (Gullion and Schaefer, 1989a). REDOR provides a direct measure of heteronuclear dipolar coupling between isolated pairs of labeled nuclei. In a solid with a ^{13}C-^{15}N labeled pair, for example, the ^{13}C rotational echoes that form each rotor period following a ^{1}H-^{13}C cross-polarization transfer can be prevented from reaching full intensity by insertion of a ^{15}N π pulse each half rotor period. The REDOR difference (the difference between a ^{13}C NMR spectrum obtained under these conditions and one obtained with no ^{15}N π pulses) has a strong dependence on the ^{13}C-^{15}N dipolar coupling, and hence, the ^{13}C-^{15}N internuclear distance (Gullion and Schaefer, 1989b). REDOR is described as *double-resonance* even though three radio frequencies (typically ^{1}H, ^{13}C, and ^{15}N) are used because the protons are removed from the important evolution part of the experiment by resonant decoupling. The dephasing of magnetization in REDOR arises from a local dipolar ^{13}C-^{15}N field gradient and involves no polarization transfer. REDOR has no dependence on ^{13}C or ^{15}N chemical-shift tensors and does not require resolution of a ^{13}C-^{15}N coupling in the chemical-shift dimension (Gullion and Schaefer, 1989b).

In early applications of REDOR to C-N distance determinations in peptides, relatively few dephasing pulses were used so that sometimes large corrections were necessary to account for the one- and two-bond coupling between the

label and spins in the natural abundance background (Marshall *et al.*, 1990). More recently, increased dephasing has been possible as a result of the development of phase-routing schemes (Gullion and Schaefer, 1991) (Fig. 1) that eliminate the dependence of REDOR determinations on frequency offsets and pulse imperfections. We now routinely use hundreds of dephasing pulses with no significant accumulated error (Fig. 2) (Hing *et al.*, 1994). We have also extended REDOR applications to include ^{2}H, with either direct observation of deuterium or indirect observation through REDOR dephasing (Schmidt *et al.*, 1992).

Sometimes a natural-abundance background cannot be measured or calculated easily. In this situation, we have proposed selecting the dipolar coupled spins from among the background of uncoupled spins by a coherence transfer from one spin of the heteronuclear pair to the other. The pulse sequence for this selection is shown in Fig. 3. Transverse magnetization is first established in the I-spin system by a cross-polarization transfer from abundant protons. The dephasing of I-spin magnetization by dipolar coupled S spins through rotor-synchronized S-spin π pulses leads to a coherence that is transferred by an I,S pair of $\pi/2$ pulses.

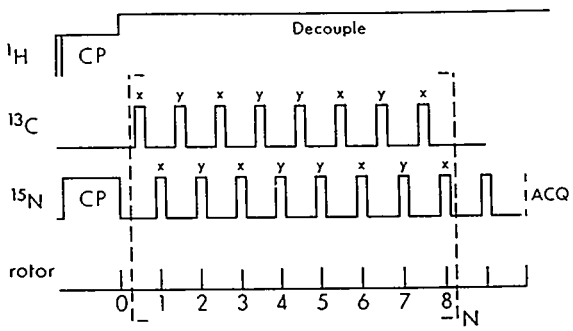


Fig. 1. REDOR pulse sequence with dephasing 1π pulses on both the nitrogen and carbon channels. The pulses are applied using an xy-8 phase-cycling scheme to eliminate offset effects and pulse imperfections. Signal acquisition begins 2 rotor cycles after the completion of the full 8N rotor cycles of dephasing.

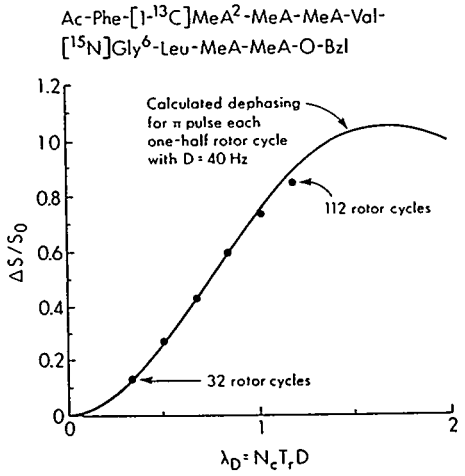


Fig. 2. Observed dependence of ~S/S$_0$ on the universal parameter, λ$_D$, for a ^{13}C-^{15}N double-labeled fragment of emerimicin. The pulse sequence of Fig. 1 was used. The experimental results (●) are in agreement with calculation (solid line) assuming a ^{13}C-^{15}N dipolar coupling of 40 Hz corresponding to a C-N distance of 4.1 ± 0.1Å. The natural-abundance correction to ΔS/S$_0$ at N$_C$ = 8 is 50% and at N$_C$ =100 is 4%. (MeA = methylalanine)

These pulses occur together at the completion of a rotor cycle. Finally, the coherence transferred to the S spins is transformed into observable S-spin transverse magnetization by rotor-synchronized I-spin π pulses. The S-spin signal can be observed either synchronously (one sampling per rotor cycle) or as part of a two-dimensional experiment in which a full S-spin spectrum is obtained by normal acquisition starting with a rotor cycle. We call the entire procedure *transferred-echo double-resonance,* or TEDOR (Hing *et al.*, 1992). TEDOR is a rotor-synchronized solid-state experiment that is based on some of the same coherence-transfer principles as the INEPT solution-state experiment.

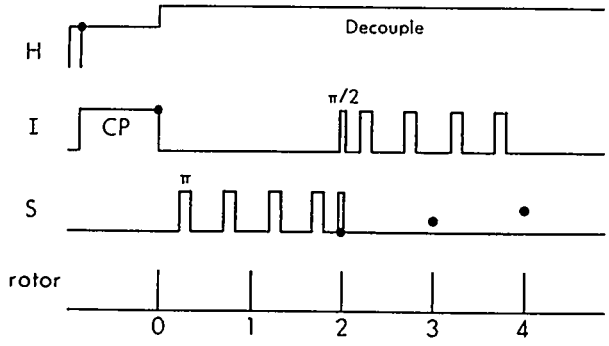


Fig. 3. Pulse sequence for transferred echo, double-resonance I-S NMR. Following a cross-polarization transfer to generate magnetization, the protons are removed by resonant decoupling. The solid circles represent observable magnetization. The π pulses are shown separated by half rotor cycles. All pulses are assumed to be applied on resonance. If there are multiple shifts for either I or S spin systems, additional refocusing pulses are required.

TEDOR can be used alone or in combination to make highly selective TEDOR-TEDOR and TEDOR-REDOR experiments (Holl *et al.*, 1992). The latter is illustrated in Fig. 4. A TEDOR sequence is tailored to select the carbonyl-carbon signal of residue 4 of a $^{19}F,^{13}C,^{15}N$ triple-labeled emerimicin fragment. The ^{15}N-^{13}C selected magnetization is then dephased by REDOR ^{19}F π pulses, resulting in the

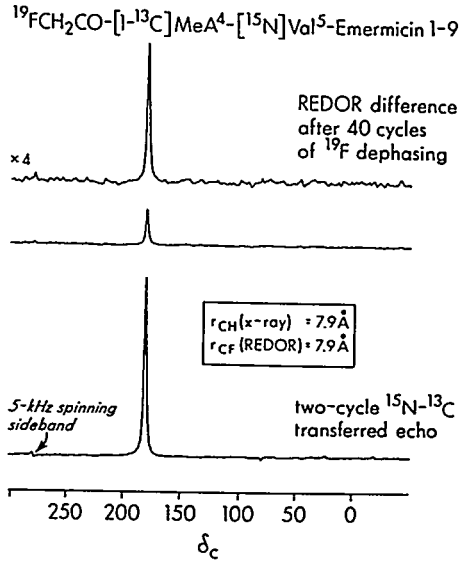


Fig. 4. TEDOR-REDOR ^{13}C NMR spectrum of a triple-labeled peptide. The TEDOR spectrum (bottom) was obtained using a ^{15}N-^{13}C coherence transfer sandwiched between two rotor cycles of dephasing and refocusing pulses. This is an optimum transfer for a directly bonded ^{13}C-^{15}N pair. The TEDOR-REDOR difference spectrum (the difference between spectra with and without dephasing pulses) is shown at the top of the figure. This spectrum was obtained by dephasing the ^{15}N TEDOR-selected ^{13}C magnetization by 40 rotor cycles of ^{19}F π pulses.

determination of a selected 8-Å C-F distance. These combination experiments involve four radio frequencies in the same experiment, which we achieved using transmission-line technology (Holl *et al.*, 1990). All tuning elements in our probes are remote from the immediate vicinity of the coil, which makes handling high power on four channels practical. In addition, there is no significant sacrifice in sensitivity: the ^{13}C channel of a ^{1}H-^{13}C double-tuned transmission line probe has 90% of the efficiency of a conventional probe with matching and tuning capacitors at the coil (McKay, 1993). A ^{1}H-^{19}F-^{31}P-^{13}C four-channel probe has also been built that allows ^{19}F detection with ^{1}H dipolar decoupling. This means that REDOR detection of rare-spin ^{13}C-^{19}F labeled proteins, for example, can be performed with the sensitivity advantage of the high-γ nucleus.

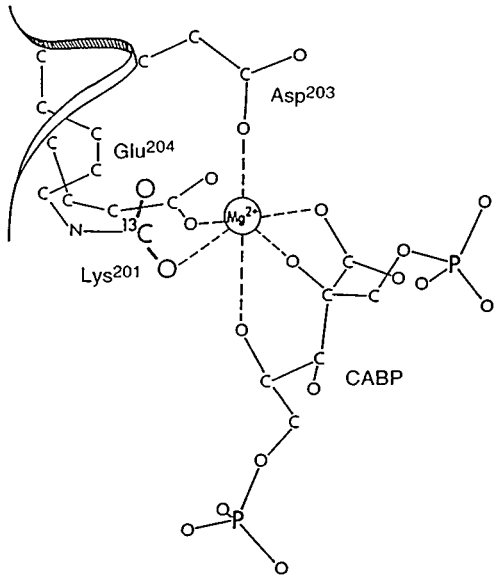


Fig. 5. Geometry of the RuBisCo binding site showing the orientation of CO_2 to the inhibitor, 2-carboxy-D-arabinito 1,5 bisphosphate (CABP). Crystallography has shown that three residues and magnesium are needed to stabilize the complex.

Confirmation of the REDOR Method for a Complex of Ribulose 1,5-Bisphosphate Carboxylase

The application of REDOR to a complicated protein complex is illustrated for the CO_2 binding site of ribulose 1,5-bisphosphate carboxylase (RuBisCo), a 540-kD enzyme with 8 equivalent subunits, as shown in Fig. 5 (Knight *et al.*, 1990). The resonance from bound $^{13}CO_2$ is resolved at 75 MHz (Fig. 6, bottom). REDOR dephasing by the two ^{31}P's of the inhibitor in the binding site agrees with expectations based on x-ray-determined ^{13}C-^{31}P distances (Fig. 6) (Mueller *et al.*, in preparation)

Uniform ^{15}N Labeling and Glutamine Binding Protein

Uniform ^{15}N-labeling of a protein complex makes it possible to use REDOR and TEDOR to measure distances from ^{31}P, ^{19}F, and ^{13}C labels in substrates and inhibitors to sidechain nitrogens in arginine, histidine, and lysine residues. In addition, REDOR difference signals from the peptide backbone are sometimes resolved as well. Uniform ^{15}N-labeling of proteins is inexpensive and relatively easy to do. Distance information about the binding sites of noncrystallizable

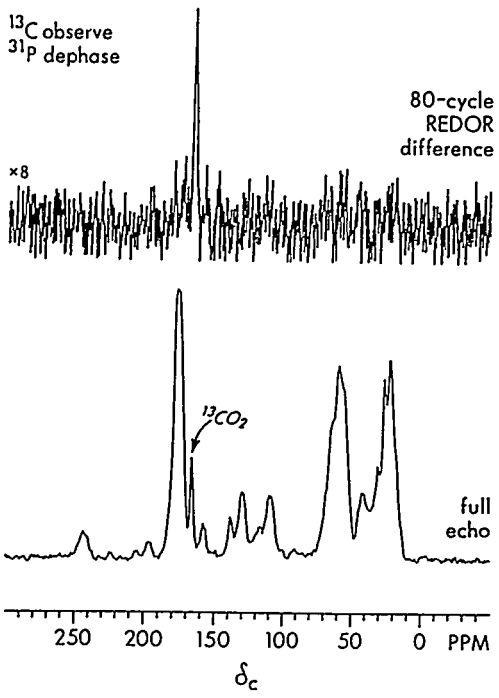


Fig. 6. 75-MHz ^{13}C NMR REDOR spectra of lyophilized RuBisCo with bound $^{13}CO_2$ and CABP after 80 rotor cycles of ^{31}P dephasing with 5-kHz magic-angle spinning. The only significant REDOR difference signal arises from the CO_2. REDOR determines an average r_{cp} of 7.5 Å, compared to an x-ray value of 7.7 Å.

proteins and protein complexes can therefore be obtained with specific labeling limited to substrates and inhibitors.

Hing *et al.* (1994) measured carbon-nitrogen distances up to 6.3 Å between ^{13}C labels in L-glutamine and histidine and lysine residues of uniformly ^{15}N-labeled glutamine-binding protein (GlnBP), an essential component of the glutamine transport system in *E. coli*. No crystal structure for the complex is presently available. Altogether, seven distance constraints have been established by REDOR (two of which are illustrated in Fig. 7) and these have unambiguously determined the ligand orientation with respect to the imidazole ring of His^{156}.

Characterization of the Binding Site of Enolpyruvylshikimate-3-phosphate Synthase

The 46-kDa enzyme 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase catalyzes the reversible condensation of shikimate-3-phosphate (S3P) and phosphoenolpyruvate (PEP) to form EPSP (Fig. 8) in the synthesis of aromatic amino acids in plants and microorganisms (Anderson and Johnson, 1990). This reaction is inhibited by the commercial herbicide N-(phosphonomethyl)glycine (glyphosate or Glp), $HO_3PCH_2NHCH_2COOH$, which, in the presence of S3P, binds to EPSP synthase and forms a stable, ternary complex. A crystal structure for EPSP synthase has been published (Stallings *et al.*, 1991), but there is no structure for the ternary complex, which does not form crystals suitable for diffraction studies. Our strategy for the characterization of the ternary complex is in two parts. First, ^{13}C and ^{15}N labels are introduced into Glp and distances are measured by REDOR from these labels to ^{31}P in S3P and Glp. These distances will help define the geometry of the two substrates relative to one another. Second, ^{15}N labels are introduced into the basic

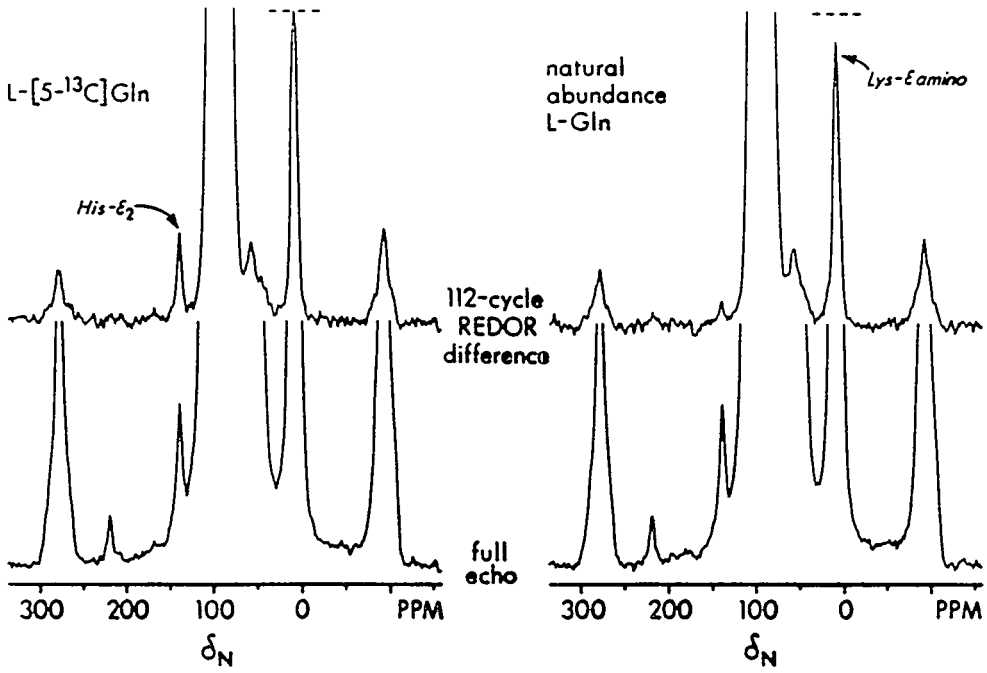


Fig. 7. ^{15}N NMR REDOR spectra of lyophilized, uniformly ^{15}N-labeled glutamine-binding protein complexed to L-[5-^{13}C]glutamine (left) and natural-abundance glutamine (right) after 112 rotor cycles of ^{13}C dephasing with 3.759-kHz magic-angle spinning. The ε nitrogen of the single histidine residue of the 25-kDa protein has a $\Delta S/S_0$ (after natural-abundance corrections have been made) corresponding to a 4.1-Å distance from the glutamine ^{13}C label. At least one lysine residue is within 4.3 Å of the ^{13}C label as well.

residues of the protein (Lys, Arg, and His) to determine which are near the ^{31}P of the negatively charged phosphate and phosphonate groups of S3P and Glp, respectively. The protein labeling is done using an *E. coli* expression system capable of producing 50 mg of purified protein from 1 ℓ of defined media containing 100 mg/ℓ of ^{15}N-labeled Lys, Arg, or His.

Christensen and Schaefer (1993) observed high-resolution ^{31}P NMR spectra of lyophilized powders of the ternary complex when the complex is quick frozen from dilute solution (50 μM or less in protein and 2 mM in buffer). An example is shown in Fig. 9. The solid-state and solution-state phosphorous chemical shifts match (Christensen and Schaefer, 1993), indicating that the local pH and extent of hydration of the binding site are unaffected by the lyophilization; this is consistent with preservation of the native conformation of the binding site.

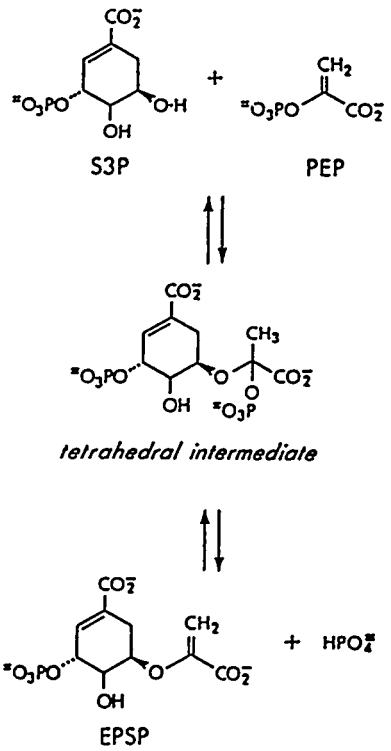


Fig. 8. Reaction scheme for the condensation of shikimate-3-phosphate and phosphoenolpyruvate catalyzed by EPSP synthase.

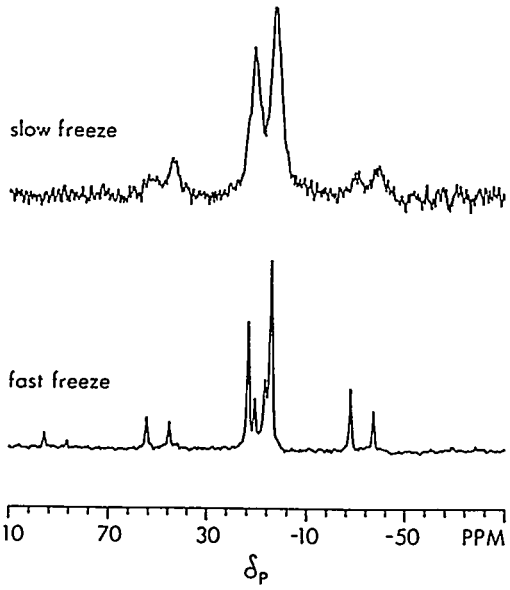


Fig. 9. A 121.3-MHz CPMAS ^{31}P NMR spectra of EPSPS-S3P-Glp ternary complex frozen slowly from 1-mM solution (top) and quick frozen from 50-μM solution (bottom). The substrates were present in slight excess relative to protein. The ^{31}P chemical shift of bound S3P is 2 ppm upfield from that of free S3P. The ^{31}P chemical shift of bound glyphosate is 3 ppm downfield from that of free glyphosate. The sharp well-resolved ^{31}P NMR spectrum of the quick-frozen lyophilized ternary complex indicates binding-site homogeneity.

We believe that individual protein molecules are encased in a buffer glass. This method of sample preparation was also used for both the RuBisCo and GlnBP complexes described above. REDOR experiments (Fig. 10) with ^{31}P observation and ^{13}C dephasing give two ^{31}P distance measurements to each ^{13}C label—one from S3P and one from Glp (Christensen and Schaefer, 1993). The contribution of the natural-abundance ^{13}C background to the dephasing is taken into account in a separate experiment using unlabeled Glp. Altogether, there are 6 useful distance measurements (McDowell *et al.*, in preparation), which indicate that Glp has an extended conformation in the ternary complex (P to C1 distance of 4.9 ± .2 Å compared to a maximum allowed value of 5.11 Å) and that S3P and Glp are in proximity (P to Glp distances vary between 6 and 7 Å, depending on the Glp label).

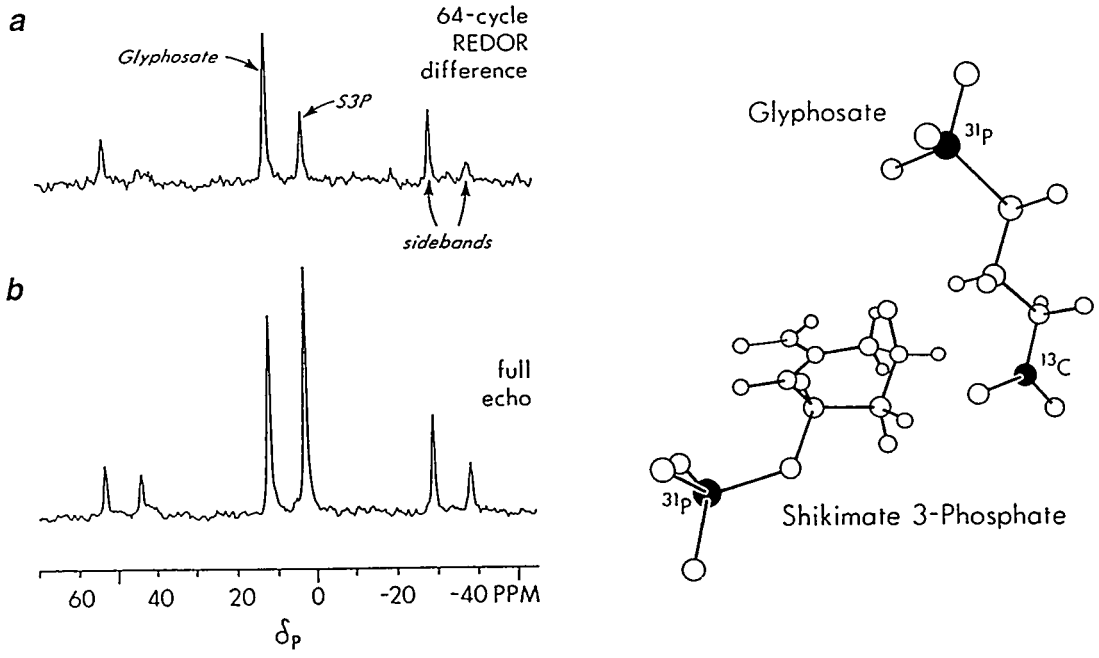


Fig. 10. ^{31}P NMR spectra of EPSPS-S3P-[1-^{13}C]Glp lyophilized all-bound ternary complex with ^{13}C dephasing pulses (a) and without dephasing pulses (b) after 64 rotor cycles with 5-kHz magic-angle spinning. The Glp ^{31}P is closer to the ^{13}C label than is the S3P ^{31}P as evidenced by the larger –ΔSIS for glyphosate. The S3P-Glp drawing on the right is a cartoon.

The phosphate of S3P and the phosphonate of Glp carry double negative charges, and the carboxyl carbons of the two substrates each carry a single negative charge (Anderson and Johnson, 1990). Thus, on the order of 6 basic residues must be in the EPSPS binding site for charge balance. Some of these basic residues are likely to be lysines. Figure 11 shows the 30.3-MHz CPMAS ^{15}N NMR spectrum of [ε-^{15}N]Lys-EPSPS-S3P-Glp ternary complex. The ε nitrogens of the 17 lysines (Stallings *et al.*, 1991) of EPSPS have been more than 95% isotopically enriched. A fraction of the lysine ^{15}N amine peak at 10 ppm is shifted to low field, and the shifted component is responsible for both of the lines observed in the REDOR difference spectrum (^{15}N observe, ^{31}P dephase; Fig. 12, top) (McDowell *et al.*, in preparation). An estimated spin count for the shifted lines based on the dependence of the REDOR difference intensities on the number of dephasing cycles suggests that three lysines are within 4 Å of the ^{31}P's of S3P and Glp. To decide on the distribution of lysines between the two ^{31}P sites, we used

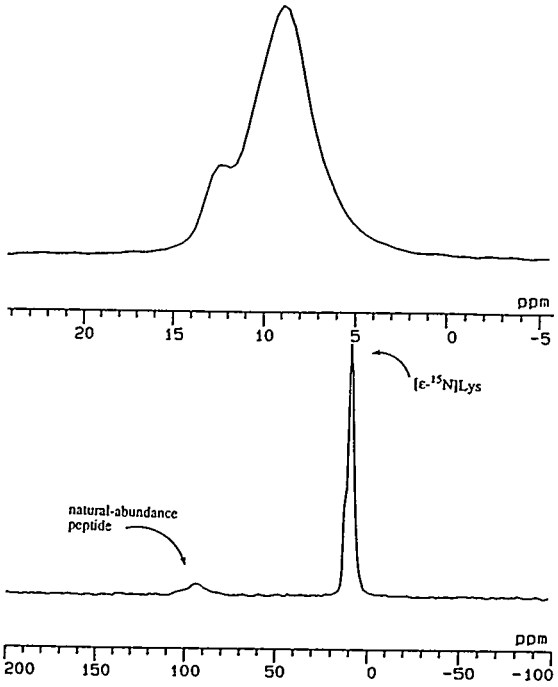


Fig. 11. CPMAS ^{15}N NMR full spectrum (bottom) and expanded lysine-region spectrum (top) of a lyophilized EPSPS S3P-Glp ternary complex. The EPSPS was overexpressed by engineered E. coli *grown on media containing 100 mg/ℓ of* L*-[ε-^{15}N]lysine.*

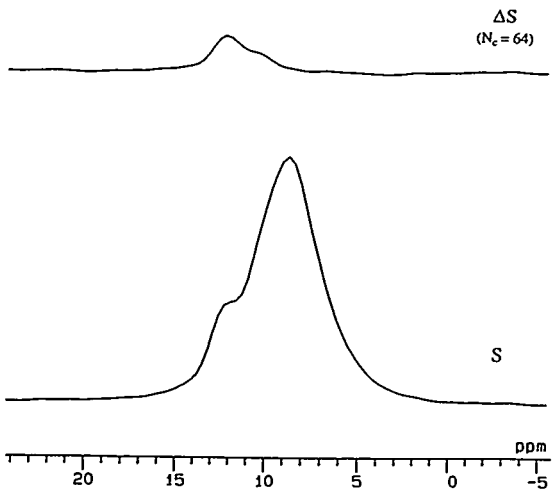


Fig. 12. REDOR ^{15}N NMR spectra of the ternary complex of Fig. 11 after 64 rotor cycles of ^{31}P dephasing with 5-kHz magic-angle spinning. The REDOR difference spectrum (top) arises exclusively from lysine lines shifted to lowfield.

a selective P—>N TEDOR coherence transfer (McDowell *et al.*, in preparation). In this experiment, the sign of the ^{31}P magnetization from Glp is altered every other scan by using a rotor-synchronized delay. Thus, the coherence transferred from the Glp ^{31}P to any nearby lysine ^{15}N alternates in sign from one scan to the next, as does the observed ^{15}N magnetization. Alternate scans are stored in separate buffers throughout the experiment and are subsequently added and subtracted. The sum gives a lysine ^{15}N spectrum arising only from those lysines near S3P, while the difference gives a spectrum arising only from those lysines near Glp. This is a version of a 2D experiment connecting chemical shifts with dipolar couplings. We conclude that there is one lysine near S3P and two near Glp. After TEDOR has established the stoichiometry, quantitative REDOR distance measurements are possible (Gullion and Schaefer, 1989b; Hing *et al.*, 1992) and have established that ε nitrogens of each of the two lysines near Glp are 4.0 Å from the Glp ^{31}P, whereas the ε nitrogen of the single lysine near S3P is 3.7 Å from the S3P ^{31}P.

We have also performed the kinds of REDOR and TEDOR experiments described above for [*ring*-^{15}N]His-EPSPS-S3P-Glp and [*guanidino*-^{15}N]Arg-EPSPS-S3P-Glp (McDowell *et al.*, in preparation). The combination of all of the REDOR/TEDOR experiments provides 16 distance constraints on the binding site: 6 between nuclei of S3P and Glp and 10 between the ^{31}P's of S3P and Glp and the ^{15}N's of lysine, arginine, and histidine residues of EPSPS. Using these constraints and the crystal coordinates of free EPSPS, we have found only a few plausible structures for the ternary complex (McDowell *et al.*, in preparation). We plan to confirm and refine the model for the complex by making additional REDOR measurements on [ε-^{15}N, ε-^{13}C]Lys-EPSPS-S3P-Glp and [ε-^{15}N]Lys-EPSPS-S3P-[l(or 2 or 3)-^{13}C]Glp, observing and dephasing with various combinations of ^{31}P, ^{15}N, and ^{13}C. The results of these experiments should define a sufficient number of distance constraints so that molecular dynamics computer modeling (starting with the coordinates of EPSPS, docking the S3P-Glp pair, and imposing the TEDOR/REDOR determined distance constraints) will establish unambiguously the structure of the ternary complex. An internal self-consistency check on this structure will involve using resolved ^{15}N to identify chemical shifts of all of the peptide nitrogens that are within 6 Å of ^{31}P in S3P and Glp as well as employing ^{31}P REDOR to identify dephasing of an ^{15}N uniformly labeled ternary complex. These experiments outline a general strategy for determining the structure of a noncrystallizable protein complex, using solid-state NMR if the crystal coordinates of the uncomplexed protein (or a homologue) are available.

Acknowledgement

This work was supported by NIH Grant 40634.

References

Anderson, K.S. and Johnson, K.A. (1990) *Chem. Rev. 90*, 1 1.

Anderson, K.S., Sikorski, J.A., and Johnson, K.A. (1988) *Biochemistry27*, 7395.

Christensen, A.M. and Schaefer, J. (1993) *Biochemistry 27, 32*, 2868.

Gullion, T. and Schaefer, J. (1989a) *J. Magn. Res. 1*, 196.

Gullion, T. and Schaefer, J. (1989b) *Adv. Magn. Res. 13*, 55.

Gullion, T. and Schaefer, J. (1991) *J. Magn. Reson. 92*, 439.

Hing, A.W., Vega, S., and Schaefer, J. (1992) *J. Magn. Reson. 96*, 205.

Hing, A.W., Tjandra, N., Cottam, P.F., Schaefer, J., and Ho, C. (1994) *Biochemistry 33*, 8651.

Holl, S.M., Marshall, G.R., Beusen, D.D., Kociolek, K., Redlinski, A.S., Leplawy, M.T., McKay, R.A., Vega, S., and Schaefer, J. (1992) *J. Am. Chem. Soc. 114*, 4830.

Holl, S.M., McKay, R.A., Gullion, T., and Schaefer, J. (1990) *J. Magn. Reson. 89*, 620.

Knight, S., Andersson, I., and Branden, C.-I. (1990) *J. Biol. Chem. 215*, 113.

Marshall, G.R., Beusen, D.D., Kociolek, K., Redlinski, A.S., Leplawy, M.T., Pan, Y., and Schaefer, J. (1990) *J. Am. Chem. Soc. 112*, 963.

McKay, R.A. (1993) "Efficiency of Transmission-Line Probes," Plenary presentation, 34th ENC, St. Louis.

Schmidt, A., McKay, R.A., and Schaefer, J. (1992) *J. Magn. Reson. 96*, 644.

Stallings, W.C., Abdel-Meguid, S.S., Lim, L.W., Shief, H.S., Dayringer, H.E., Leimgruber, N.K., Stegman, R.A., Anderson, K.S., Sikorski, J.A., Padgette, S.R., and Kishore, G.M. (1991) *Proc. Nat. Acad. Sci. USA 88*, 5046.

Stable-Isotope-Labeled Carbohydrates and Nucleosides: Synthesis and Applications in Chemistry and Biology

Anthony S. Serianni

Department of Chemistry and Biochemistry
University of Notre Dame
Notre Dame, IN 46556

Carbohydrates play important roles in many key biochemical processes in living cells. For example, they are metabolized to produce energy, mediate cell-cell recognition, and play an indirect role (as constituents of DNA and RNA) in DNA replication, RNA transcription and protein synthesis. These roles, and others of comparable biochemical significance, have been studied to varying extents with the use of stable isotopically labeled molecules, usually in conjunction with NMR spectroscopy and/or mass spectrometry. For example, carbohydrate metabolism has been monitored *in vitro* and *in vivo* with the use of isotopically labeled compounds (Beckmann *et al.*, 1991; Rothman *et al.*, 1985; Podlasek and Serianni, 1994; Kukal *et al.*, 1988, 1989). Molecular aspects of cell-cell recognition, mediated by cell-surface glycoproteins and glycolipids, have been probed through NMR studies of isotopically labeled oligosaccharides (Nunez and Barker, 1980; Rosevear *et al.*, 1982; Ichikawa *et al.*, 1992). More recently, the solution behavior of DNA and RNA has been examined through the use of labeled oligonucleotides (Hines *et al.* , 1993; Nikonowicz and Pardi, 1992; Pardi and Nikonowicz, 1992; Legault *et al.*, 1994; Marino *et al.*, 1994; Lancelot *et al.*, 1993; Kellenbach *et al.*, 1992; Sklenar *et al.*, 1993). In all of these pursuits, the effort and expense to prepare labeled molecules, both of which can be substantial, are more than offset by the wealth of information derived from these studies. This information often cannot be accessed, or can be accessed only with great difficulty, using natural (unlabeled) compounds.

The recent advances in three-dimensional (3D) structure determination of proteins by NMR can be, in part, attributed to the use of ^{13}C and/or ^{15}N enrichment (Otting and Wüthrich, 1990; Fesik and Zuiderweg, 1990; Clore and Gronenborn, 1991). Isotopic labeling not only permits

signal discrimination and assignment through heteronuclear multidimensional NMR techniques, but also allows the measurement of spin-coupling constants (for example, ^{13}C-^{1}H, ^{15}N-^{1}H) and nuclear spin-relaxation rates that may be related to backbone and sidechain conformation and overall/local motion, respectively (Vuister *et al.* , 1992; Mierke *et al.*, 1992; Palmer *et al.*, 1991; Kushlan and LeMaster, 1993). Similar benefits have been derived from studies of labeled oligosaccharides and oligonucleotides, although progress to date has been more modest than that for proteins.

More widespread use of stable isotopes (for example, ^{2}H, ^{13}C, ^{15}N) in structural studies has been hindered by high cost and less-than-convenient access to desired labeled compounds. Cost is determined by the price of the primary isotope (for instance, ^{13}CO), costs of conversion of the primary isotope to labeled precursors (such as $^{13}CO_2$, $K^{13}CN$), and costs for the conversion of labeled precursors to more complex biological compounds. This article discusses chemical and enzymic methods to prepare biologically important monosaccharides containing ^{13}C, ^{2}H and/or $^{17,18}O$ isotopes and routes to assemble these labeled monomers into isotopically labeled oligosaccharides and oligonucleotides. In addition, some examples taken from our work will be discussed in which stable isotopes have been used to probe the solution behavior of monosaccharides, oligosaccharides, and oligonucleotides.

Approaches to Labeling of Carbohydrates: Guiding Principles

The high cost of isotopic labeling, especially ^{13}C-labeling, demands access to synthetic methods that are reliable and efficient. In addition, synthetic routes are most attractive when the isotope is introduced late in the reaction sequence in order to maximize yield with respect to the isotope. In this laboratory, synthetic routes to labeled sugars commonly involve a combination of chemical and enzymic steps. The diverse roles of carbohydrates in cellular processes make this approach realistic, because a wide variety of enzyme-catalyzed reactions is available for potential use in the synthesis of labeled sugars.

The attractiveness of enzymic approaches lies in their normally high yields and specificity in comparison to related chemical methods. However, the convenience of an enzymic route will be determined, in part, by the availability of the required enzyme(s). A survey of *Enzyme Nomenclature* (1992) reveals many potentially useful enzymes for labeled carbohydrate synthesis, but some of these are not commercially available, are not isolable from convenient sources, have low activities, and/or have tedious purification protocols. Of course, for many laboratory syntheses, the enzyme preparation(s) does not need to be homogeneous. Often one or more contaminating activities do not interfere with the synthesis if cofactors necessary for these side-reactions can be excluded from the reaction mixture. It should be appreciated,

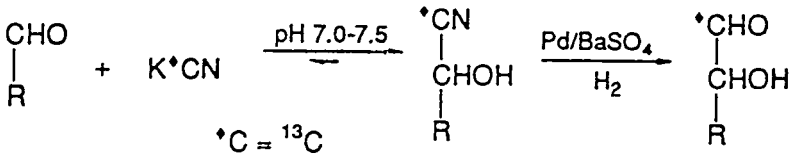
however, that enzyme-catalyzed reactions are not always superior to chemical transformations. There are, and will continue to be, efficient chemical reactions that are easy to perform on large scales and give excellent yields. Thus, optimal synthetic protocols for the synthesis of simple labeled monosaccharides and their assembly into oligomers (for example, oligosaccharides, oligonucleotides) will frequently be composed of integrated chemical and enzymic reactions.

Cyanohydrin Reduction Reaction

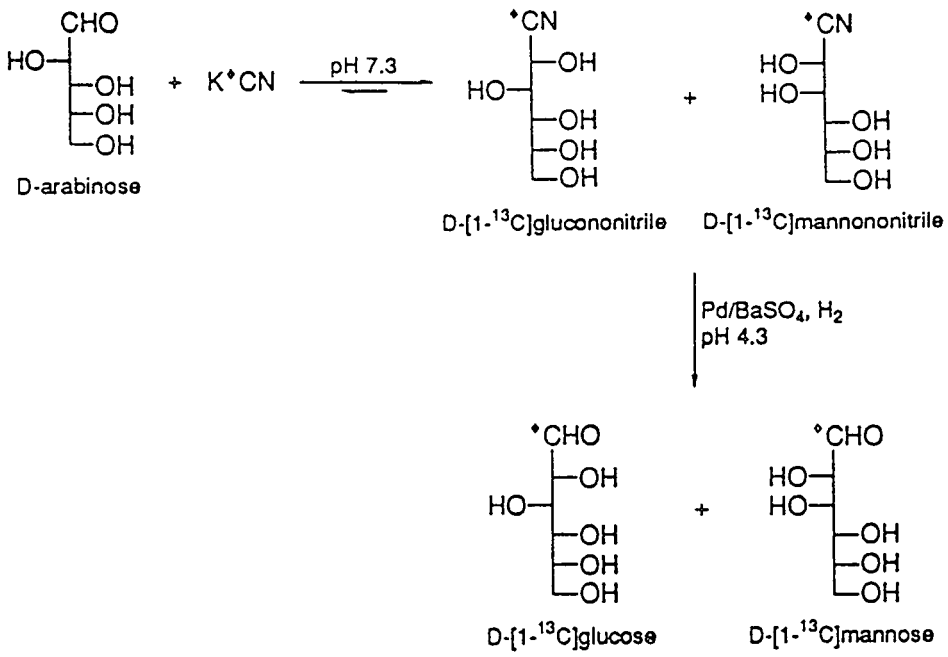
Isotopes of carbon, hydrogen, and oxygen can be introduced into aldoses by means of the cyanohydrin reduction reaction (Serianni *et al.*, 1979a, 1990) (Scheme 1). A parent aldose is treated with $K^{13}CN$ at pH 7.3 to generate C2 epimeric [1-^{13}C]aldononitriles (cyanohydrins) in high yield. According to the classical Kiliani-Fischer chain-extension reaction (Fischer, 1889), these nitriles are hydrolyzed to the corresponding aldonic acids, and the latter are lactonized and reduced to the product C2 epimeric aldoses. In the cyanohydrin reduction scheme, however, the nitriles are reduced directly to aldoses with a palladium catalyst ($Pd/BaSO_4$) and H_2 at pH 1.7-4.3. Thus, for example, application of this reaction with D-arabinose as the parent sugar yields a mixture of D-[1-^{13}C]mannose (70%) and D-[1-^{13}C]glucose (30%) (Scheme 2). The ratio of epimeric aldoses depends on the structure of the starting sugar (Serianni *et al.*, 1979a). The reduction proceeds in acceptable yields (80%), and the sole by-products (1-amino-1-deoxy-alditols) are removed by treatment with cation exchange resin (Serianni *et al.*, 1979a, 1990). The product C2 epimeric aldoses are separated by chomatography on Dowex 50 x 8 (200 – 400 mesh) ion-exchange resin in the Ca^{2+} or Ba^{2+} form (Angyal *et al.*, 1979; Jones and Wall, 1960).

Short- and long-chain parent aldoses may be used in the cyanohydrin reduction reaction, since lactonization is not required. This versatility has been extensively documented over the past decade (Serianni *et al.*, 1979b, 1979c; Snyder and Serianni, 1987a, 1987b, 1991; King-Morris *et al.*,1988; Vuorinen and Serianni, 1990a, 1990b; Wu and Serianni 1991a). Parent

Scheme 1



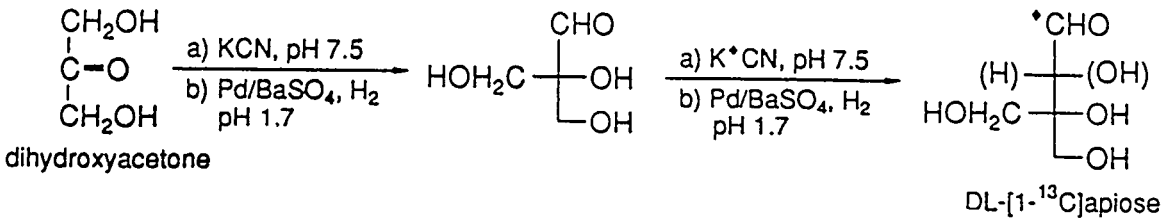
Scheme 2



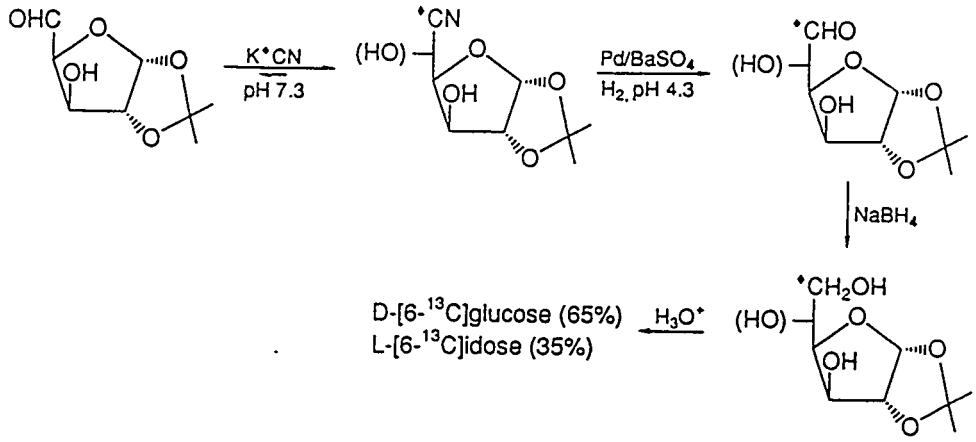
aldoses ranging from formaldehyde to reducing disaccharides have been used as reactants. The optimal solution conditions for Pd-catalyzed reduction depend on the structure of the nitrile, with lower pH values (pH 1.7) preferred for short-chain nitriles and higher values (pH 4.3) for long-chain nitriles. The reaction is not restricted to aldoses as reactants; any carbonyl-containing compound can, in principle, be employed, although the extent to which the equilibrium favors nitrile formation is affected by the structure of the parent aldehyde or ketone. For example, branched-chain sugars have been prepared using ketones as reactants (Snyder and Serianni, 1987b) (Scheme 3), and dicarbonyl compounds have been used to generate aldoses labeled at the terminal hydroxymethyl carbon (for example, D-[6-^{13}C]glucose) (Scheme 4) (King-Morris *et al.*, 1988). An automated chemical synthesizer has been designed to perform the cyanohydrin reduction reaction with minimal human intervention (Stafford *et al.*, 1990), and the reaction can be scaled to accommodate the synthesis of kilogram quantities of labeled aldoses.

A potential drawback of the cyanohydrin reduction reaction is the production of epimeric product mixtures (for example, D-glucose and D-mannose; D-galactose and D-talose), which compromises overall yields. This problem can be minimized through the application of molybdate-catalyzed epimerization and/or chemical or enzymic isomerization/epimerization.

Scheme 3



Scheme 4



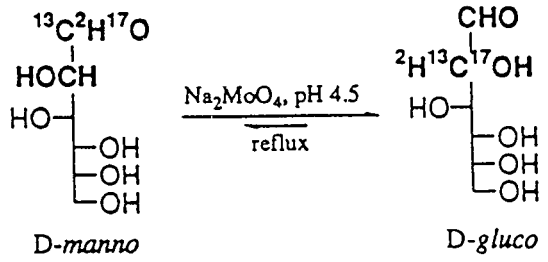
In the latter instance, base-catalyzed isomerization (Lobry DeBruyn-van Ekenstein reaction) (Lobry De Bruyn and van Ekenstein, 1895) is performed in the presence of complexing agents (usually phenylboronic acid) to prevent degradation of the sugar during the reaction (Barker *et al.*, 1977). It has been shown that base-catalyzed isomerization in the absence of these agents can lead to label scrambling; for example, treatment of D-[1-^{13}C]mannose in mild base produces small quantities of [6-^{13}C]glucose, [6-^{13}C]mannose and [6-^{13}C]fructose, presumably through the formation of enediol intermediates (King-Morris and Serianni, 1986). Phenylborate-mediated isomerization typically generates a mixture containing the two C2 epimeric aldoses and the corresponding 2-ketose (for example, D-mannose, D-glucose, D-fructose), which are subsequently separated by chromatography. The reaction can be applied on a large scale and proceeds in good yield (Serianni *et al.*, 1990). Alternatively, molybdate-catalyzed epimerization may be used to interconvert aldoses and generate new labeling patterns simultaneously, as discussed below.

Molybdate-Catalyzed Epimerization

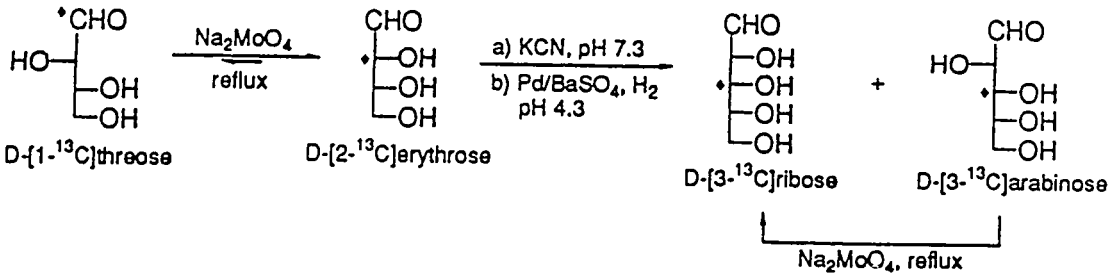
Molybdate-catalyzed epimerization (Hayes *et al.*, 1982a) is an example of a chemical transformation having characteristics similar to those commonly associated with enzyme-catalyzed reactions, namely, it is simple to perform, reliable, proceeds in high yield, and is highly stereospecific. The reaction reduces the effect of producing less desirable epimers from the cyanohydrin reduction reaction. For example, the synthesis of D-[1-^{13}C]glucose is accompanied by the formation of its less desirable C2 epimer, D-[1-^{13}C]mannose, with the latter formed in significantly greater amounts. Molybdate-catalyzed epimerization of D-[1-^{13}C]mannose provides a convenient route to a more desirable product (glucose) and permits ready access to aldoses labeled at C2. The latter is achieved because the reaction not only causes epimerization at C2, *but is also accompanied by C1—C2 transposition* (Scheme 5). Thus, treatment of D-[1-^{13}C]-mannose with a catalytic amount of sodium molybdate in aqueous solution at pH 4.5 for 3 hours under reflux results in an equilibrium mixture of D-[1-^{13}C]-mannose and D-[2-^{13}C]glucose (Hayes *et al.*, 1982a). These aldoses are separated by chromatography on cation-exchange resins (Angyal *et al.*, 1979; Jones and Wall, 1960). The mechanism of this unique reaction has been explored in some detail and apparently involves the formation of a complex between dimolybdate and the acyclic aldehyde form of the aldose (Hayes *et al.*, 1982a). In addition to the aldehyde group at C1, hydroxyl groups at C2 and C3 are absolutely required for the reaction, whereas an OH group at C4 is not required but facilitates the reaction. The reaction can be conducted with molybdate ion free in solution (Hayes *et al.*, 1982a) or complexed to a resin (Clark *et al.*, 1986), or may be conducted in non-aqueous solution with appropriate molybdate complexes (Hayes *et al.*, 1982a). It is interesting to note that a similar epimerization reaction accompanied by C1—C2 transposition is catalyzed by Ni^{2+}-complexes (London, 1987; Tanase *et al.*, 1988).

The synthetic options made available through the application of molybdate epimerization are extensive, as the reaction permits ready incorporation of ^{13}C (and ^{2}H and ^{17,18}O as discussed below) labels at internal sites within carbohydrates. Thus, for example, D-[3-^{13}C]ribose can be prepared by epimerizing D-[1-^{13}C]-threose (generated by cyanohydrin reduction using D-glyceraldehyde as the parent

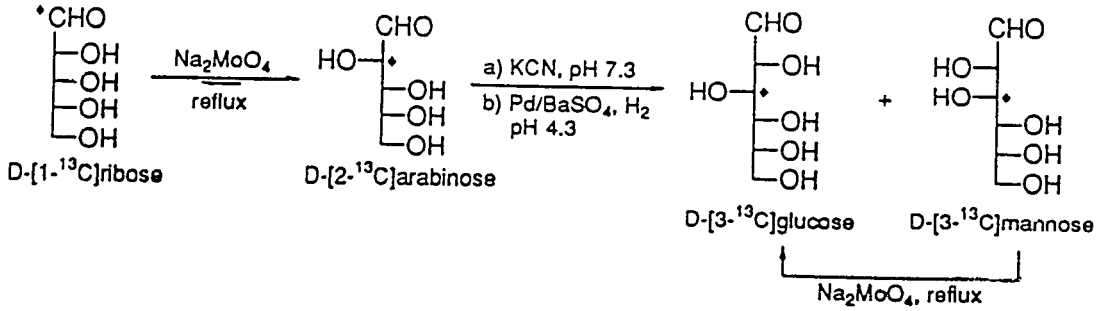
Scheme 5



Scheme 6



Scheme 7

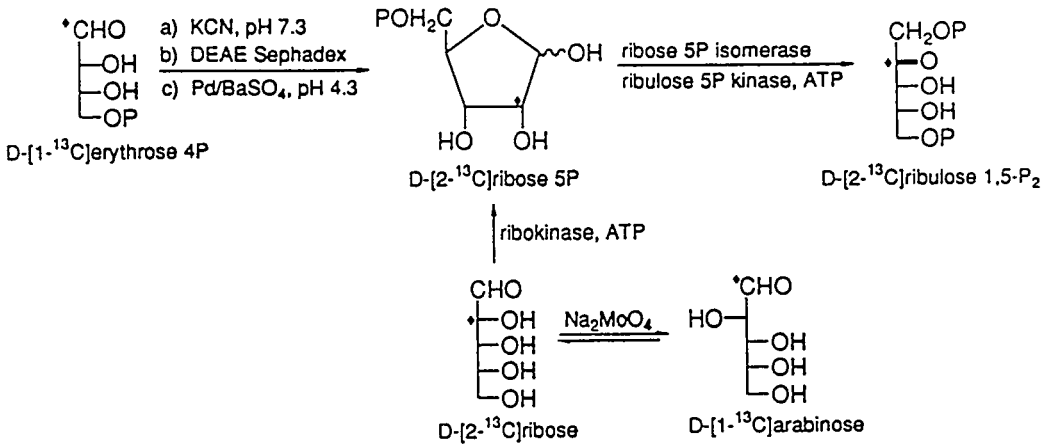


aldose) to D-[2-^{13}C]erythrose and by extending the latter by one carbon to give D-[3-^{13}C]ribose (Scheme 6). Note that the by-product in this case, D-[3-^{13}C]-arabinose, may be epimerized to improve the yield of the *ribo* epimer. Likewise, D-[3-^{13}C]glucose may be prepared using a similar strategy starting from D-[2-^{13}C]-arabinose (Scheme 7). Although D-[3-^{13}C]-glucose might be prepared enzymically from [3-^{13}C]dihydroxyacetone phosphate (DHAP) (or a DHAP metabolic precursor) by using aldolase as discussed below, the integrated cyanohydrin reduction/molybdate epimerization route can be performed with greater ease on a large scale and is more economical.

Integration of Cyanohydrin Reduction and Molybdate Catalyzed Epimerization into Chemi-Enzymic Synthesis

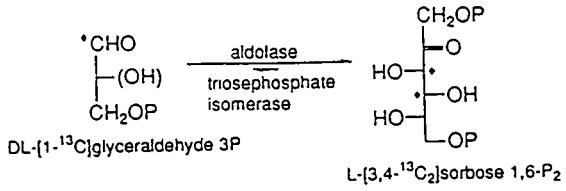
The above reactions have been integrated into a number of chemi-enzymic syntheses of biologically relevant compounds. An early application of this approach involved the use of several glycolytic enzymes to prepare D-[1,6-$^{13}C_2$]fructose 1,6-bisphosphate from D-[1-^{13}C]glucose (Nunez *et al.*, 1977). D-[2-^{13}C]Ribose 5-phosphate, generated from the chain extension of D-[1-^{13}C]erythrose 4-phosphate, was treated with ribose 5-phosphate isomerase (E.C. 5.3.1.6) and ribulose 5-phosphate kinase (E.C. 2.7.1.19) to produce D-[2-^{13}C]ribulose 1,5-bisphosphate, a substrate of the enzyme, ribulose

Scheme 8



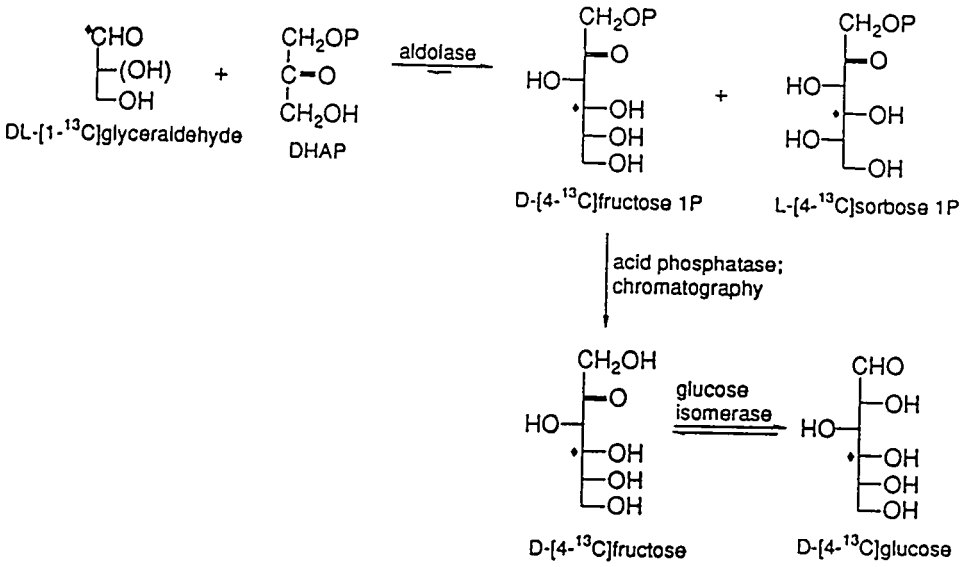
bisphosphate carboxylase/oxygenase (E.C. 4.1.1.39) (Serianni *et al.*, 1979c) (Scheme 8). DL-[1-^{13}C]Glyceraldehyde was treated with glycerolkinase (E.C. 2.7.1.30) and ATP to prepare L-[1-^{13}C]glyceraldehyde 3-phosphate (Serianni *et al.*, 1979b), and α-D-[1-^{13}C]galactose 1-phosphate was prepared from D-[1-^{13}C]galactose with the use of galactokinase (E.C. 2.7.1.6) and ATP (Serianni *et al.*, 1982a). DL-[1-^{13}C]-Glyceraldehyde 3-phosphate has been treated with triosephosphate isomerase (E.C. 5.3.1.1) (TPI) and fructose 1,6-bisphosphate aldolase (E.C. 4.1.2.13) to generate, after equilibration, L-[3,4-$^{13}C_2$]-sorbose 1,6-bisphosphate (Serianni *et al.*, 1979c) (Scheme 9).

Scheme 9



Chemi-enzymic routes have been used to prepare various isotopomers of D-glucose and D-ribose. For example, D-[4-^{13}C]glucose has been synthesized from DL-[1-^{13}C]glyceraldehyde and DHAP in the presence of aldolase (Serianni *et al.*, 1982a) (Scheme 10); D-[5-^{13}C]glucose is prepared in a similar fashion using DL-[2-^{13}C]glyceraldehyde (Serianni *et al.*, 1979b). Because all ^{13}C-isotopomers of DL-glyceraldehyde can be prepared by consecutive cyanohydrin reduction with [^{13}C]formaldehyde and K^{13}CN, the aldolase route provides access to D-glucose labeled at C4, C5, and/or C6. These labeled forms of D-glucose are precursors in the synthesis of labeled D-ribose (Serianni and Bondo, 1994).

Scheme 10



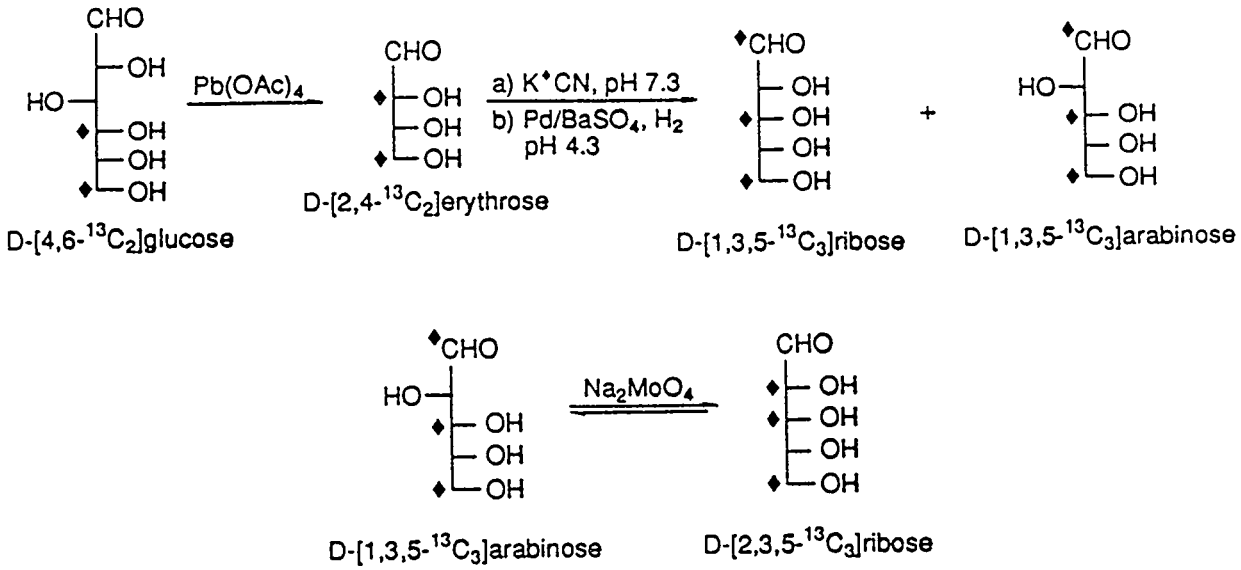
Thus, for example, D-[1,3,5-$^{13}C_3$]ribose can be prepared by treatment of D-[4,6-$^{13}C_2$]-glucose with $Pb(OAc)_4$, giving D-[2,4-$^{13}C_2$]-erythrose in good yield (80%) (Scheme 11). Chain extension of this tetrose gives D-[1,3,5-$^{13}C_3$]ribose. The by-product, D-[1,3,5-$^{13}C_3$]arabinose, can be epimerized with molybdate to yield D-[2,3,5-$^{13}C_3$]-ribose (Scheme 11). Access to 26 of the 32 ^{13}C-isotopomers of D-ribose is possible by integration of cyanohydrin reduction, molybdate-catalyzed epimerization and a few simple carbohydrate transformations, as discussed in a recent review (Serianni and Bondo, 1994). The remaining six ^{13}C-isotopomers may be accessed via chemi-enzymic approaches using appropriately labeled glycerol and/or DL-glyceraldehyde (Serianni and Bondo, 1994) (Scheme 12). D-Mannose labeled at C1 and/or C2 can be converted to D-ribose labeled at C5 and/or C4, respectively (Scheme 13) (Wu *et al.*, 1992a).

A number of ^{13}C- and ^{2}H-labeled oligosaccharides have been prepared with the use of appropriately labeled sugar nucleotides and specific glycosyltransferases. For example, ^{13}C-labeled methyl β-lactosides (Hayes *et al.*, 1982b; Kline *et al.*, 1990) and *N*-acetyllactosamines (Nunez and Barker, 1980; Rosevear *et al.*, 1982) have been prepared with isotopically labeled acceptors (methyl β-D-glucopyranoside or 2-acetamido-2-deoxy-D-glucose) and/or labeled sugar nucleotides (UDP-D-galactose) (Nunez and Barker, 1980) and lactose synthase (E.C. 2.4.1.22) (with or without α-lactalbumin present in the reaction mixture) (Scheme 14). Several L-fucose-containing trisaccharides have been

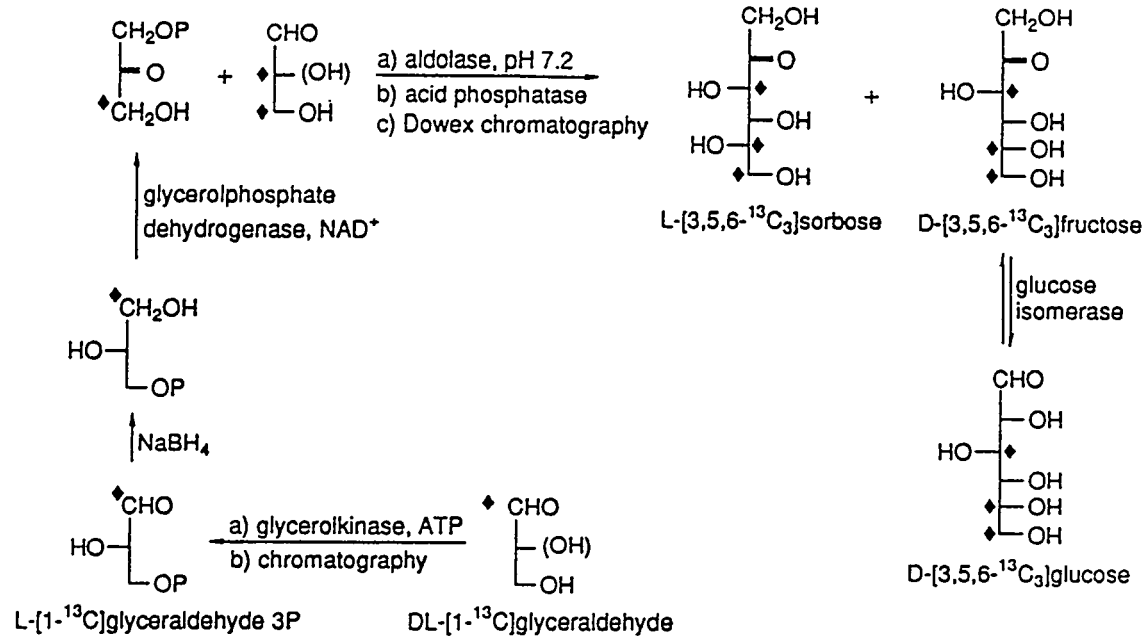
prepared with galactoside 2-α-L-fucosyltransferase (E.C. 2.4.1.69) and isotopically labeled GDP-L-fucose (Nunez *et al.*, 1981). Sucrose has been synthesized with single sites of ^{13}C-enrichment within the furanosyl ring using labeled D-fructose, UDP-D-glucose and sucrose synthase (E.C. 2.4.1.13) (Duker and Serianni, 1993).

Access to ^{13}C-labeled riboses (discussed above) permits the synthesis of ribo- and 2′-deoxyribonucleosides labeled within the furanose component. Adenosine, cytidine, guanosine and uridine have been prepared chemically with single sites of ^{13}C-enrichment at C1′ or C2′ (Kline and Serianni, 1990b), according to the general outline shown in Scheme 15. Ribonucleoside derivatives (erythroadenosine, erythrocytidine, erythrouridine) have been prepared with ^{13}C-labeling at C1′ (Kline and Serianni, 1992). Labeled ribonucleosides have been converted chemically to labeled 2′-deoxyribonucleosides (Scheme 15); in this fashion, 2′-deoxyadenosine, 2′-deoxycytidine and thymidine have been synthesized with ^{13}C-labeling at C1′ or C2′ (Bandyopadhyay *et al.*, 1993). Several studies have appeared in which ^{13}C-labeled 2′-deoxyribonucleosides have been converted to their corresponding phosphoramidites and assembled into specific-sequence DNA oligomers (Lancelot *et al.*, 1993; Kellenbach *et al.*, 1992; Wu and Serianni, 1994). Chemically synthesized ribonucleosides, after phosphorylation to give the corresponding 5′-monophosphates (Yoshikawa *et al.*,

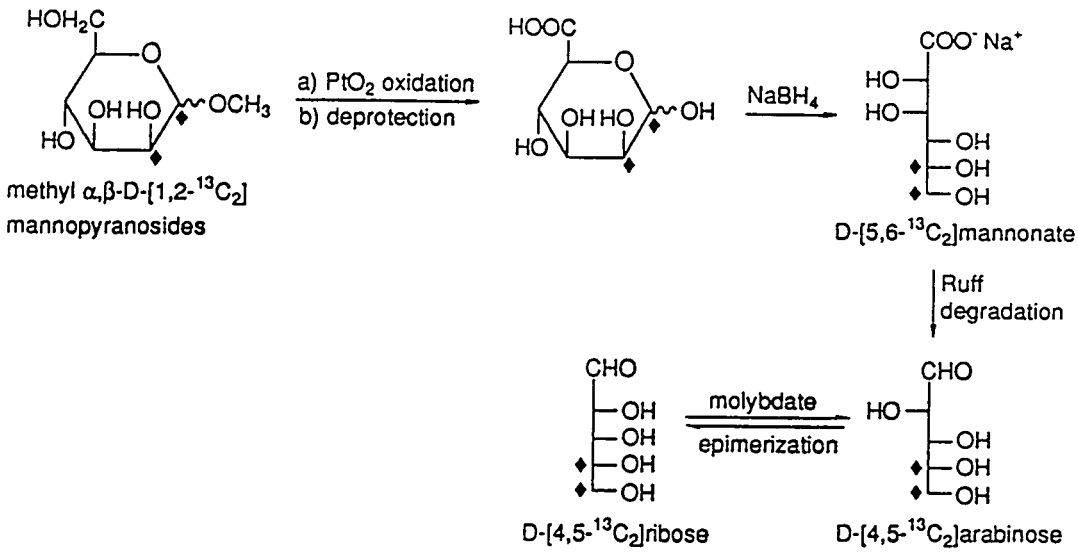
Scheme 11



Scheme 12



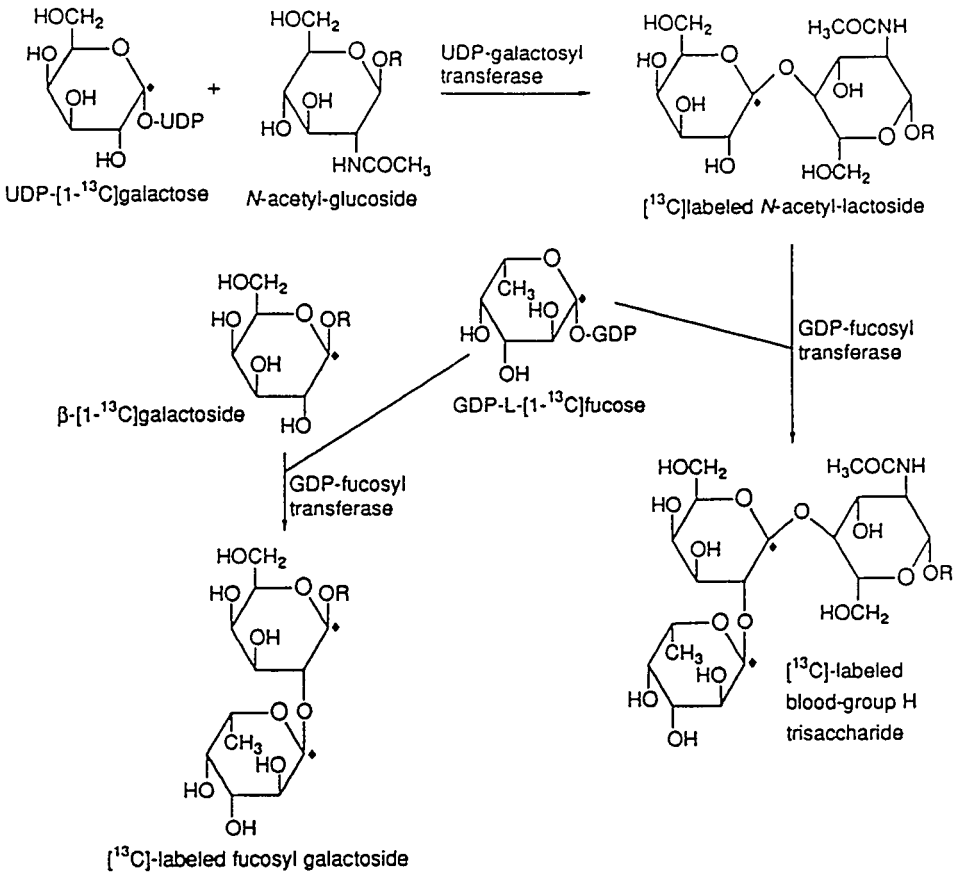
Scheme 13



1967; Sowa and Ouchi, 1975), can be assembled into RNA oligomers using RNA T7 polymerase (Nikonowicz *et al.*, 1992). Alternatively, the preparation of labeled RNA oligomers via solid-phase methods may soon be realistic (Wu and Ogilvie, 1990).

The synthesis of labeled ribonucleosides and 2'-deoxyribonucleosides can be improved by employing enzymes to catalyze sugar-base exchange or by chemically interconverting nucleosides. Thus, for example, thymidine phosphorylase (E.C. 2.4.2.4) and purine nucleoside phosphorylase (PNPase)(E.C. 2.4.2.1) can be used to convert labeled thymidine to labeled 2'-deoxyadenosine (Krenitsky *et al.*, 1981). Amination of uridine gives cytidine in good yield (Vorbrüggen *et al.*, 1975), and uridine can be converted to adenosine using uridine phosphorylase (E.C. 2.4.2.3) and PNPase (Krenitsky *et al.*, 1981).

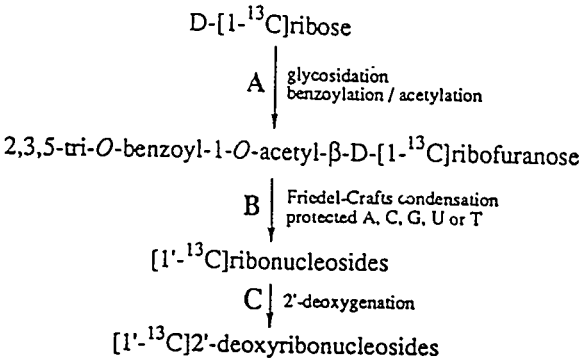
Scheme 14



Labeling with Hydrogen and Oxygen Isotopes

The cyanohydrin reduction reaction can be applied to prepare aldoses labeled with ^{2}H at H1 (Serianni and Barker, 1979d; Serianni *et al.*, 1982b) (Scheme 16). Likewise, molybdate-catalyzed epimerization of [1-^{2}H]aldoses yields corresponding C2 epimers containing ^{2}H at C2 (Hayes *et al.*, 1982a) (Scheme 5). Thus, the integrated approaches discussed above for the preparation of ^{13}C-labeled compounds can be applied with appropriate modifications to prepare ^{2}H-labeled products (for example, D-[4-^{2}H]glucose can be prepared from DL-[1-^{2}H]glyceraldehyde and DHAP with the use of aldolase as outlined in Scheme 10 (Serianni *et al.*, 1982a)).

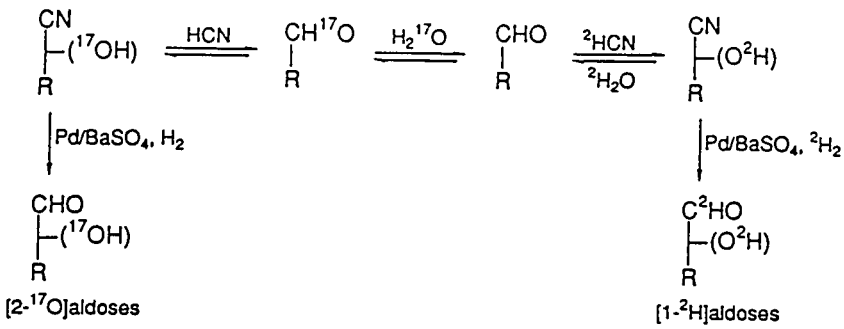
Scheme 15



Step A: Recondo and Rinderknecht, *Helv. Chim. Acta* 1959, 42, 1171.
Step B: Vorbruggen, Krolikiewicz and Bennua, *Chem. Ber.* 1981, 114, 1234.
Step C: Robins, Wilson and Hansske, *J. Am. Chem. Soc.* 1983, 105, 4059.

The cyanohydrin reduction reaction may be applied to introduce oxygen isotopes ($^{17,18}O$) at C2 of aldoses (Serianni *et al.*, 1982b; Clark and Barker, 1986) (Scheme 16). Isotope incorporation is achieved through pre-equilibration of the parent aldose in oxygen-labeled water to facilitate exchange at O1, followed by trapping of the label with cyanide. Alternatively, molybdate-catalyzed epimerization of an aldose that has been pre-equilibrated with labeled water gives its C2 epimer containing the oxygen isotope at C2 (Hayes *et al.*, 1982a) (Scheme 5).

Scheme 16



In some instances, multiple incorporation of deuterium into carbohydrates is desirable. For example, multiple deuterium labeling can simplify NMR spectral data (Snyder et al., 1989), eliminate potential relaxation pathways in order to facilitate the observation and/or measurement of NOEs (Cumming *et al.*, 1986), and/or permit ^{1}H-^{1}H internuclear distance measurements in solution via DESERT methods (Kline *et al.*, 1990). A convenient route to multiple deuterium labeling involves the use of Raney nickel as a catalyst, as first described by Koch and Stuart (1977, 1978). The reaction must be applied to nonreducing compounds, otherwise substantial degradation of the molecule occurs. The rate and extent of deuterium exchange of the carbon-bound protons depends on chemical structure and can lead to useful products, as demonstrated in a comprehensive study of this reaction applied to tetrofuranosides and pentofuranosides (Wu *et al.*, 1983). Treatment of methyl β-D-erythrofuranoside with Raney nickel and $^{2}H_2O$ yields methyl β-D-[3,4,5*S*-$^{2}H_3$]-erythrofuranoside as the major product. Interestingly, deuterium incorporation at the prochiral C4 position is stereospecific (that is, at the pro-*S* site), leading to a protected chiral CH_2OH group. Methyl β-D-[3,4,5*S*-$^{2}H_3$]erythrofuranoside was subsequently used as a precursor to ribonucleosides and 2′-deoxyribonucleosides stereospecifically deuterated at H5′*S* (Scheme 17), thereby allowing stereochemical assignments of the hydroxymethyl proton signals in ^{1}H NMR spectra and an analysis of C4′—C5′ bond rotamer populations in solution (Fig. 1) (Kline and Serianni, 1988, 1990a). The Raney nickel hydrogen-deuterium exchange reaction has been applied to oligosaccharides to assist in structure/dynamics studies by NMR (Cumming *et al.*, 1986).

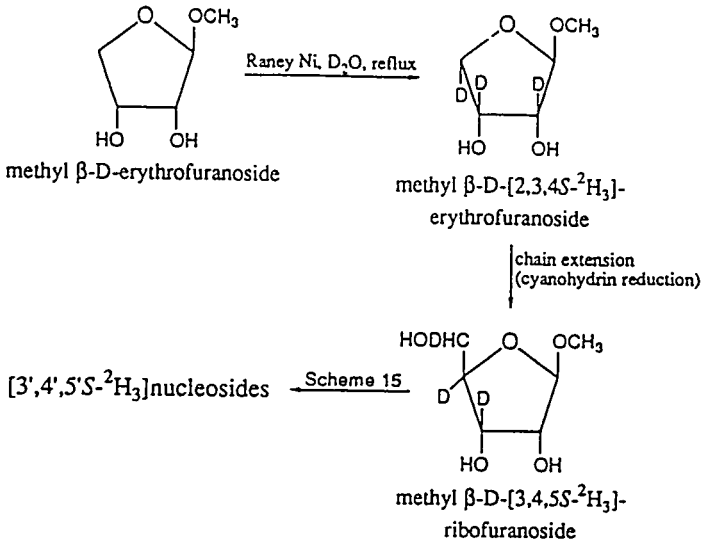
Some Applications of Labeled Carbohydrates

In this section, several current problems will be considered which have been addressed with the use of isotopically labeled carbohydrates and their derivatives.

Solution Composition of Simple Reducing Sugars and Rates of Anomerization

Simple reducing sugars, when dissolved in aqueous solution, undergo spontaneous ring-opening and -closing reactions to yield equilibrium mixtures of tautomers (Scheme 18). These mixtures typically contain α- and β-furanoses, α- and β-pyranoses, and acyclic hydrate and carbonyl forms (Angyal, 1969, 1991). NMR spectroscopy has played a central role in detecting and quantifying these forms in solution (Angyal, 1991), although detection of those present in minor amounts is not straightforward (Maple and Allerhand, 1987). ^{13}C-Enrichment at the anomeric carbon of reducing sugars enhances the detection and quantification of these minor forms; for example, the acyclic carbonyl and/or hydrate forms of D-ribose (King and Morris, 1987), D-idose

Scheme 17



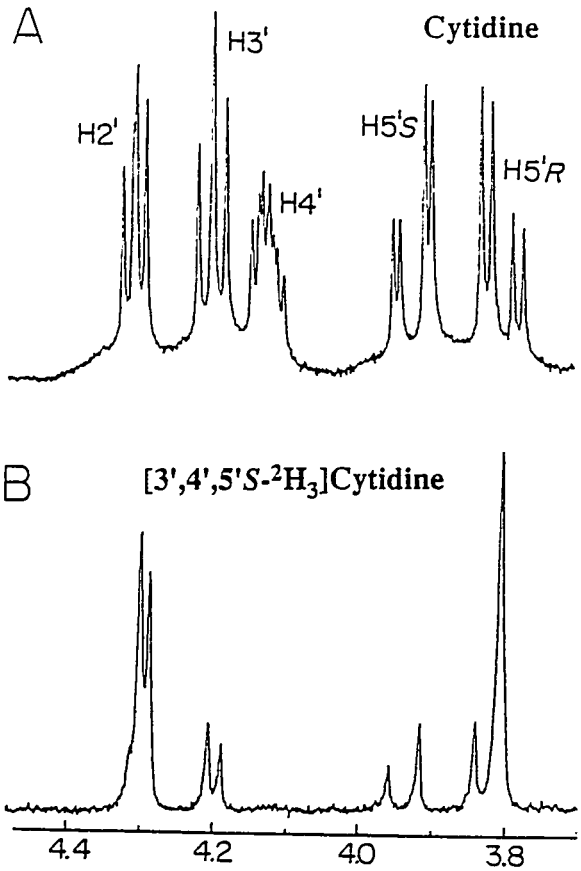
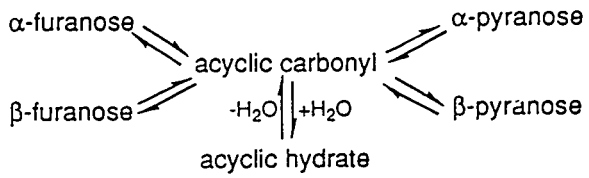


Fig. 1. (A) 300 MHz ^1H NMR spectrum of cytidine in ^2H_2O (H2'-H5'R region), showing signal assignments. The stereochemical assignments of the C5' protons were made by inspection of a similar spectrum obtained on [3',4',5'S-$^2H_3$]-cytidine (B). The spectrum of the trideuterated compound shows that H3' and H5'S are not completely exchanged with deuterium, and reveals a significant upfield shift of the H5'R signal (~4.5 Hz) in molecules containing deuterium at H5'S . (Figure taken from Kline and Seranni, 1988)

(Snyder and Serianni, 1986), D-talose (Snyder *et al.*, 1989), D-erythrose (Serianni *et al.*, 1982c), D-threose (Serianni *et al.*, 1982c), 5-*O*-methylpentoses (Snyder and Serianni, 1988), 5-deoxypentoses (Snyder and Serianni, 1988), ribulose (Wu *et al.*, 1990), xylulose (Wu *et al.*, 1990), ribulose 1,5-bisphosphate (Serianni *et al.*, 1979c), pentose 5-phosphates (Pierce *et al.*, 1985) and apiose (Snyder and Serianni, 1987b) have been detected and quantified in this fashion, as summarized in a recent review (Angyal, 1991). More importantly, the detection of the acyclic carbonyl forms of reducing sugars permits the measurement of unidirectional rate constants for the ring-opening and -closing reactions of anomerization by saturation-transfer NMR methods (Snyder *et al.*, 1989; Serianni *et al.*, 1982c). This approach has been applied to a wide range of furanoses, and some pyranoses, in an attempt to better understand the effects of molecular structure and configuration on anomerization kinetics. In aldofuranoses, ring-opening rate constants depend on the relative orientation of hydroxyl groups at C1 and C2, with *cis* orientations facilitating the reaction (Serianni *et al.*, 1982c; Snyder and Serianni, 1988). This effect may be caused by the anchimeric assistance of O2 in abstracting the hydroxyl proton at O1, either directly or through the participation of solvent water (Snyder and Serianni, 1988) (Scheme 19). Substitution of the anomeric proton of aldofuranoses by a CH_2OH group (that is, conversion to the corresponding 2-ketose) results in a substantial decrease in ring-opening rate

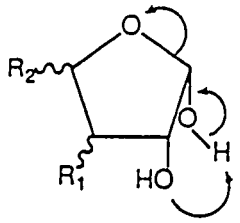
Scheme 18



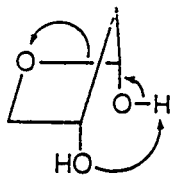
constants (Wu *et al.*, 1990), whereas alkylation at sites other than the anomeric site has little effect on ring-opening rate constants but a measurable effect on ring-closing (Snyder and Serianni, 1991). The latter result presumably represents a manifestation of the Thorpe-Ingold effect on anomerization kinetics (Snyder and Serianni, 1991). Some evidence has been obtained to suggest that anchimeric assistance other than that involving O2 may occur during aldofuranose ring-opening. This conclusion was based on the observed effects of anomeric configuration in 2-deoxy-aldofuranoses that could not be explained by the above-noted *cis*-1,2 effect. An analysis of the preferred conformations of these deoxyfuranose rings in solution suggested that O3 could be involved as a participant in the abstraction of the hydroxyl at C1 (Scheme 20).

The effect of phosphate and carboxylate substituents on furanose ring anomerization has also been inspected. The *cis*-1,2 effect observed in simple (neutral) aldofuranoses is not maintained in aldose phosphates at pH > 6.0 (Pierce *et al.*, 1985) and penturonic acids at pH > 3.0 (Wu and Serianni, 1991b), whereas the effect is

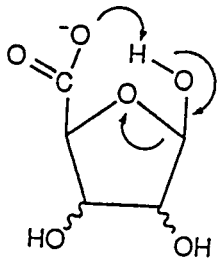
Scheme 19



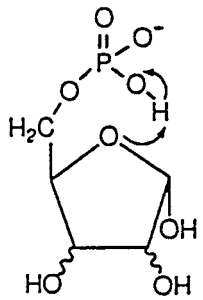
Scheme 20



Scheme 21



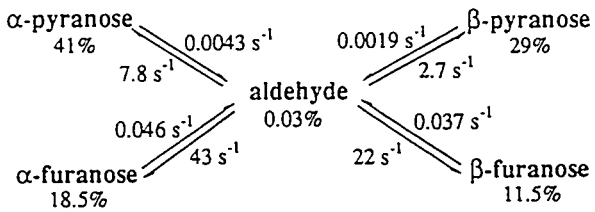
Scheme 22



observed in the protonated penturonic acids (pH < 1.5) (Wu and Serianni, 1991b). These data suggest different mechanisms of ring-opening of the ionized species, as illustrated in Schemes 21 and 22 (Wu and Serianni, 1991b).

A few studies of aldohexoses have been conducted using saturation-transfer and/or 2D ^{13}C-exchange spectroscopy (Snyder *et al.*, 1989; Snyder and Serianni, 1986). Some effects of ring structure on pyranose ring-opening and -closing rate constants have been observed, but more data are needed to draw firm structure/reactivity correlations. The available data show, however, that the anomerization of furanoses is more kinetically favored than that of pyranoses, as manifested by larger ring-opening and -closing rate constants for the furanose tautomers of a given aldose compared to those for pyranose ring forms (Scheme 23) (Snyder *et al.*, 1989; Snyder and Serianni, 1986). Thus, not unexpectedly, the thermodynamically and kinetically favored ring forms are different. In addition, although ring-opening is always considered to be rate-determining in the interconversion of aldose ring forms, this assumption may not be valid for 2-ketofuranoses. For example, the ring-opening rate constant

Scheme 23



[50 mM Na acetate, pH 4.0, 15% v/v ^2H_2O, 28°C]

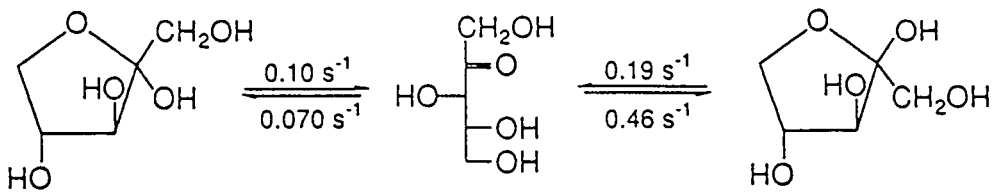
for β-D-*threo*-2-pentulofuranose is larger than that for the closure of the acyclic carbonyl form to α-D-*threo*-2-pentulofuranose at pH 4.0 (acetate buffer, 15%v/v ^2H_2O) and 50°C (Wu *et al.*, 1990) (Scheme 24).

These studies not only contribute to an improved understanding of anomerization kinetics of simple reducing sugars, but also have potential value in assessing the nature of more complex biologically important processes involving reducing centers. For example, DNA repair frequently involves the generation of reducing centers within the DNA molecule whose characteristics may affect DNA structure and thus recognition by other biological factors involved in the repair process (Sancar and Sancar, 1988). The above-noted experimental approaches can also be applied to these complex cases to examine the thermodynamics and kinetics of their constituent reducing centers. Such work would require the incorporation of ^{13}C at the appropriate anomeric centers to permit the selective detection and analysis of the solution chemistry at these sites (Wilde *et al.*, 1989).

NMR Studies of Metabolic Processes

^{13}C-labeled molecules are excellent non-invasive probes of metabolic processes, as amply demonstrated over the past 15 years (London, 1988; Bottomley, 1989). Our work has focused on studies of polyol metabolism in insects capable of adapting and surviving low temperatures.

Scheme 24



α-D-*threo*-2-pentulofuranose
(α-D-xylulose)

β-D-*threo*-2-pentulofuranose
(β-D-xylulose)

[0.3 M pentulose, 15% v/v ^2H_2O, 50 mM acetate buffer, pH 4.0, 50 °C]

The organisms under investigation, *Gynaephora groenlandica* and *Hyalophora cecropia*, are freeze tolerant because mechanisms are present in these organisms that allow nonlethal ice-formation within the extracellular spaces of their tissues. It is believed that studies of freeze-adapted organisms will lead to an improved understanding of the effects of cold temperature on cell structure and function, and to improved methods of tissue preservation.

The injection of ^{13}C- and/or ^{19}F-labeled carbohydrates into live *Gynaephora* larvae or *Hyalophora* pupae permits the *in vivo* monitoring of polyol (for example, glycerol, sorbitol) metabolism (Podlasek and Serianni, 1994; Kukal *et al.*, 1988, 1989) (Fig. 2). Although a detailed summary of the results of these studies is not possible here, a few findings are worthy of mention. In *Gynaephora* larvae, the turnover of the glycerol pool was found to depend on oxygen availability at normal temperatures (Kukal *et al.*, 1989), with lower oxygen availability resulting in a slower turnover of this pool. Ice formation within the insect will induce hypoxic or anoxic conditions and may be a cause—other than or in addition to metabolic controls—of glycerol accumulation under these conditions. Mitochondrial density within the tissue was also observed to decrease at lower body temperatures (Kukal *et al.*, 1989), and this decrease may represent an anticipatory physiological response to freezing conditions and the subsequent hypoxic/anoxic metabolism.

3-Deoxy-3-fluoro-D-glucose, and D-glucose singly labeled with ^{13}C at C1, C2, or C3, have been used as metabolic probes in *cecropia* pupae to detect the presence of enzymes required for the production of sorbitol, a common cryoprotectant, and

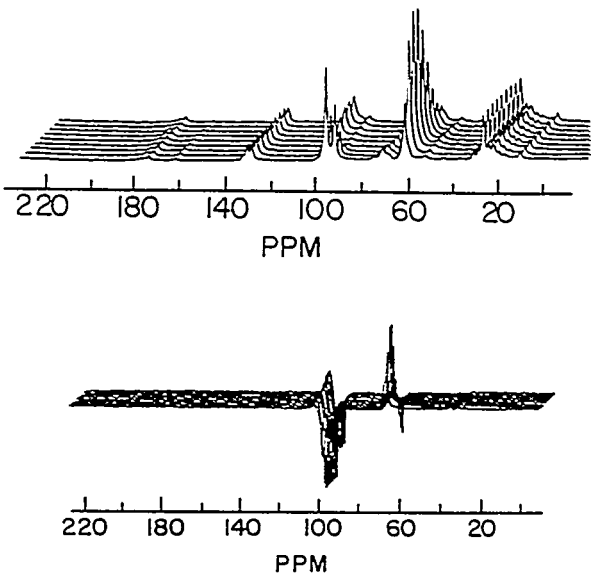


Fig. 2. **In vivo** ***^{13}C NMR spectra of a representative warm-acclimated (15°C)*** **G. groenlandica** ***larva after injection with 15 mg of D-[1-^{13}C]-glucose. Time-lapse spectra collected approximately hourly for 10 hr (upper stacked plots) at a body temperature of 25°C, showing the degradation of labeled glucose (97.3 and 93.6 ppm), the formation of intermediate labeled glycerol (64.2 ppm), and eventual labeling of glycogen (101.3 ppm) and trehalose (94.8 ppm). Difference spectra (lower stacked plots) obtained by subtracting the first spectrum from each subsequent spectrum show that no detectable ^{13}C label is incorporated into lipids over this 10-hr time period. (Figure taken from Kukal*** **et al.,** ***1988).***

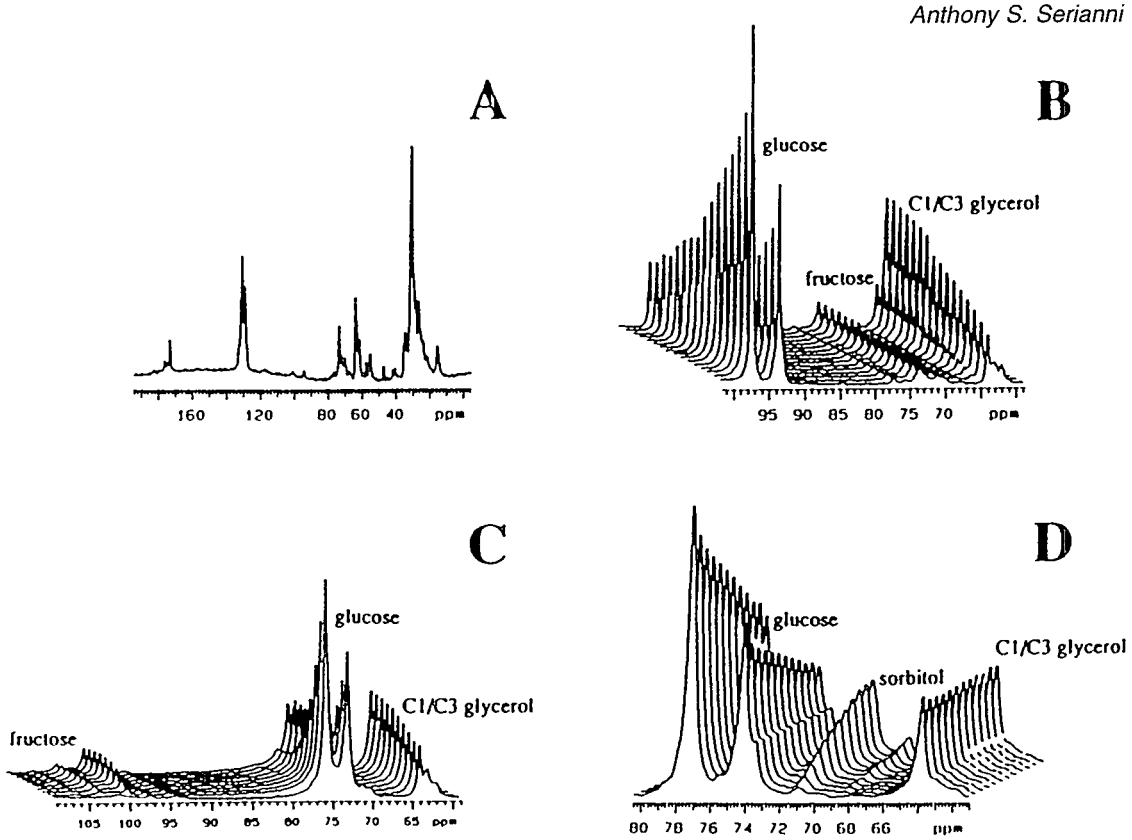


Fig. 3. (A) Natural abundance ^{13}C NMR spectrum of an intact H. cecropia *pupa. Signals observed at 30, 130, and 180 ppm were attributed to the saturated, unsaturated, and carbonyl carbons of the fatty acids of triglycerides. Two additional resonances at 64.0 and 73.5 ppm arise from the C1/C3 and C2 carbons of glycerol. (B) Time-lapse ^{13}C NMR spectra of an intact pupa injected with D-[1-^{13}C]glucose, showing metabolism to D-[1-^{13}C]fructose (65.4 ppm). (C) Time-lapse ^{13}C NMR spectra of a pupa injected with D-[2-^{13}C]glucose. Fructose synthesis is apparent from the signals at 99.5, 103.0, and 106.0 ppm, which are attributed to the ß-pyranose, ß-furanose, and α-furanose forms, respectively. (D) Time-lapse ^{13}C NMR spectra of an intact pupa injected with D-[3-^{13}C]glucose. Time-lapse ^{13}C label incorporation into C3 of sorbitol (71.3 ppm), fructose (69.1 ppm), and glycerol (64.0 ppm) was observed. (Figure taken from Podlasek and Serianni, 1994.)*

D-fructose *in vivo* (Fig. 3) (Podlasek and Serianni, 1994). Sorbitol production was observed at a wide range of body temperatures, despite the fact that this organism fails to accumulate this polyol at lower temperatures, unlike other insects (Storey *et al.*, 1981). This result suggests that mechanisms regulating sorbitol and glycerol biosynthesis in freeze-tolerant insects are species-dependent.

Oligosaccharide Conformation

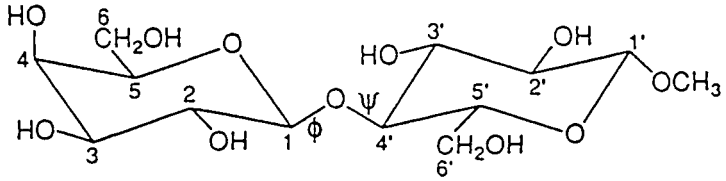
Using chemi-enzymic approaches, a number of biologically important oligosaccharides have been prepared that

contain one or more sites of isotopic enrichment in order to study their solution conformations (Nunez and Barker, 1980; Rosevear *et al.*, 1982; Hayes *et al.*, 1982b; Kline *et al.*, 1990; Duker and Serianni, 1993). The main justification for ^{13}C and/or ^{2}H-labeling within this class of biomolecules lies in the additional information provided with respect to *O*-glycoside linkage geometry, although, as found in the labeling of other classes of biomolecules, such labeling can greatly facilitate signal assignments through the use of multidimensional methods (for example, HMQC-TOCSY (Wijmenga *et al.*, 1989).

In general, the least well understood structural component of oligosaccharides which exerts a major effect on the conformations of these molecules is the *O*-glycoside linkage. Although the forces that govern geometry about these linkages have been identified (for example, stereoelectronic forces such as the anomeric (Lemieux, 1963; Tvaroska and Bleha, 1989) and exoanomeric (Tvaroska and Bleha, 1989; Limieux, 1971; Lemieux *et al.*, 1979) effects), the experimental assessment of conformation and dynamics about these linkages is not straightforward. The origin of the difficulty lies, in part, in the absence of three-bond ^{1}H-^{1}H spin-coupling pathways across these linkages, whose corresponding couplings can be correlated with molecular dihedral angles through appropriate Karplus relationships. In addition, for many oligosaccharides, NOEs between carbon-bound protons are small or absent (Serianni,1992), thereby preventing 3D structure determination through the use of ^{1}H-^{1}H internuclear distance constraints. Some relief might be provided by the exchangeable hydroxyl protons which lie on the periphery of these molecules and should thus experience NOEs that are sensitive to molecular geometry. With the development of pulsed field gradients (Hurd, 1990) and shaped pulses (Smallcombe, 1993; Kessler *et al.*, 1991), oligosaccharide analyses in H_2O solvent should become easier to perform, thereby making hydroxyl proton NOEs more easily measurable. However, the effects of conformational averaging on NOEs are complex, and thus they may be difficult to interpret structurally.

The presence of ^{13}C-labeling in the vicinity of the *O*-glycoside linkages of oligosaccharides permits the measurement of trans-glycoside ^{13}C-^{1}H and ^{13}C-^{13}C spin-coupling constants that can be used to assess linkage geometry. These couplings have been measured in methyl β-lactoside (Hayes *et al.*, 1982b) (Scheme 25). Using Karplus-type relationships derived for $^{3}J_{COCH}$ and $^{3}J_{CCCH}$ (Schwarcz *et al.*, 1972; Tvaroska *et al.*, 1989; Mulloy *et al.*, 1988), and more limited data for $^{3}J_{COCC}$ (King-Morris and Serianni, 1987), it is possible to limit the linkage geometry to a defined region of conformational space (Hayes *et al.*, 1982b) (Scheme 26), leading to the conclusion that linkage geometry, at least in this disaccharide, is comparatively rigid. In contrast, recent studies of sucrose singly labeled with ^{13}C at C1, C2, C3 and C6 of the fructofuranosyl ring have led to the conclusion

Scheme 25



methyl β-lactoside

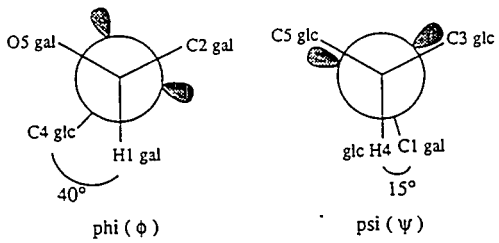
ϕ = -40°: $^{3}J_{C4',H1}$ = 3.8 Hz; $^{3}J_{C4',C2}$ = 3.1 Hz
ψ = -15°: $^{3}J_{C1,H4'}$ = 4.9 Hz; $^{3}J_{C1,C3'}$ = 0 Hz; $^{3}J_{C1,C5'}$ = 1.6 Hz

that the phi (ϕ) angle is relatively rigid, whereas the psi (ψ) angle is flexible (Duker and Serianni, 1993), a conclusion that appears consistent with prior computational predictions (Tran and Brady, 1990).

Other approaches involving use of isotopic enrichment to assess oligosaccharide conformation include selective ^{2}H-substitution near the *O*-glycoside linkage. Thus, for example, methyl β-lactoside was synthesized with a single site of ^{2}H-enrichment at H1 of the Gal residue, and the effect of this substitution on ^{1}H spin-lattice relaxation times was assessed (Kline *et al.*, 1990). These data were used to estimate the ^{1}H-^{1}H internuclear distance between H1 Gal and H4′ Glc, which was used to assess linkage geometry. This distance was found to be in good agreement with ^{13}C-^{1}H and ^{13}C-^{13}C spin-coupling data (Hayes *et al.*, 1982b) (see above). Similar deuteration experiments have been conducted on a Gal-Fuc disaccharide (Kline *et al.*, 1990).

Scheme 26

Glycoside Linkage Conformation: Methyl β-lactoside



Oligonucleotide Conformation

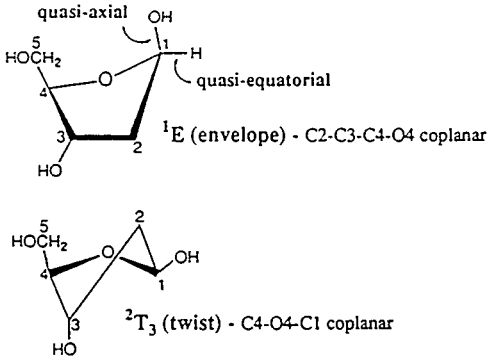
^{1}H NMR studies of the solution conformations and dynamics of DNA and RNA have been hampered by the complex spectral characteristics exhibited by these molecules. This complexity is caused by the relatively small number of highly similar monomer units found in these structures compared to proteins, which gives rise to considerable signal overlap, especially in spectral regions where the furanose ring protons resonate. RNA represents a greater problem than DNA

in that the C2′—C5′ protons have similar chemical shifts; in DNA, the C2′ proton signals lie far upfield of those of the C3′—C5′ protons, thereby reducing the spectral crowding in the latter region. The anomeric proton region is distinct in ^{1}H NMR spectra of DNA and RNA, but this region can also be difficult to decipher as oligomer length increases. Fortunately, the spectral dispersion of the proton signals for any specific residue is usually sufficient to produce first-order (or essentially first-order) spectral behavior, which permits the use of multidimensional NMR methods to evaluate chemical shifts and coupling constants with reasonable accuracy without the need for spectral simulation (in comparison, non-first-order behavior is frequently encountered in the ^{1}H NMR spectra of oligosaccharides). Nevertheless, serious limitations are encountered in DNA and RNA structure determination at molecular weights significantly lower than observed for proteins. For this reason, the labeling strategies applied successfully to proteins to decipher their spectra are of even greater importance in interpreting the spectra of oligonucleotides.

Recent progress in introducing ^{13}C isotopes into oligonucleotides has led to an interest in exploiting new spin-coupling constants made available by this labeling, such as ^{13}C-^{1}H and ^{13}C-^{13}C spin-coupling constants. These spin-couplings are especially useful in studying furanose ring conformation, which is particularly challenging to assess, although they may be exploited to inspect other conformational features (for example, *N*-glycoside and hydroxymethyl conformation).

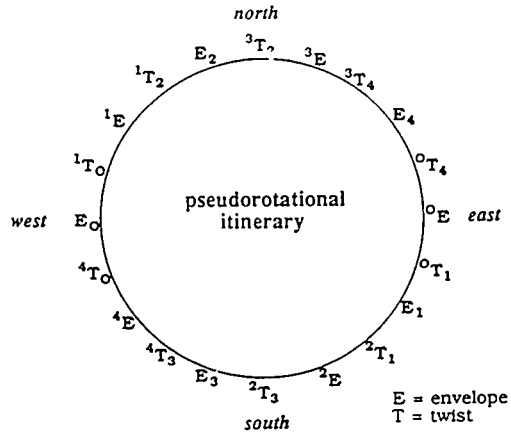
Furanose rings are conformationally flexible. They do not prefer planar geometries, but instead adopt two idealized nonplanar forms characterized by having one (envelope, E) or two (twist, T) ring atoms out of the plane defined by the remaining ring atoms (Scheme 27). Ten idealized envelope and ten twist forms can be depicted for a given furanose ring, each having energies within about 4 kcal/mol (Westhof and Sundaralingam, 1983). This small energy difference allows considerable conformational averaging in solution, with interconversion between nonplanar forms occurring via an inversion mechanism (via the planar form) or pseudorotation (Altona and Sundaralingam, 1972) (Scheme 28). The relative importance of the inversion and pseudorotation mechanisms has been studied (Westhof and Sundaralingam, 1983), and although it is commonly believed that pseudorotation is more preferred, there may be some circumstances in which interconversions occur via inversion. It should also be appreciated that the motional properties of the furanose rings of DNA and RNA contribute to the conformational entropy of these molecules (along with other factors), and that any change in this motion that may occur upon binding by appropriate receptors may potentially affect the binding affinity (that is, the free energy of binding).

Scheme 27



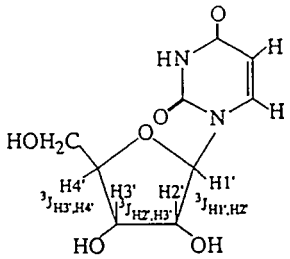
Non-planar Forms of 2-Deoxy-β-D-*erythro*-pentofuranose

Scheme 28



Conventional conformational analysis of furanose rings in oligonucleotides is based on an interpretation of $^3J_{HH}$ values, whose magnitudes are affected by molecular dihedral angles as first enunciated by Karplus (Karplus, 1959) and subsequently refined by Altona and coworkers (Haasnoot *et al.*, 1980, 1981). In RNA, the analysis is based on the magnitudes of $^3J_{H1',H2'}$ and $^3J_{H3',H4'}$ (Scheme 29) and assumes a two-state model (de Leeuw and Altona, 1983) involving exchange between north (N, ^{3}E) and south (S, ^{2}E) conformers of the pseudorotational itinerary (Scheme 28) ($^3J_{H2',H3'}$ is relatively insensitive to N/S exchange). A similar two-state model is invoked to interpret furanose $^3J_{HH}$ values in DNA (de Leeuw and Altona, 1983), although two additional $^3J_{HH}$ ($^3J_{H1',H2''}$, and $^3J_{H2'',H3'}$) are available to support the analysis. The two-state N/S model has been proposed based on the observed conformational behaviors of

Scheme 29



Uridine

simple nucleosides and nucleotides in the crystalline state (de Leeuw *et al.*, 1980), and has been applied to interpret $^3J_{HH}$ in a wide range of nucleosides, nucleotides and oligonucleotides. However, considering the flexibility of furanose rings and the numerous conformational models that can exist, it can be argued that additional parameters to assess ring conformation would be desirable. Furthermore, some evidence has been reported to suggest the absence of N/S averaging in some structures

(for example, Reid and coworkers conclude that the deoxyribose rings in a DNA-RNA hybrid duplex assume an east (°E) conformation) (Salazar *et al.*, 1993). Likewise, in furanose-containing biomolecules having configurations other than those found in DNA and RNA, it is possible, if not likely, that conformational averaging does not conform to a simple N/S two-state model.

We focus here on ^{13}C-^{1}H spin-couplings within the furanose rings of DNA and RNA as potential conformational probes, although $^{1}J_{CC}$ and longer-range ^{13}C-^{13}C spin-couplings in carbohydrates have been studied (King-Morris and Serianni, 1987; Wu *et al.*, 1992b; Carmichael *et al.*, 1993) and may also prove useful in this regard.

^{13}C-^{1}H Spin-coupling constants (J_{CH}) have been used qualitatively to examine the solution conformations of simple furanoid rings (Snyder and Serianni, 1987a;Wu and Serianni, 1991a; Vuorinen and Serianni, 1990b; Snyder and Serianni, 1991; Serianni and Barker, 1984; Angelotti *et al.*, 1987). A more quantitative treatment, however, is required in order to apply these couplings to test conformational models and dynamical properties. Inspection of the β-D-ribofuranose ring of RNA reveals the presence of 18 two- ($^{2}J_{CH}$) and three- ($^{3}J_{CH}$) bond ^{13}C-^{1}H couplings, 13 being useful for ring analysis and 5 for sidechain (hydroxymethyl) conformation (Table 1). In addition, six one-bond ($^{1}J_{CH}$) couplings exist, four of which may be sensitive to ring conformation. Thus, a total of 17 J_{CH} values report on β-D-ribofuranose ring shape. In principle, these latter couplings might be applied in an integrated quantitative approach to structure determination provided that each of their dependencies on structure can be established. These dependencies can be probed using computational and experimental (empirical) approaches, as discussed below.

Table 1. Two- and Three-Bond ^{13}C-^{1}H Coupling Pathways in β-D-Ribofuranosyl Rings[a]

Intraring (13)	Hydroxymethyl (5)
C1 - H2	C3 - H5*R*[b]
C1 - H3[b]	C3 - H5*S*[b]
C1 - H4[b]	C4 - H5*R*
C2 - H1	C4 - H5*S*
C2 - H3	C5 - H4
C2 - H4[b]	
C3 - H1[b]	
C3 - H2	
C3 - H4	
C4 - H1[b]	
C4 - H2[b]	
C4 - H3	
C5 - H3[b]	

[a]Three additional couplings are available in 2-deoxy-β-D-*erythro*-pentofuranosyl rings.
[b]Vicinal pathways.

Accurate measurements of J_{CH} values can be made from an analysis of 1D ^{1}H NMR spectra of ^{13}C-labeled compounds (for example, Serianni and Barker, 1984). However, this approach is impractical for studies of complex molecules due

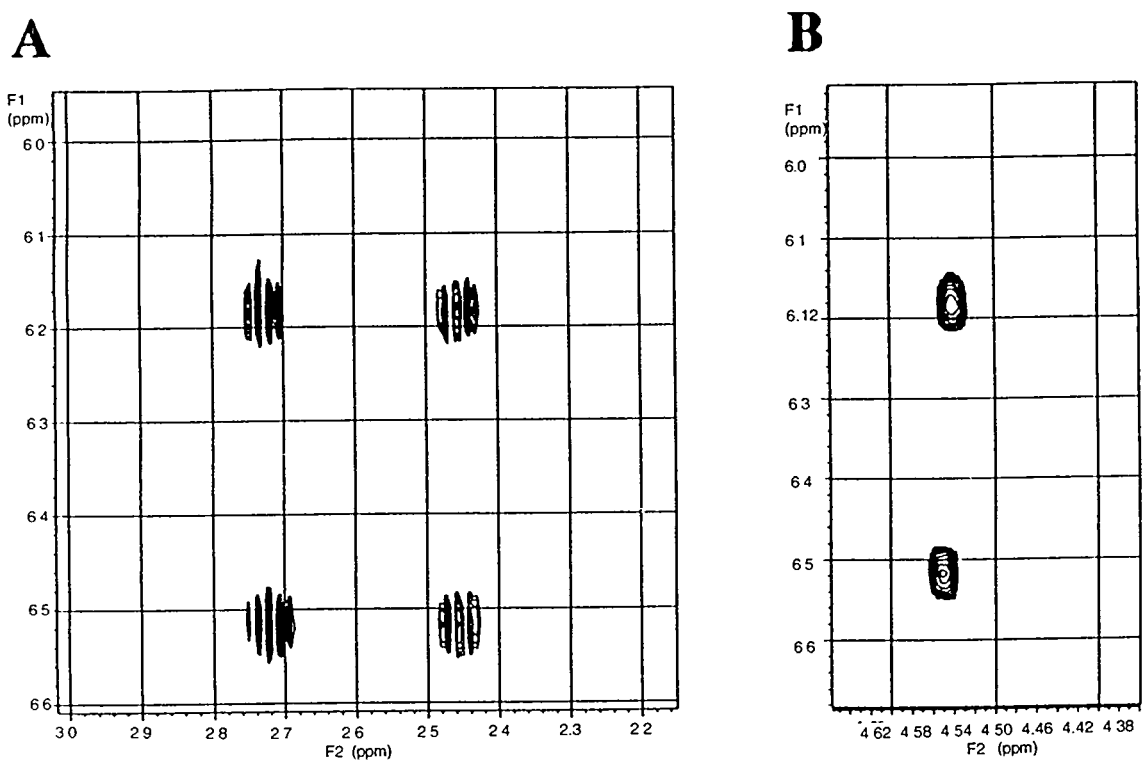


Fig. 4. (A) The partial 500-MHz TOCSY spectrum of [1'-^{13}C]2'-deoxyadenosine showing only the H1'-H2'R (F_2 = 2.45 ppm) and H1'-H2'S (F_2 = 2.72 ppm) paired crosspeaks. Only the latter are displaced (5.7 Hz). (B) The H1'-H3' paired TOCSY crosspeaks for [1'-^{13}C]2'-deoxyadenosine, showing a displacement (5.3 Hz) having the opposite sense of that for H1-H2'S. Because $^{3}J_{C1,H3}$ = +5.3 Hz, then $^{2}J_{C1',H2S}$ = -5.7 Hz. (Figure taken from Serianni and Podlasek, 1994).

to significant resonance overlap and does not yield coupling sign information readily, the latter being particularly important in the interpretation of $^{2}J_{CH}$. Homo-nuclear 2D NMR methods (for example, COSY, TOCSY), however, can provide accurate J_{CH} values and sign information based on the observation of the displacement of paired crosspeaks generated by the large $^{1}J_{CH}$ (Hines *et al.*, 1993; Montelione *et al.*, 1989; Serianni and Podlasek, 1994) (Fig. 4). The method can be extended to three dimensions (for example, HMQC-TOCSY) (Fig. 5) to achieve additional resolution via heteronuclear chemical shifts. These multidimensional methods have been applied to measure J_{CH} and J_{NH} values in proteins (Montelione *et al.*, 1989), and can be applied to make similar determinations in oligonucleotides (Hines *et al.*, 1993).

Theoretical insight into the behavior of J_{CH} in furanose rings can be gained

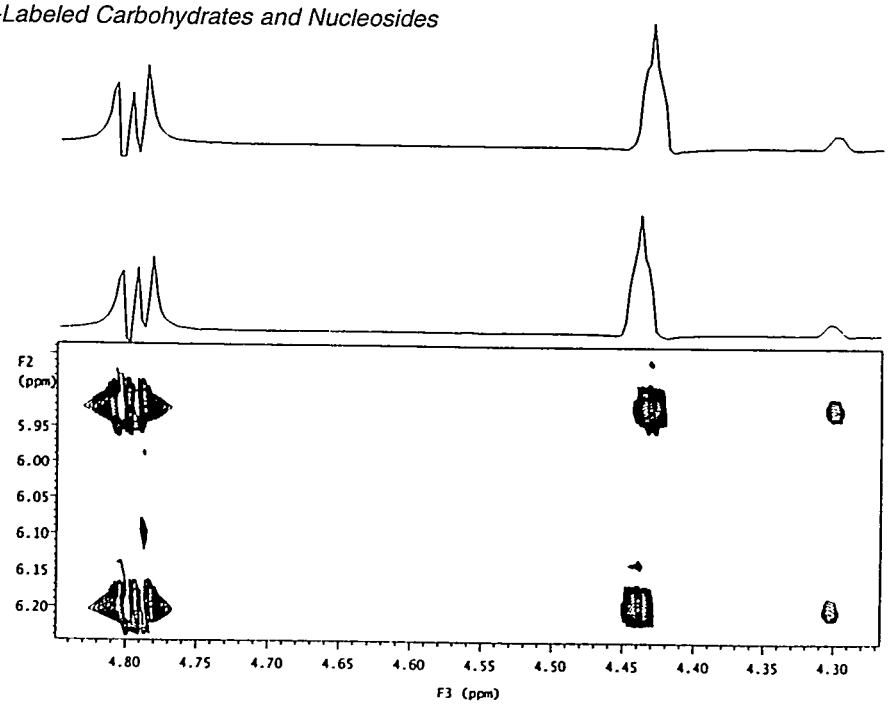
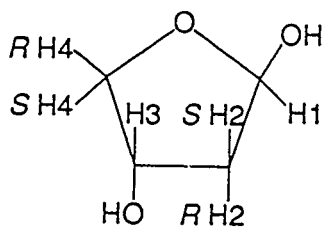


Fig. 5. A TOCSY plane of a 3D HMQC-TOCSY data set (no ^{13}C-decoupling) obtained on an equimolar mixture of [1'-^{13}C]adenosine, [1'-^{13}C]cytidine, [1'-^{13}C]ribothymidine, and [1'-^{13}C]uridine (10 mM each) in $^{2}H_2O$, edited with respect to the C1' chemical shift of [1'-^{13}C]adenosine. Thus, only the H2', H3', and H4' crosspeaks of [1'-^{13}C]adenosine are observed. As shown in Fig. 4, displacement of the paired crosspeaks (split by the large $^{1}J_{CH}$) is used to measure the magnitude and sign of $^{2}J_{C1',H2'}$ (-3.2 Hz), $^{3}J_{C1',H3'}$ (~+5.1 Hz), and $^{3}J_{C1',H4'}$ (+1.3 Hz).

through the use of *ab initio* molecular orbital calculations. The ten envelope forms and the planar form of the model furanose, 2-deoxy-β-D-*glycero*-tetrofuranose 1 (Scheme 30), have been studied using the *HF/STO-3G, *HF/3-21G, *HF/6-31G* and MP2/6-31G* basis sets (Garrett and Serianni, 1990a, 1990b). Geometric optimization was complete except for a single endocyclic torsion angle (in envelope forms) or two such angles (in the planar form), which were held fixed to confine the computation to a specific ring form.

Scheme 30



2-deoxy-β-D-*glycero*-tetrofuranose (1)

The results of this treatment indicate that the C-H bond lengths in furanose rings appear to depend on ring conformation (Figure 6A,B). In general, a given C-H bond is maximal in length when quasiaxial, and minimum in length when quasiequatorial, although the sensitivity to bond orientation is not identical for each C-H bond (for example, the C3—H3 bond in 1 appears insensitive to ring conformation in comparison to the remaining C-H bonds) (Figure 6B).† These data suggest that, in general, quasi-equatorial C-H bonds may have more s-character than quasi-axial C-H bonds, leading to the suggestion that the former should exhibit larger $^1J_{CH}$ values (Müller and Pritchard, 1959). This prediction is consistent with observations made in conformationally

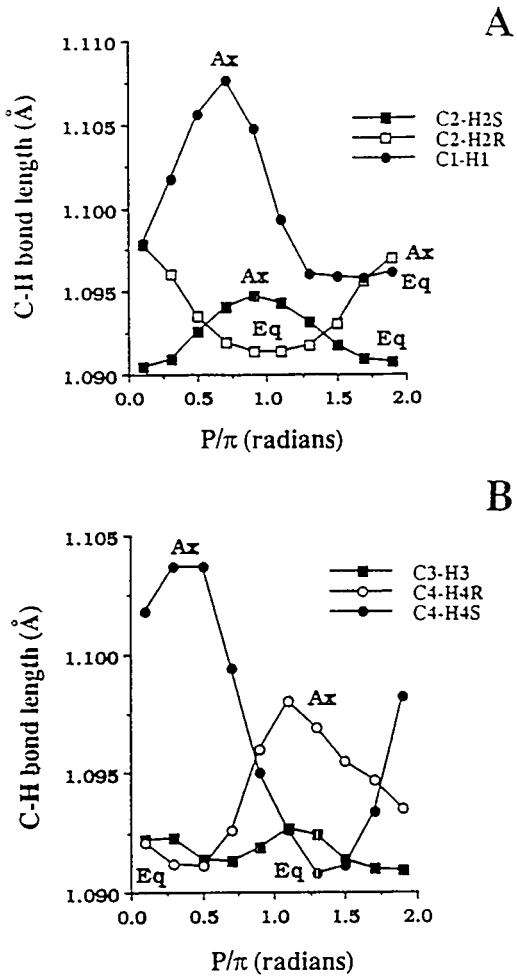


Fig. 6. The effect of ring conformation on the six C-H bond lengths of 1. Data are derived from ab initio *molecular orbital calculations (Gaussian) on the ten envelope forms of 1 using the MP2/6-31G* basis set.*

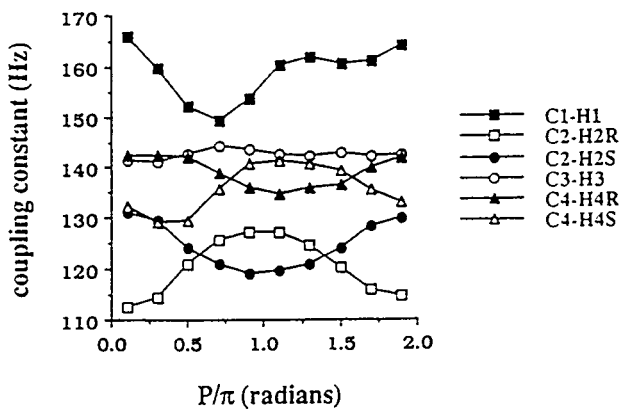


Fig. 7. Predicted change in $^1J_{CH}$ for the six C-H bonds in 1. Couplings are QCISD-scaled values (Carmichael et al., *1993). Thus, trends are expected to be correct, but absolute values will require further refinement.*

†The magnitude of C-H bond length change may be affected by exocyclic C-O rotations and intermolecular interactions (for example, solvation).

rigid aldopyranosyl rings, where $^1J_{C1,H1}$ for equatorial C1—H1 bonds is about 10 Hz larger than $^1J_{C1,H1}$ for axial C-H bonds (Bock *et al.*, 1973; Bock and Pedersen, 1974, 1975). Computed values of $^1J_{CH}$ in **1** for each of the six C-H bonds are shown in Fig. 7. These data suggest that $^1J_{CH}$ values within furanose rings may be useful conformational probes, although all $^1J_{CH}$ do not appear to be equally sensitive to changes in ring conformation. More complete calculations and additional experimental data will be required to fully assess the value of $^1J_{CH}$ to evaluate ring geometry.

Two-bond J_{CH} have been computed using the MP2/6-31G*-optimized structures of **1** and are predicted to have both negative and positive signs (Fig. 8). Several $^2J_{CH}$

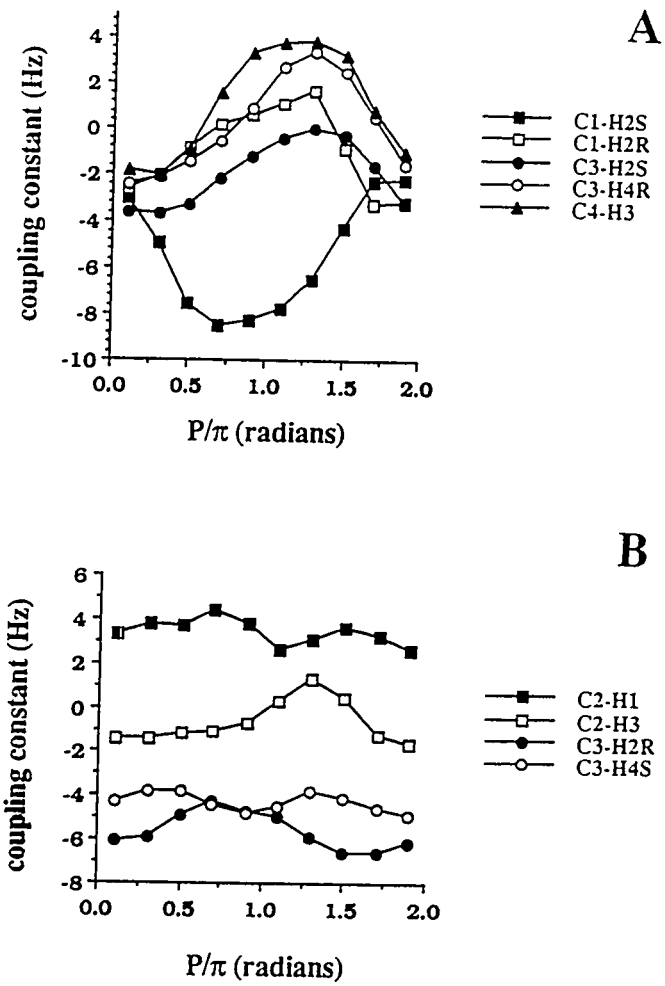


Fig. 8. (A) Predicted change in five $^2J_{CH}$ in 1 that are sensitive to ring conformation. (B) Predicted change in four $^2J_{CH}$ in 1 that are relatively insensitive to ring conformation. Couplings are QCISD-scaled values (Carmichael et al., *1993). Thus, trends are expected to be correct, but absolute values will require further refinement.*

appear sensitive to ring shape (Fig. 8A), especially $^2J_{C1,H2S}$, whereas others (for example, $^2J_{C2,H1}$) do not change significantly with conformation (Fig. 8B). $^2J_{C1,H2S}$ has been observed in 2'-deoxyribonucleosides (Bandyopadhyay *et al.*, 1993) and DNA oligomers (Wu and Serianni, 1994) to be large and negative, in agreement with these computations and with predictions based on empirical rules correlating $^2J_{CH}$ and carbohydrate structure (Schwarcz *et al.*, 1975; Bock and Pedersen, 1977).

Three-bond J_{CH} may be estimated for each envelope form of **1** using torsional information derived from MP2/6-31G*-optimized structures (Fig. 9). If the torsional behavior of **1** is assumed to be similar to that observed in the β-*ribo*

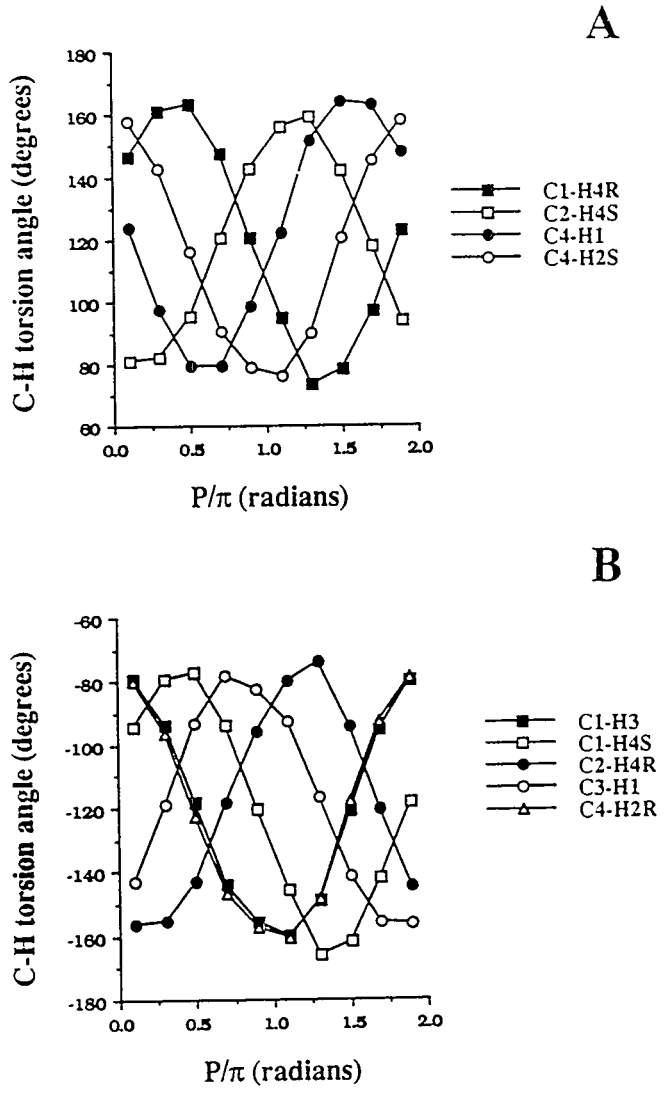


Fig. 9. The effect of ring conformation on the C-C-C-H and C-O-C-H endocyclic torsion angles of 1. Data are derived from ab initio *molecular orbital calculations (Gaussian) on the ten envelope forms of 1 using the MP2/6-31G* basis set.*

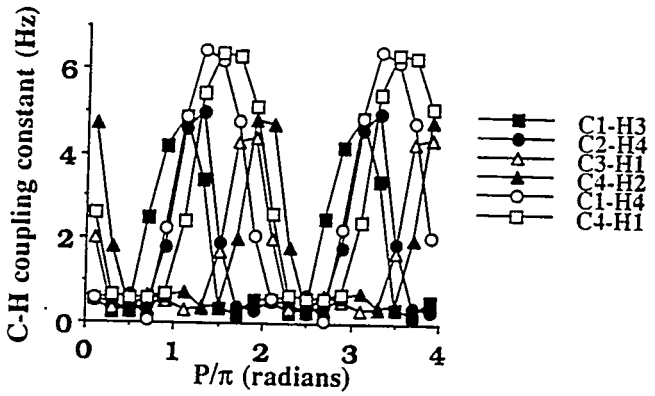
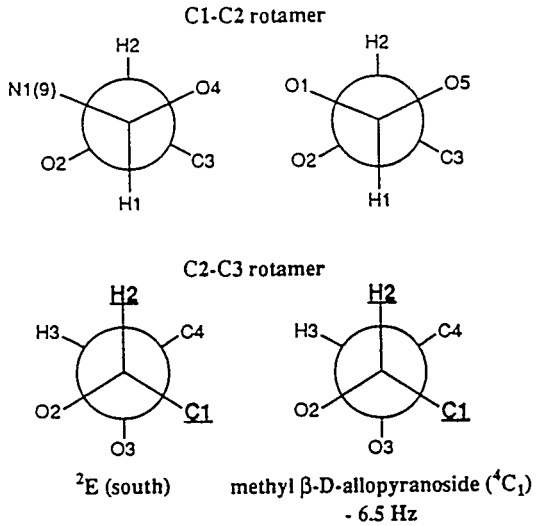


Fig. 10. Predicted effect of ring conformation on the six $^3J_{CH}$ values in a ß-D-ribofuranosyl ring. Torsions were taken from data in Fig. 9 and converted to J values using crude Karplus relationships reported previously (Schwarcz et al., 1972; Tvaroska et al., 1989a; Mulloy et al., 1988).

ring (that is, substitution at C2 and C4 does not significantly affect ring behavior), then appropriate Karplus curves for C-C-C-H and C-O-C-H coupling pathways derived using model carbohydrates (Schwarcz *et al.*, 1972; Tvaroska *et al.*, 1989; Mulloy *et al.*, 1988) may be used to convert torsion data into *J*-plots (Fig. 10). It is not possible to construct a *J*-plot for the 2-deoxy-D-*ribo* ring since Karplus curves for coupling pathways involving methylene carbons have not been reported. The data in Fig. 10 nevertheless indicate a reasonable sensitivity of $^3J_{CH}$ to ring conformation that should prove useful in furanose structure determination in RNA.

Although computational methods provide a means to evaluate the sensitivity of J_{CH} to furanose structure, some experimental verification of computed couplings is necessary. This information can be obtained through the use of conformationally rigid molecules containing C-H coupling pathways that mimic those found in discrete non-planar conformers of the β-D-*ribo* and 2-deoxy-β-D-*ribo* rings. For example, a comparison of the Newman projections about the C1—C2 and C2—C3 bonds of the ^{2}E (south) form of the β-D-ribofuranosyl ring of RNA and methyl β-D-allopyranoside shows an identical relative orientation of substituents (Scheme 31). Thus, $^2J_{C1,H2}$ observed in methyl β-D-allopyranoside (-6.5 Hz) should be similar in magnitude and sign to that in the ^{2}E conformer of the β-D-ribofuranosyl ring. Likewise, $^2J_{C1,H2}$ in methyl α-D-mannopyranoside should be similar in magnitude and sign to $^2J_{C1,H2}$ in the E_2 conformer of the β-D-ribofuranosyl ring. The α- and β-anomers of methyl 2-deoxy-D-glucopyranoside have been shown to be reasonable models for the C1-H2 coupling

Scheme 31



^{2}E (south)

methyl β-D-allopyranoside (^{4}C$_1$)
- 6.5 Hz

Scheme 32

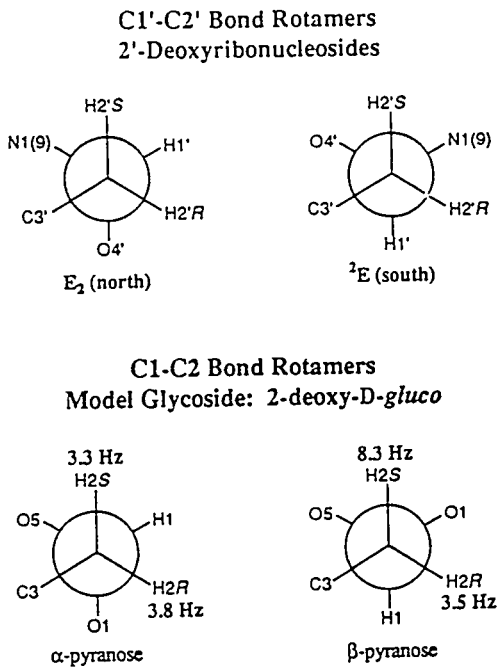


Table 2. Predicted ^{13}C-^{1}H Spin-Couplings in North (E_2) and South (^{2}E) Forms of β-D-Ribofuranosyl Rings[a]

Coupling	North (E_2) Conformer	South (^{2}E) Conformer
C1-H2	α-*manno* (-1.2 Hz)	β-*allo* (-6.5 Hz)
C1-H3	α-*manno* (0 Hz)	β-*allo* (6.0 Hz)
C1-H4 (NE,SW)[b]	β-*gluco* (2.3 Hz)	β-*arabino* (7.6 Hz)
C2-H1	α-*manno* (~ -1.8 Hz)	β-*allo* (0.3 Hz)
C2-H3	α-*manno* (1.4 Hz)	β-*allo* (-4.8 Hz)
C2-H4	β-*allo* (-2.0 Hz)	β-*arabino*
C3-H1	α-*manno* (4.6 Hz)	β-*allo* (0 Hz)
C3-H2	α-*manno* (-3.7 Hz)	β-*allo* (1.3 Hz)
C3-H4	β-*allo* (α-*gluco*)	α-*arabino*
C4-H1 (E,W)[b]	β-*gluco* (-1.0 Hz)	β-*arabino* (6.1 Hz)
C4-H2	β-*altro* (*allo*)	α-*arabino*
C4-H3	β-*allo* (*gluco*)	α-*arabino*
C5-H3 (NE,SW)[b]	β-*allo* (*gluco*)	1,6-anhydro-*altro*

[a]Estimated from model methyl D-aldohexopyranosides.
[b]Model systems are mimics of nonplanar forms in parentheses rather than E_2 and ^{2}E.

pathways in the ^2E and E_2 conformers of the 2-deoxy-β-D-ribofuranosyl ring (52) (Scheme 32).

Essentially all of the C-H coupling pathways in E_2 and ^2E conformers of the β-*ribo* ring can be modeled with appropriate ^{13}C-labeled methyl aldopyranosides (Table 2). The available data to date indicate that J_{CH} within this ring are not equally sensitive to N/S exchange; thus, for example, $^2J_{C2,H1}$ assumes similar values in these two ring forms (Table 2) and is not expected to be useful in distinguishing between these forms in solution. On the other hand, $^2J_{C1,H2}$, being significantly affected (Table 2), is expected to be a useful probe for this purpose.

The above discussion describes some experiments designed to evaluate the usefulness of ^{13}C-^1H spin-couplings as potential probes of DNA and RNA structure. It is clear that the present understanding of these parameters is rudimentary and that further work is needed to fully appreciate their dependence on molecular structure. The anticipated increase in the use of isotopically labeled oligonucleotides in structural biology will inevitably result in the more routine measurement of J_{CH} in these molecules, leading to the expectation that they will become more prominent in assigning solution conformation with greater confidence.

References

Altona, C. and Sundaralingam, M. (1972) *J. Am. Chem. Soc. 94,* 8205.

Angelotti, T., Krisko,M., O'Connor, T., and Serianni,A.S. (1987) *J. Am. Chem. Soc. 109,* 4464.

Angyal, S.J. (1969) *Angew. Chem. Int. Ed. Engl. 8,* 157.

Angyal, S.J. (1991) *Adv. Carbohydr. Chem. Biochem. 49,* 19.

Angyal, S.J., Bethell, G.S., and Beveridge, R. (1979) *Carbohydr. Res. 73,* 9.

Bandyopadhyay, T., Wu, J., and Serianni, A.S. (1993) *J. Org. Chem. 58,* 5513.

Barker, S.A., Somers, P.J., and Woodbury, R.R. (1977) *German Patent,* DT2726535.

Beckmann, N., Turkalj, I., Seelig, J., and Keller, U. (1991) *Biochemistry 30,* 6362.

Bock, K. and Pedersen, C. (1974) *J. Chem. Soc. Perkin II,* 293.

Bock, K. and Pedersen, C. (1975) *Acta Chem. Scand., B29,* 258.

Bock, K. and Pedersen, C. (1977) *Acta Chem. Scand. Ser. B, B31,* 354.

Bock, K., Lundt, I., and Pedersen, C. (1973) *Tett. Let.,* 1037.

Bottomley P. A., (1989) *Radiology 170,,* 1.

Carmichael, I., Chipman, D.M., Podlasek, C.A., and Serianni, A.S., (1993) *J. Am. Chem. Soc. 115,* 10863.

Clark, E.L. Jr. and Barker, R. (1986) *Carbohydr. Res. 153,* 253.

Clark, E.L., Hayes, M.L., and Barker, R. (1986) *Carbohydr. Res. 153,* 263.

Clore, G.M. and Gronenborn, A.M. (1991)*Prog. NMR Spectros. 23,* 43.

Cumming, D.A., Dime, D.S., Grey, A.A., Krepinsky, J.J., and Carver, J.P. (1986) *J. Biol. Chem. 261,* 3208.

Duker, J. and Serianni, A.S. (1993) *Carbohydr. Res. 249,* 281.

Enzyme Nomenclature (1992), Academic Press, San Diego.

Fesik, S.W. and Zuiderweg, E.R.P. (1990) *Quart. Rev. Biophys. 23,* 94.

Fischer, E. (1889) *Ber. Dtsch. Chem. Ges. 22,* 2204.

Garrett, E.C. and Serianni, A.S. (1990a,) *Carbohydr. Res. 206,* 183.

Garrett, E.C. and Serianni, A.S. (1990b), "*Ab Initio* Molecular Orbital Calculations on Carbohydrates: Conformational Properties of Deoxygenated Furanose Sugars," in *Computer Modeling of Carbohydrate Molecules,* J. Brady and A. French, eds., ACS Symposium Series 430, American Chemical Society, p. 91.

Haasnoot, C.A.G., de Leeuw, F.A.A.M., and Altona, C. (1980) *Tetrahedron 36,* 2783.

Haasnoot, C.A.G., de Leeuw, F.A.A.M., de Leeuw, H.P.M., and Altona, C. (1981) *Org. Magn. Reson. 15,* 43.

Hayes, M.L., Pennings, N.J., Serianni, A.S., and Barker, R. (1982a) *J. Am. Chem. Soc. 104,* 6764.

Hayes, M.L., Serianni, A.S., and Barker, R. (1982b) *Carbohydr. Res. 100,* 87.

Hines, J.V., Varani,G., Landry, S.M., and Tinoco, I., Jr. (1993) *J . Am. Chem. Soc. 115,* 11002.

Hurd, R.E. (1990) *J. Magn. Reson. 87,* 422.

Ichikawa, Y., Lin, Y.-C., Dumas, D.P., Shen, G.-H., Garcia-Junceda, E., Williams, M.A., Bayer, R., Ketcham, C., Walker, L. E., Paulson, J.C., and Wong, C.-H. (1992) *J. Am. Chem. Soc. 114,* 9283.

Jones, J.K.N. and Wall, R.A. (1960) *Can. J. Chem. 38,* 2290.

Karplus, M. (1959) *J. Chem. Phys. 30,* 11.

Kellenbach, E.R., Remerowski, M.L., Eib, D., Boelens, R., van der Marel, G.A., van den Elst, H., van Boom, J.H., and Kaptein, R. (1992) *Nucl. Acids. Res. 20,* 653.

Kessler, H., Mronga, S., and Gemmecker, G. (1991) *Magn. Reson. Chem. 29,* 527.

King-Morris, M.J. and Serianni, A.S. (1986) *Carbohydr. Res. 154,* 29.

King-Morris, M.J. and Serianni, A.S. (1987) *J. Am. Chem. Soc. 109,* 3501.

King-Morris, M.J., Bondo, P,. Mrowca, R., and Serianni, A.S. (1988) *Carbohydr. Res. 175,* 49.

Kline, P.C. and Serianni, A.S. (1988) *Magn. Res. Chem. 26,* 120.

Kline, P.C. and Serianni. A.S. (1990a) *Mag. Res. Chem. 28,* 324.

Kline, P.C. and Serianni. A.S. (1990b) *J. Am. Chem. Soc. 112,* 7373.

Kline, P.C. and Serianni, A.S. (1992) *J. Org. Chem. 57,* 1772.

Kline, P.C., Huang, S.-G., Hayes, M.L., Barker, R., and Serianni, A.S. (1990) *Can. J. Chem. 68,* 2171.

Koch, H.J. and Stuart, R.S. (1977) *Carbohydr. Res. 59,* C1.

Koch, H.J. and Stuart, R.S. (1978) *Carbohydr. Res. 67,* 341.

Krenitsky, T.A., Koszalka, G.W., and Tuttle, J.V. (1981) *Biochemistry 20,* 3615.

Kukal, O., Serianni, A.S., and Duman, J.G. (1988) *J. Comp. Physiol. B158,* 175.

Kukal, O., Duman, J.G., and Serianni, A.S. (1989) *J. Comp. Physiol. B158,* 661.

Kushlan, D.M. and LeMaster, D.M. (1993) *J. Am. Chem. Soc. 115,* 11026.

Lancelot, G., Chanteloup, L., Beau, J.M., and Thuong, N.T. (1993) *J. Am. Chem. Soc. 115,* 1599.

de Leeuw, F.A.A.M. and Altona, C. (1983) *J. Comp. Chem. 4,* 428.

de Leeuw, H.P.M., Haasnoot, C.A.G., and Altona, C. (1980) *Isr. J. Chem. 20,* 108.

Legault, P., Farmer, B.T., II, Mueller, L., and Pardi, A. (1994) *J. Am. Chem. Soc. 116,* 2203.

Lemieux, R.U. (1963) in *Molecular Rearrangements,* P. de Mayo, ed., Wiley-Interscience, New York, p. 713.

Lemieux, R.U. (1971) *Pure Appl. Chem. 25,* 527.

Lemieux, R.U., Koto,S., and Voisin D. (1979) in Anomeric Effect: Origin and Consequences, W.A. Szarek and D. Horton, eds., ACS Symposium Series 87, American Chemical Society, Washington, D.C., p. 17.

Lobry De Bruyn,C.A. and Alberda van Ekenstein, (1895) W. *Recl. Trav. Chim. Pays-Bas 14,* 203.

London, R.E. (1987) *J. Chem. Soc. Chem. Commun. 661,* .

London, R.E. (1988) *Prog. NMR Spectrosc. 20,* 337.

Maple, S.R. and Allerhand, A. (1987) *J. Am. Chem. Soc. 109,* 3168.

Marino, J.P., Prestegard, J.H., and Crothers, D.M. (1994) *J. Am. Chem. Soc. 116,* 2205.

Mierke, D.F., Grdadolnik, S.G., and Kessler, H. (1992) *J. Am. Chem. Soc. 114,* 8283.

Montelione, G.T., Winkler, M.E., Rauenbuehler, P., and Wagner, G. (1989) *J. Magn. Reson. 82,* 198.

Müller, N. and Pritchard, D.E. (1959) *J. Chem. Phys. 31,* 768.

Mulloy, B., Frenkiel, T.A., and Davies, D.B. (1988) *Carbohydr. Res. 184,* 39.

Nikonowicz, E.P. and Pardi, A. (1992) *J. Am. Chem. Soc. 114,* 1082.

Nikonowicz, E.P., Sirr, A., Legault, P., Jucker, F.M., Baer, L.M., and Pardi, A. (1992) *Nucl. Acids. Res. 20,* 4507.

Nunez, H.A. and Barker, R. (1980) *Biochemistry 19,* 489.

Nunez, H.A., Walker, T.E., Fuentes, R., O'Connor, J., Serianni, A., and Barker, R. (1977) *J. Supramol. Struct. 6,* 535.

Nunez, H.A., O'Connor, J., Rosevear, P.R., and Barker, R. (1981) *Can. J. Chem. 59,* 2086.

Otting, G. and Wüthrich, K. (1990) *Quart. Rev. Biophys. 23,* 39.

Palmer, A.G., III, Rance, M., and Wright, P.E. (1991) *J. Am. Chem. Soc. 113,* 4371.

Pardi, A. and Nikonowicz, E.P. (1992) *J. Am. Chem. Soc. 114,* 9202.

Pierce, J., Serianni, A., and Barker, R. (1985) *J. Am. Chem. Soc. 107,* 2448.

Podlasek, C.A. and Serianni, A.S. (1994) *J. Biol. Chem. 269,* 2521.

Rosevear, P.R., Nunez, H.A., and Barker, R. (1982) *Biochemistry 21,* 1421.

Rothman, D.L., Behar, K.L., Hetherington, H.P., den Hollander, J.A., Bendall, M.R., Petroff, O.A.C., and Shulman, R.G. (1985) *Proc. Natl. Acad. Sci. USA 82,* 1633.

Salazar, M., Federoff, O.Y., Miller, J.M., Ribeiro, N.S., and Reid, B.R. (1993) *Biochemistry 32,* 4207.

Sancar, A. and Sancar, G.B (1988) *Ann. Rev. Biochem. 57,* 29.

Schwarcz, J.A., Cyr, N., and Perlin, A.S., (1972) *Can. J. Chem. 50,* 3667.

Schwarcz, J.A., Cyr, N., and Perlin, A.S., (1975) *Can. J. Chem. 53,* 1872.

Serianni, A.S. (1992) in *Glycoconjugates: Composition, Structure and Function,* H.J. Allen and E.C. Kisailus, eds., Marcel-Dekker, p. 71.

Serianni, A.S. and Barker, R. (1979d) *Can. J. Chem. 57,* 3160.

Serianni, A.S. and Barker, R. (1984) *J. Org. Chem. 49,* 3292.

Serianni, A.S. and Bondo, P.B. (1994) *J. Biomol. Struct. Dyn.* 11, 1133.

Serianni, A.S. and Podlasek, C.A. (1994) Carbohydr. Res. 259, 277.

Serianni, A.S., Nunez, H.A., and Barker, R. (1994) *Carbohydr. Res. 72,* 1979a, 71.

Serianni, A.S., Clark, E.L., Jr., and Barker, R. (1979b) *Carbohydr. Res. 72,* 79.

Serianni, A.S., Pierce, J., and Barker, R. (1979c) *Biochemistry 18,* 1192.

Serianni, A.S., Cadman, E., Pierce, J., Hayes, M.L., and Barker, R. (1982a) *Meth. Enzymol. 89,* 83.

Serianni, A.S., Clark, E.L. Jr., and Barker, R. (1982b) *Methods Enzymol. 89,* 79.

Serianni, A.S., Pierce, J., Huang, S.-G., and Barker, R. (1982c) *J. Am. Chem. Soc. 104,* 4037.

Serianni, A.S., Vuorinen, T. and Bondo, P.B. (1990) *J. Carbohydr. Chem. 9,* 513.

Sklenar, V., Peterson, R.D., Rejante, M.R., and Feigon, J. (1993) *J. Biomol. NMR 3,* 721.

Smallcombe, S.H. (1993) *J. Am. Chem. Soc. 115,* 4776.

Snyder, J.R. and Serianni, A.S. (1986) *J. Org. Chem. 51,* 2694.

Snyder, J.R. and Serianni, A.S. (1987a) *Carbohydr. Res. 163,* 169.

Snyder, J.R. and Serianni, A.S. (1987b) *Carbohydr. Res. 166,* 85.

Snyder, J.R. and Serianni, A.S. (1988) *Carbohydr. Res. 184,* 13.

Snyder, J.R. and Serianni, A.S. (1991) *Carbohydr. Res. 210,* 21.

Snyder, J.R., Johnston, E.R., and Serianni, A.S. (1989) *J. Am. Chem. Soc. 111,* 2681.

Stafford, U.L., Serianni, A.S., and Varma, A. (1990) *Am. Inst. Chem. Eng. J. 36,* 1822.

Sowa, T. and Ouchi, S. (1975) *Bull. Chem. Soc. Jpn. 48,* 2084.

Storey, K.B., Baust, J.G., and Storey, J.M. (1981) *J. Comp. Physiol. 144,* 183.

Tanase, T., Shimizu, F., Kuse, M., Yano, S., Hidai, M., and Yoshikawa, S. (1988) *Inorg. Chem. 27,* 4085.

Tran, V.H. and Brady, J.W. (1990) *Biopolymers 29,* 961.

Tvaroska, I. and Bleha, T. (1989) *Adv. Carbohydr. Chem. Biochem. 47,* 45.

Tvaroska, I., Hricovini, M., and Petrakova, E. (1989) *Carbohydr. Res. 189,* 359.

Vorbrüggen, H., Krolikiewicz. K., and Niedballa, U. (1975) *Liebigs Ann. Chem.* 988.

Vuister, G.W., Delaglio, F., and Bax, A. (1992) *J. Am. Chem. Soc. 114,* 9674.

Vuorinen, T. and Serianni, A.S. (1990a) *Carbohydr. Res. 207,* 185.

Vuorinen, T. and Serianni, A.S. (1990b) *Carbohydr. Res. 209,* 13.

Westhof, E. and Sundaralingam, M. (1983) *J. Am. Chem. Soc. 105,* 970.

Wijmenga, S.S., Hallenga, K., and Hilbers, C.W. (1989) *J. Magn. Reson. 84,* 634.

Wilde, J.A., Bolton, P.H., Mazumder, A., Manoharan,M., and Gerlt, J.A. (1989) *J. Am. Chem. Soc. 111,* 1894.

Wu, T. and Ogilvie, K.K. (1990) *J. Org. Chem. 55,* 4717.

Wu, J. and Serianni, A.S. (1991a) *Carbohydr. Res. 210,* 51.

Wu, J. and Serianni, A.S. (1991b) *Carbohydr. Res. 211,* 207.

Wu, G., Serianni, A.S., and Barker, R. (1983) *J. Org. Chem. 48,* 1750.

Wu, J., Serianni, A.S., and Vuorinen, T. (1990) *Carbohydr. Res. 206,* 1.

Wu, J., Bondo, P.B., and Serianni, A.S. (1992a) *Carbohydr. Res. 226,* 261.

Wu, J., Bondo, P.B., Vuorinen, T,. and Serianni, A.S. (1992b) *J. Am. Chem. Soc. 114,* 3499.

Wu, J. and Serianni, A.S. (1994) *Biopolymers 34,* 1175.

Yoshikawa, M., Kato, T., and Takenishi, T. (1967) *Tett. Let. 50,* 5065.

RNA Structure and Scalar Coupling Constants

Ignacio Tinoco, Jr., Zhouping Cai, Jennifer V. Hines,
Stacy M. Landry, John SantaLucia, Jr., L. X. Shen, and Gabriel Varani

Department of Chemistry, University of California
and
Division of Structural Biology, Lawrence Berkeley Laboratory
Berkeley, CA 94720

Signs and magnitudes of scalar coupling constants—spin-spin splittings—comprise a very large amount of data that can be used to establish the conformations of RNA molecules. Proton-proton and proton-phosphorus splittings have been used the most, but the availability of ^{13}C- and ^{15}N-labeled molecules allow many more coupling constants to be used for determining conformation. We will systematically consider the torsion angles that characterize a nucleotide unit and the coupling constants that depend on the values of these torsion angles. Karplus-type equations have been established relating many three-bond coupling constants to torsion angles (Altona, 1982). However, one- and two-bond coupling constants can also depend on conformation. Serianni and coworkers measured carbon-proton coupling constants in ribonucleosides and have calculated their values as a function of conformation (Kline and Serianni, 1990; Wu and Serianni, 1992; Serianni, this volume). The signs of two-bond coupling constants can be very useful (Hines *et al.*, 1993) because it is easier to measure a sign than an accurate magnitude.

UUCG Tetraloop

We will use the extra-stable RNA tetraloop [5′-GGAC(UUCG)GUCC] as a model molecule. The 12-nucleotide RNA has a stem of four base pairs with a hairpin loop of four nucleotides. It is a compact molecule with a well-defined conformation that includes a wide range of torsion angle values. The conformation was first determined from NOEs and scalar couplings (Varani *et al.*, 1991) using distance

geometry (DSPACE, Hare Research, Inc.) with 415 distance constraints: 125 intranucleotide interproton distances, 157 internucleotide proton distances, 12 hydrogen bonds, and 121 distances calculated from 86 torsion angle constraints. A similar conformation was also obtained from the same NMR data by using restrained molecular dynamics (Wimberly, 1992). NOE distance constraints and torsion angle constraints from the coupling constant data were used in X-PLOR (Nilges *et al.*, 1988) to obtain possible structures starting from a conformation with random torsion angles. The structures obtained by the two methods agreed well; a typical structure is shown in Fig. 1.

The stem and loop contain a wide range of torsion angles. There is, of course, typical A-form double-helix geometry with C3′-endo ribose sugars and *anti* bases. But a *syn* guanosine also occurs in the loop, two loop riboses are C2′ endo as in B-form DNA, and there is an abrupt bend in the backbone in the loop. All of these different torsion angles can be related to characteristic scalar coupling constants. Figure 2 illustrates the six torsion angles that characterize the polynucleotide backbone in RNA. The qualitative description of each angle in A-form double-helix geometry is indicated on the left as gauche minus ($g^- = -60°$), *trans* ($t = \pm 180°$), or gauche plus ($g^+ = +60°$). Newman projections show the conformations found in A-form double helices; the torsion angles are given on the right. The Newman projections indicate the three-bond couplings that can characterize the torsion angles. We expect large coupling constants for nuclei that are *cis* or *trans* to each other and small values for those nuclei that are close to 90°.

One-, two-, and three-bond coupling constants were measured for the UUCG hairpin from natural abundance samples (Varani *et al.*, 1991; Varani and Tinoco, 1991a), and from a sample enriched to 30% in ^{13}C (Hines *et al.*, 1993; Hines *et al.*, 1994). The 30% enriched sample makes it possible to measure two- and three-bond carbon-proton coupling constants without interference from carbon-carbon splittings. The methods used for measuring the splittings are given in the original papers cited and in a review (Varani and Tinoco, 1991b).

Three-Bond Coupling Constants

Table I gives the average torsion angles found for the two internal base pairs of the four base-pair stem in the UUCG hairpin (G^2, A^3, U^{10}, and C^{11}). They are very similar to the standard A-form geometry obtained from x-ray diffraction (Saenger, 1984). Average three-bond coupling constants that characterize the torsion angles are also given (see Fig. 2). Angles α and ζ on either side of phosphorus do not affect any three-bond couplings. They are related qualitatively to phosphorus chemical shifts (Gorenstein, 1984), so for A-form helices, all phosphorus resonances should be within 1 ppm. Torsion angle β (O5′-C5′) for an A-form helix leads to equal and small (<3 Hz)

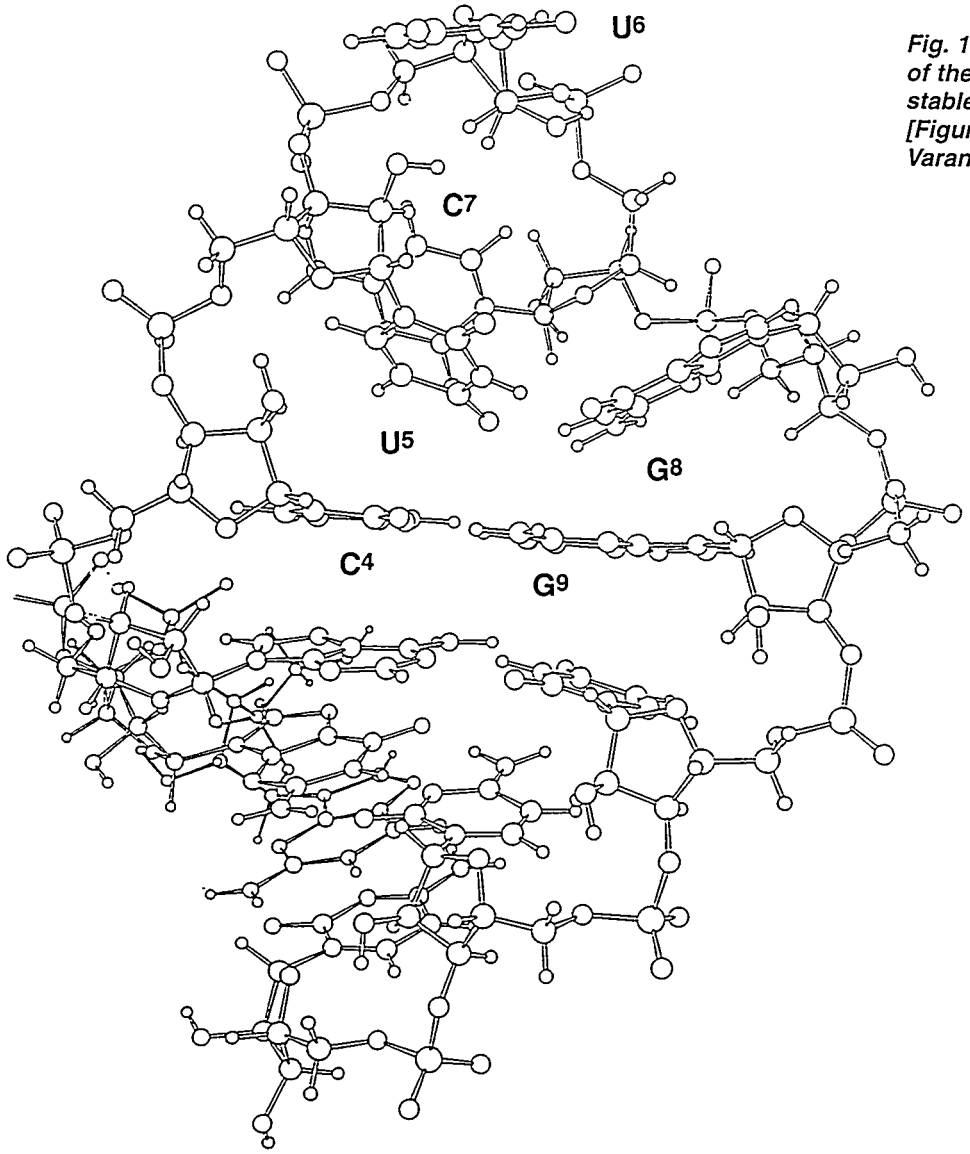


Fig. 1. The structure of the UUCG extra-stable hairpin [Figure taken from Varani et al. *(1991].*

splittings for P-H5′ and P-H5″. Similarly, γ (C5′-C4′) has equal and small (<3 Hz) splittings for H4′-H5′ and H4′-H5″. We could not measure $^3J_{C3',H5'}$ because of overlap of cross peaks, but it should be larger than the 1-Hz splitting measured for C3′-H5″; C3′ is *trans* to H5′ but gauche to H5″. Torsion angle δ (C4′-C3′) corresponds to the C3′-endo ribose pucker of A-form helices; it is characterized by a large (10-Hz) H3′-H4′ splitting. Other ribose coupling constants that also depend on sugar pucker will be discussed in a later section. Torsion angle ε for

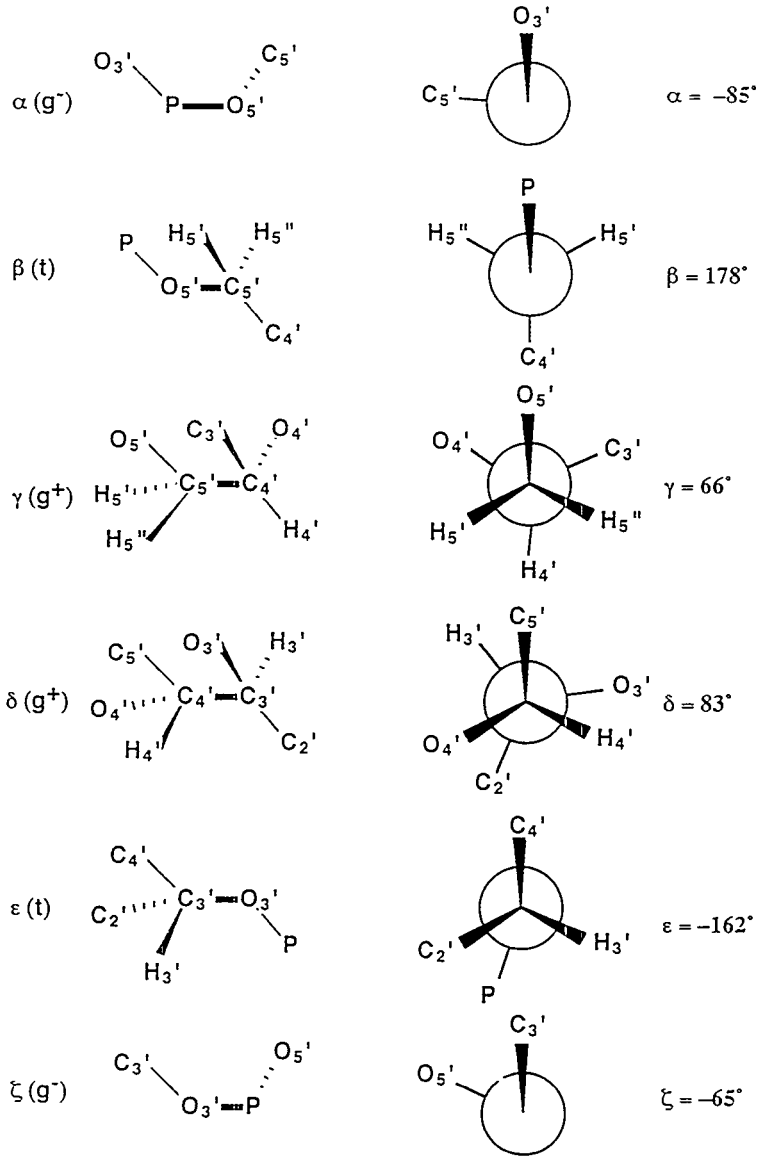


Fig. 2. The six torsion angles that determine the backbone conformation of an RNA. Newman projections on the right indicate which three-bond coupling constants depend on each torsion angle. The angles found in A-form double helices are also shown.

Table I. Double-Stranded A-Form Conformation and Three-Bond Scalar Coupling Constants (Hz) [a]

	α (P-O5')	β (O5'-C5')	γ (C5'-C4')	δ (C4'-C3')	ε (C3'-O3')	ζ (O3'-P)
A-Form X-Ray	–85°	178°	66°	83°	–162°	–65°
A-Form NMR	–68°	178°	54°	84°	–153°	–71°
^3J (Hz)	none	P-H5' = <3	H4'-H5' = <3	H3'-H4' = 10	P-H3' = 8	none
^3J (Hz)	none	P-H5" = <3	H4'-H5" = <3	C2'-H4' = ?	P-C2' = ?	none
^3J (Hz)	none	P-C4' = ?	C3'-H5' = ?	C5'-H3' = ?	P-C4' = ?	none
^3J (Hz)	none		C3'-H5" = 1			none

[a]The torsion angles (labeled A-form NMR) and splittings are averages of two base pairs (A•U, G•C) in the middle of the stem of the UUCG hairpin molecule.

A-form geometry has an 8-Hz H3'-P splitting. In A-form double helices, there is a four-bond H4'-P coupling of ~3 Hz; this provides further evidence for the *trans*, gauche conformation of β and γ. ^{13}C-P coupling constants are also very informative in determining the backbone torsion angles (Schmieder *et al.*, 1992).

As is evident from Table I, there are many carbon-proton and carbon-phosphorus three-bond coupling constants that we did not obtain. However, a sufficient criterion for A-form duplexes in RNA with C3'-endo sugars is $^3J_{P,H5'} \approx {}^3J_{P,H5''} < 3$ Hz; $^3J_{H4',H5'} \approx {}^3J_{H4',H5''} < 3$ Hz; $^3J_{H3',H4'} \approx 10$ Hz; $^3J_{P,H3'} \approx 8$ Hz. Coupling constants outside of this range, as seen in the loop nucleotides and in the loop-closing base pair, indicate torsion angles other than β(t), γ(g$^+$), δ(g$^+$), and ε(t). G^9, in the loop-closing base pair, has $^3J_{P,H5'} \approx 18$ Hz and $^3J_{P,H5''} \approx 8$ Hz; it has β = –116° ± 16°. G^8 has γ = 169° ± 10°, but overlap prevented obtaining useful ^3J values. U^6 and C^7 are C2' endo (δ ≈ 150°); they have $^3J_{H3',H4'} < 2$ Hz.

Other proton-proton three-bond coupling constants further characterize the sugar conformation. This is best seen in Fig. 3, where characteristic splittings are shown for H1'-H2', H2'-H3', and H3'-H4' for C3'-endo and C2'-endo conformations. de Leeuw and Altona (1982) obtained empirical equations relating each of these ^3J values to P, the pseudorotation phase angle for the ribose ring. A value of P = 18° corresponds to C3'-endo; P = 162° is C2'-endo. Their results are shown in

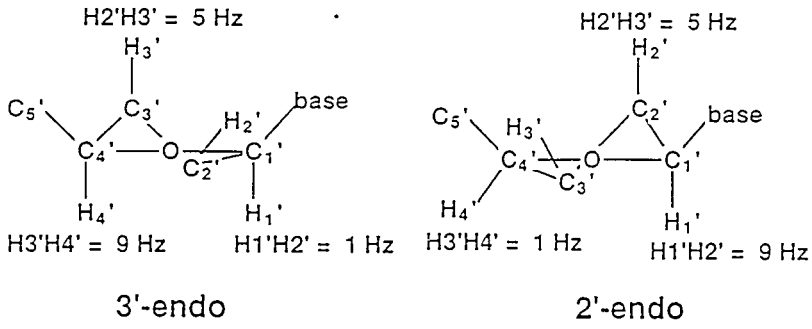


Fig. 3. Typical three-bond proton-proton coupling constants in C3'-endo and C2'-endo conformations.

Fig. 4 for an amplitude of pucker of 40°; the amplitude has been found to be nearly constant for RNA mono- and oligonucleotides. When the three coupling constants are not consistent with one value of P, we assume a mixture in fast exchange of C2′-endo and C3′-endo conformers. This usually occurs for the nucleotides at the 5′ and 3′ ends of the RNA and may occur in loop nucleotides. Even if a single conformer (not C2′-endo or C3′-endo) is consistent with the three proton-proton ^3J values, NOE magnitudes may be able to rule it out. However, it will be also helpful to have carbon-proton coupling constants to further characterize the

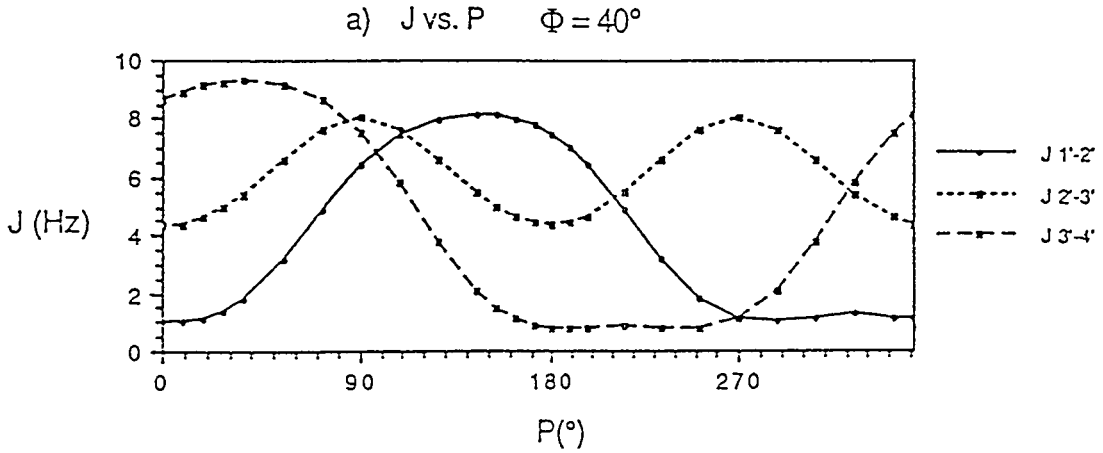


Fig. 4. The variation of three-bond proton-proton coupling constants with pseudorotation phase angle for H1'-H2', H2'-H3' and H3'-H4' [Figure from deLeeuw and Altona (1982) as presented in Davis (1989)].

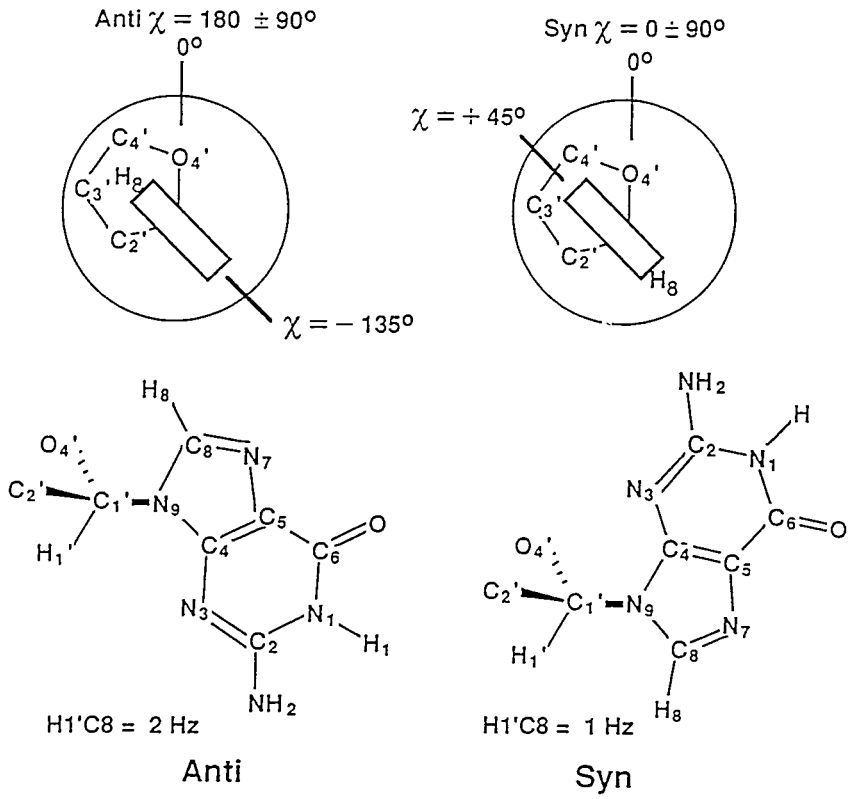


Fig. 5. The definition of the glycosidic angle, chi, for anti *and* syn *orientations of a base relative to a ribose in a nucleotide. The three-bond carbon-proton coupling constants do not differ much for the two* anti *and one* syn *conformations of the guanine nucleotides in the UUCG hairpin.*

ribose conformations and dynamics. The two loop nucleotides U^6 and C^7 that are C2′-endo have ${}^3J_{C1',H3'} = 6.6$ Hz and ${}^3J_{C3',H1'} = 0.6$ Hz. These values are consistent with expected relative magnitudes (see Fig. 3), but weak crosspeaks and spectral overlap prevented us from obtaining values for the C3′-endo nucleotides for comparison.

In principle, the glycosidic chi angle can be obtained from ${}^3J_{H1',C8}$, as shown in Fig. 5. However, we found that all nucleotides in the UUCG hairpin had values of ${}^3J_{H1',C8} = 2$ Hz ± 1 Hz. The nucleotide with the *syn* base (G^8) did not have a coupling constant that was significantly different from the others.

Two-Bond Coupling Constants

Two-bond coupling constants have the advantage of being positive or negative. Thus, determination of sign may be sufficient to obtain useful information. The most important new information that can be determined from the sign of carbon-proton two-bond splittings is the identification of H5′ and H5″ (Hines *et al.*,

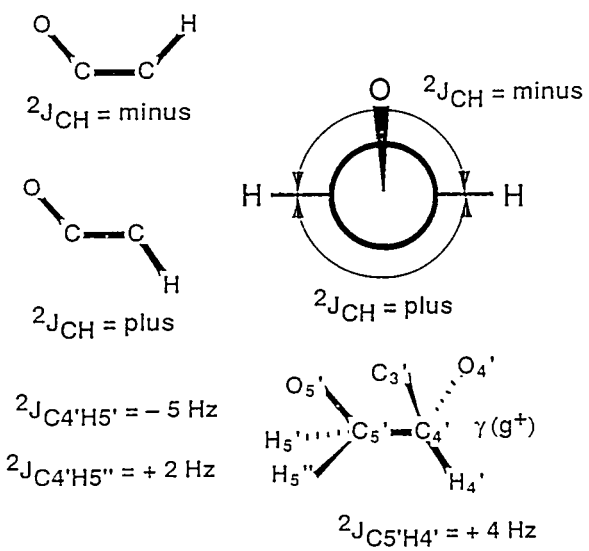


Fig. 6. The signs of two-bond carbon-proton coupling constants and the torsion angle X-C-C-H, where X is an electronegative substituent. $^2J_{CH}$ is negative for torsion angle = 0° ± 90°; it is positive for torsion angle 180° ± 90°. This fact allows stereospecific assignment of H5' and H5".

1993), as is illustrated in Fig. 6. An electronegative substituent on a carbon produces a negative $^2J_{CH}$ when the proton is *cis* to the substituent and a positive $^2J_{CH}$ when it is *trans*. The signs of $^2J_{C5',H4'}$, $^2J_{C5',H4'}$ and $^2J_{C4',H5''}$ allow the characterization of torsion angle γ and stereospecific assignment of H5' and H5". This greatly enhances the usefulness of NOEs involving H5' and H5" in the determination of structure.

In standard A-form double helices, the chemical shift of H5' is more downfield than H5". But in loop regions, or where there is tertiary structure or interactions with other molecules, a definitive assignment method is needed. This is illustrated for the two-bond carbon-proton coupling constants in Table II. We see that in loop nucleotides C^7 and G^8, the chemical shift of H5' is at higher field than H5"; all the other nucleotides have the usual order (H5' is at lower field than H5"). The large upfield shifts of both H5' and H5" for C^7 are caused by the ring current of G^8; H5' is 0.9 ppm more shielded than H5". For G^8, the smaller difference of 0.2 ppm probably results from the *trans* value of torsion angle γ.

The signs of two-bond coupling constants also characterize the ribose pucker, as is shown in Fig. 7. The signs depend on whether each proton is *cis* or *trans* to O2', O3', or O4', as illustrated in Fig. 6. The five torsion angles around the ribose ring are related to the pseudorotation phase angle P by

$$\nu_j = \phi \cos [P + 144 (j - 2)]$$

$$j = 0, 1, 2, 3, 4 \quad .$$

Table II. Two-Bond Scalar Coupling Constants (Hz) and Stereospecific Assignment of H5' and H5"

Nucleotide	$^2J_{C5',H4'}$	$^2J_{C4',H5'}$	$^2J_{C4',H5"}$	H5' (ppm)	H5" (ppm)
U^6 (g = g^+)	+3.8	–5.3	+1.9	4.19	4.01
C^7 (g = g^+)	+4.3	–5.4	+1.9	2.67	3.57
G^8 (g = t)	–3.7	+1.2	–4.8	4.15	4.38

The amplitude ϕ is the maximum torsion angle; it is nearly constant at a value of 40°. Torsion angle ν_0 refers to C4′-O4′-C1′-C2′; ν_1 = O4′-C1′-C2′-C3′, ν_2 = C1′-C2′-C3′-C4′, etc. We can now calculate the sign of each $^2J_{C',H'}$ as a function of conformation. $^2J_{C2',H1'}$ and $^2J_{C3',H4'}$ should always be negative, because the configuration of the substituents on C2′ and C1′ makes the O2′-C2′-C1′-H1′ torsion angle equal to ν_1 and similarly the O3′-C3′-C4′-H4′ torsion angle is equal to ν_3. These torsion angles are 0° ± 40°.

Experimentally, we find that all the $^2J_{C2',H1'}$ and $^2J_{C3',H4'}$ values we could obtain for the UUCG loop are indeed negative. The conformationally dependent two-bond couplings are shown below.

- $^2J_{C3',H2'}$ should be negative for P = 0° ± 40°
- $^2J_{C2',H3'}$ should be negative for P = 180° ± 40°
- $^2J_{C4',H3'}$ should be negative for P = 35° ± 40°

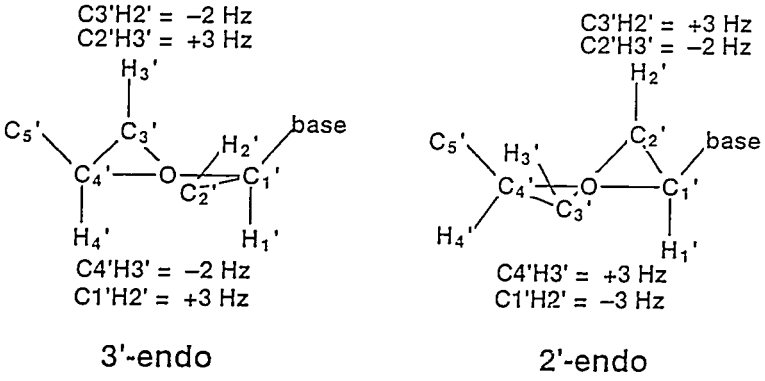


Fig. 7. The signs of two-bond carbon-proton coupling constants and ribose conformations. The signs of each of four $^2J_{CH}$ coupling constants indicate a C3'-endo or C2'-endo conformation.

Outside these ranges, the coupling constants should be positive. As C3′-endo has a P = 18°, and C2′-endo has a P = 162°; these signs are consistent with the experimental measurements shown in Fig. 7. The sign of $^2J_{C1',H2'}$ is more difficult to assess because C1′ has two electronegative substituents: O4′ and N1 or N9. Experimentally, we found that $^2J_{C1',H2'}$ was negative for C3′-endo and positive for C2′-endo; this indicates that O4′ is dominant.

It is clear that measurements of $^2J_{CH}$ can be very useful in determining the conformation of the ribose ring.

One-Bond Coupling Constants

One-bond coupling constants depend on conformation—probably through slight changes in bond lengths with changes in structure. Table III shows the changes that occur when ribose switches from C3′-endo to C2′-endo. The average of one-bond coupling constants for four pyrimidine nucleotides (C^4, U^5, U^{10}, and C^{11}) in C3′-endo conformation are compared with the two pyrimidine nucleotides that are C2′-endo. There are significant differences in three of the four types of $^1J_{CH}$ values measured.

To test for a possible effect of a *syn* base on one-bond coupling constants, we compared two C3′-endo, *anti* nucleotides (G^2, G^9) with the C3′-endo, *syn* nucleotide (G^8). The only significant difference was a 6-Hz decrease in the C1′-H1′ coupling for the *syn* G^8; $^1J_{C1',H1'}$ = 175 Hz for G^2, G^9 and $^1J_{C1',H1'}$ = 169 Hz for G^8.

A Pseudoknot and a Ribozyme

We are studying two RNA molecules which have significant biological functions, to relate their structure to their function. (1) Many retroviruses use controlled frameshifts to synthesize essential enzymes for their replication and viability. A pseudoknot is often required for this frameshifting (Atkins *et al.*, 1990). We want to learn what makes a pseudoknot special. Why is a stable hairpin loop, or other structures that cause pausing of the ribosome, not sufficient? (2) The structures of ribozymes are being extensively studied by NMR and x-ray diffraction,

Table III. One-Bond Scalar Coupling Constants (Hz) and Ribose Conformation

	$^1J_{C1',H1'}$	$^1J_{C2',H2'}$	$^1J_{C3',H3'}$	$^1J_{C4',H4'}$
3′-endo (C^4, U^5, U^{10}, C^{11})	180 ± 1	159 ± 2	143 ± 4	149 ± 4
2′-endo (U^6, C^7)	170 ± 1	147 ± 2	156 ± 3	152 ± 3
Difference	10	11	− 13	− 3

and many kinetic mechanistic studies are being done to understand how the catalytic step occurs. We chose a hairpin ribozyme that has been engineered to cleave the HIV-1 virus (Ojwang *et al.*, 1992). Both these projects are in progress; they are far from the level of resolution of the UUCG hairpin.

Pseudoknot

We have applied isotopic labeling to structural studies of a 34-nucleotide RNA pseudoknot that promotes frameshifting in a mouse mammary tumor virus messenger RNA. The proposed pseudoknot contains two stems, S1(5 bp) and S2(6 bp), and two loops, L1(2 nt) and L2(8 nt). The two stems are separated by a single A nucleotide, as shown in Fig. 8. Both stems are G•C rich, which results in severe proton spectral overlap. Initial assignment of proton resonances was done by comparing the NOESY spectrum of the pseudoknot to the spectra of two tetraloop hairpin molecules that contain pseudoknot stems S1 and S2, respectively. However, the pseudoknot and the hairpins have different loops, and the resonance assignment for the pseudoknot loop and loop-stem junction regions cannot be obtained by direct comparison of the NOESY data. To reduce spectral overlap, we used a ^{13}C-editing strategy in two-dimensional NMR experiments to confirm assignments and to obtain further proton assignments for the pseudoknot.

Two selectively ^{13}C-labeled pseudoknot samples were made. One was ^{13}C-labeled at cytosine-C6; the other was labeled at

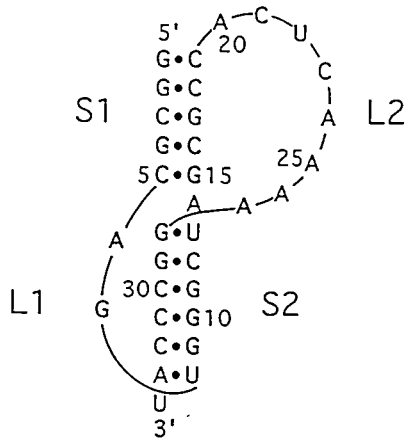


Fig. 8. A pseudoknot that causes an efficient minus-one frameshift during translation of a mouse mammary tumor virus RNA.

adenine-C8 and uracil-C6. In addition, 1/2 X-filtered NOESY (Otting and Wüthrich, 1990) experiments were done on each of the selectively ^{13}C-labeled pseudoknot samples. Two subspectra were obtained from each experiment, a ^{13}C and a ^{12}C subspectrum. Both of the spectra were greatly simplified compared to the NOESY spectrum of the unlabeled pseudoknot. The ^{13}C spectrum of the [6-^{13}C]cytosine-labeled pseudoknot contains only NOEs from 6Hs of cytosines to other protons. These NOEs firmly established 6H, 5H, and H1' assignments for all C nucleotide residues except C^5 and C^{12}. The ^{13}C spectrum of the [8-^{13}C]-adenine and [6-^{13}C]uracil-labeled pseudoknot contains NOEs from ^{13}C-bound 8H and 6H protons of A and U residues to other sugar protons. The ^{12}C spectrum of this sample has NOEs from 8H and

6H protons of G and C residues to other protons. Comparison of the two spectra distinguished A 8Hs from G 8Hs/A 2Hs. This led to the sequential 8H-H1′ assignments for residues A^{24} through A^{27} in loop 2, and for residues G^{9} through G^{11} in stem 2. Out of four U residues in the pseudoknot sequence, only two have strong 6H-5H NOEs and corresponding 6H-H1′ NOEs in the ^{13}C spectrum. These NOEs confirmed assignments of 6H, 5H, and H1′ resonances for U^{22} and U^{34}. Weak crosspeaks from residues U^{8} and U^{13} were attributed to broad linewidths of their proton resonances. Comparison of 1/2 X-filtered NOESY spectra with the NOESY spectrum of the unlabeled pseudoknot facilitated sequential assignments of H2′ protons.

Spectral overlap in the sugar proton region made resonance assignment difficult beyond H2′ protons. We therefore hoped to spread out proton resonances by introducing a ^{13}C dimension. Uniformly ^{13}C- and ^{15}N-labeled pseudoknot RNA molecules were synthesized by using labeled nucleotides obtained from bacterial *Methylophilus methylotrophus* cells. Two-dimensional-constant time heteronuclear single quantum coherence (CT-HSQC) experiments were done to help identify proton resonances by dispersing them along the ^{13}C dimension. Unlike proteins and smaller RNA molecules, the ^{13}C dimension of the pseudoknot CT-HSQC spectrum has severe overlap for each type of carbon resonance (C1′, C2′, etc.). The two-dimensional spectral resolution is poor because of the overlap of resonances in both dimensions. Improvement in resolution can be achieved by using three-dimensional heteronuclear techniques. It remains to be seen how useful 100% uniformly labeled molecules will be in structure determination of large RNA molecules (>40 nucleotides).

A Hairpin Ribozyme

Ribozymes catalyze site-specific RNA cleavage and ligation reactions. We are studying the structure of a hairpin ribozyme derived from the minus strand of tobacco ring spot virus satellite RNA [(-)sTRSV], which has been engineered to specifically cleave the HIV-1 RNA in the leader sequence (Ojwang *et al.*, 1992). The essential sequence and the secondary structure of the ribozyme have been identified by *in vitro* selection experiments (Berzal-Herranz *et al.*, 1993). The minimal sequence of the system involves a 50-nucleotide ribozyme that specifically cleaves and ligates a 14-nucleotide substrate. The proposed secondary structure of the ribozyme-substrate complex consists of four short helices separated by two internal loops. The 64-nucleotide ribozyme-substrate complex is very large for obtaining a detailed three-dimensional structure by NMR. Therefore, the smaller structural subunits of the complex are being studied first. We are determining the high-resolution solution structure of a 28-nucleotide RNA that includes the substrate binding site of the ribozyme and the cleavage site of the substrate. It contains a symmetric internal loop of

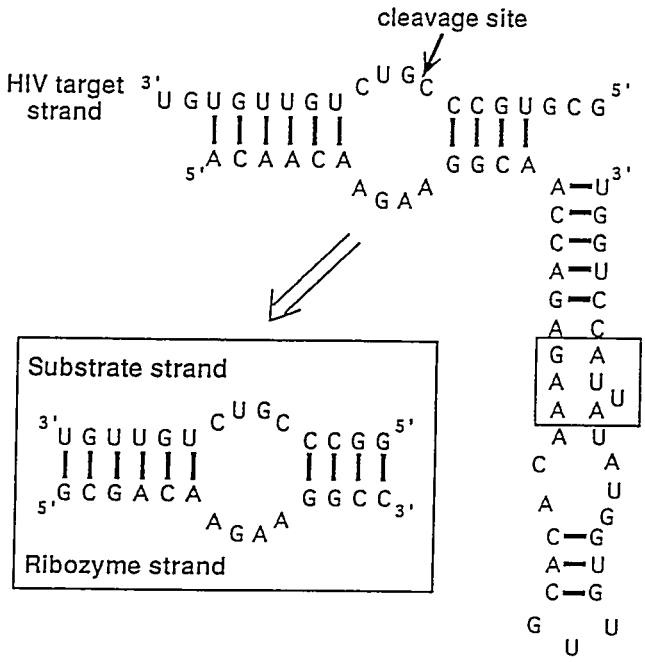


Fig. 9. A ribozyme from tobacco ring spot virus satellite that specifically cleaves HIV-1 viral RNA.

eight nucleotides and its flanking helices (Fig. 9). Spectral overlap is a substantial problem for the NMR studies of RNA molecules of this size. The NOESY spectrum for the 28-nucleotide internal loop is very crowded in some regions. For example, eight aromatic protons—seven of which come from pyrimidines—have resonances between 7.60 and 7.69 ppm. The large 5H - 6H crosspeaks from pyrimidines have made this spectral area even more crowded. The H1′ - 8H crosspeak from G^8 is completely buried underneath the 5H - 6H crosspeak from U^{28}. As a result of the severe spectral overlap, only two out of the eight resonances above were assigned. In addition, four H1′ protons resonate between 5.81 and 5.82 ppm. Therefore, the H1′ to 8/6H NOE connectivities could only extend from G^1 to A^7 on the ribozyme strand and from G^{15} to G^{17} and G^{20} to U^{28} on the substrate strand. Only two of the five 2H aromatic protons from adenine were definitively assigned.

To overcome the spectral overlap problems, we selectively ^{13}C-labeled the ribozyme strand of the internal loop at C8 of each purine and C6 of each pyrimidine. Isotope-edited two- and three-dimensional experiments were applied to facilitate resonance assignments. The ^{13}C-edited NOESY experiment greatly simplified the spectrum and allowed unambiguous

sequential assignments of aromatic and H1′ protons along each strand. The ^{13}C-edited NOESY also helped assign and confirm the other sugar proton resonances. Three-dimensional ^{1}H-^{13}C HMQC-NOESY experiments spread out the normal NOESY spectrum in the ^{13}C dimension and further confirmed some of the sugar resonance assignments.

Structure modeling based mainly on NOE constraints is proceeding with the pseudoknot and the ribozyme. Specific ^{13}C-labeling was helpful in assigning the spectra, particularly in the ribozyme made of two strands. We plan to use specific labeling to measure dynamics in different parts of each molecule by examining the relaxation of each ^{13}C by its bound proton. It is too early for us to demonstrate the utility of the 100% uniformly labeled samples.

Acknowledgements

This work was supported by NIH GM 10840, DOE DE-FG03-86ER60406, and instrumentation grants DOE DE-FG05-86ER75281, NSF DMB 86-09305 and NSF BBS 86-20134. We would like to thank the Stable Isotope Resource at Los Alamos National Laboratory for supplying us with the ^{13}C-labeled bacteria; their work made this research possible. We would also like to thank Dr. Jeffrey Pelton for informative NMR discussions, Mr. David Koh for DNA synthesis, and Ms. Barbara Dengler for running an efficient laboratory.

References

Altona, C. (1982) *Recl. Trav. Chim. Pays-Bas 101,* 413-433.

Atkins, J.F., Weiss, R.B, and Gesteland, R.F. (1990) *Cell 62,* 413-423.

Berzal-Herranz, A., Simpson, J., Chowrira, B.M., Butcher, S.E., and Burke, J.M. (1993) *EMBO J. 12,* 2567-2574.

P. Davis (1989) Ph. D. Thesis, Univ. of California, Berkeley.

de Leeuw, F.A.A. and Altona, C. (1982) *J. Chemical Soc., Perkin Transactions 2,* 375-384.

Gorenstein, D.G., Editor (1984) *Phosphorus-31 NMR: Principles and Applications,* Academic Press, New York.

Hines, J.V., Landry, S.M., Varani, G., and Tinoco, I., Jr. (1994) *J. Am. Chem. Soc. 116,* 5823-5831.

Hines, J.V., Varani, G., Landry, S.M., and Tinoco, I., Jr. (1993) *J. Am. Chem. Soc. 115,* 11002-11003.

Kline, P.C. and Serianni, A.S., (1990) *J. Am. Chem. Soc. 112,* 7373-7381 .

Nilges, M., Clore, G.M., and Gronenborn, A.M. (1988) *FEBS Lett. 239,* 129-136.

Ojwang, J.O., Hampel, A., Looney, D.J., Wong-Stall, F., and Rappaport, J. (1992) *Proc. Natl. Acad. Sci. USA 89,* 10802-10806.

Otting, G. and Wüthrich, K. (1990) *Quart. Rev. Biophys. 23,* 39-96.

Saenger, W. (1984) *Principles of Nucleic Acid Research,* Springer-Verlag, New York.

Schmieder, P., Ippel, J.H., van den Elst, H., van der Marel, G.A., van Boom, J.H., Altona, C., and Kessler, A. (1992) *Nucleic Acids Res. 20,* 4747-4751.

Varani, G. and Tinoco, I., Jr. (1991a) *J. Am. Chem. Soc. 113,* 9349-9354.

Varani, G. and Tinoco, I., Jr. (1991b) *Quarterly Revs. Biophys. 24,* 479-532.

Varani, G., Cheong, C., and Tinoco, I., Jr. (1991) *Biochemistry 30,* 3280-3289.

Wimberly, B. (1992) Ph. D. Thesis, University of California.

Wu, J. and Serianni, A.S. (1992) *Carbohydrate Research 226,* 209-219.

Neutron Scattering with Deuterium Labeling Reveals the Nature of Complexes Formed by Ca^{2+}-Binding Proteins and their Regulatory Targets

Jill Trewhella

Chemical Science and Technology Division
Los Alamos National Laboratory
Los Alamos, NM 87545

Small-angle neutron scattering with deuterium labeling is extremely useful for studying the structures of complex biomolecular assemblies in solution. The different neutron scattering properties of the isotopes of hydrogen combined with the ability to uniformly label biomolecules with deuterium allow one to characterize the structures and relative dispositions of the individual components of an assembly using methods of "contrast variation." We have applied these techniques to studies of the evolutionarily related dumbbell-shaped Ca^{2+}-binding proteins calmodulin and troponin C and their interactions with the target proteins whose activities they regulate (Heidorn *et al.*, 1989; Trewhella *et al.*, 1990; Olah *et al.*, 1994; Olah and Trewhella, 1994). Ca^{2+} is one of the simplest of nature's messengers used in the communication pathways between physiological stimulus and cellular response. The signaling mechanism generally involves Ca^{2+} binding to a protein and inducing a conformational change that transmits a signal to modify the activity of a specific target protein. Ca^{2+} is thus important in the regulation of a diverse array of intracellular responses, including neurotransmitter release, muscle contraction, the degradation of glycogen to glucose to generate energy, microtubule assembly, membrane phosphorylation, etc. It is the conformational language of the Ca^{2+} induced signal transduction that we have sought to understand because of its central importance to biochemical regulation and, hence, to healthy cellular function.

Small-Angle Neutron Scattering

Neutrons (and x-rays) have the properties of plane waves. Small-angle scattering of neutrons (and x-rays) results from the

constructive interference of secondary waves that are scattered when a plane wave interacts with matter (for review, see Heidorn and Trewhella, 1990; Glatter, 1982). Small-angle scattering from a particle in solution is a maximum at zero scattering angle and falls off with a rate that depends upon the size and shape of the scattering particle: the larger the particle, the faster the falloff. The intensity of the scattering from a particle in solution depends upon its "contrast," that is, the difference in scattering density between the particle and the solvent. Scattering densities are readily calculated by summing the scattering amplitudes of each atom within a volume and dividing by that volume.

Neutrons are neutral particles and therefore interact principally with the atomic nuclei in a sample. Hence, neutron scattering amplitudes depend upon the complex properties of the neutron-nucleus interaction and show no systematic dependence on atomic number. Further, isotopes of the same element can have very different neutron scattering properties. In contrast, x-rays are scattered by the electrons in a sample; therefore x-ray scattering amplitudes increase monotomically with increasing atomic number, and there are no isotope effects. For neutrons, one of the largest differences in neutron scattering amplitude is between the isotopes of hydrogen ^{1}H and ^{2}H. Table I lists the coherent, elastic neutron scattering amplitudes for the atoms commonly found in biological systems. Note the scattering amplitude for ^{1}H is negative, resulting from a 180° phase shift between the incident and scattered neutron. Thus, selective deuteration of one component of a complex in solution provides a way of altering the mean neutron scattering density of that component. Further, by changing the deuterium level in the solvent, the neutron scattering contrast of each component is varied. If internal scattering density fluctuations are negligible, then solvent matching can be achieved by adjusting the deuteration level in the solvent so that the mean solvent and particle (or component) scattering densities are the same; that is, there is zero contrast and, thus, no small-angle scattering from the particle, rendering it "invisible" in the neutron experiment. Contrast variation techniques provide methods for extracting structural information on the individual components of a complex and their relative dispositions (Ibel and Suhrmann, 1975; Moore, 1981).

Table 1. Neutron Scattering Amplitudes, b (10^{-12} cm), for Atoms Commonly Found in Biomolecules

Nucleus	b
^{1}H	-0.38
^{2}H	0.67
^{12}C	0.66
^{14}N	0.94
^{16}O	0.58
^{31}P	0.51

The scattering from a homogeneous solution of monodisperse particles can be expressed as:

$$I(Q) = | \int [\rho(r)-\rho_s] \bullet \exp(-iQ \bullet r)\, dr |^2 , \quad (1)$$

where Q is the momentum transfer or scattering vector and can be expressed as $4\pi\sin\theta/\lambda$, $\rho(r)$ and ρ_s are the scattering densities for the particle and solvent, respectively, and the integration is taken over the volume of the particle. For a two-component complex in solution, ignoring internal scattering density fluctuations, we can write the scattering as:

$$I(Q,\Delta\rho_1,\Delta\rho_2) = \Delta\rho_1^2 I_1(Q) + \Delta\rho_1\Delta\rho_2 I_{12}(Q) + \Delta\rho_1^2 I_2(Q) \quad , \quad (2)$$

where the subscripts 1 and 2 refer to each component, $\Delta\rho_{1(2)} = \rho_{1(2)}-\rho_s$, where $\rho_{1(2)}$ is the mean scattering density for component 1(2). The three terms in Eq. (2) correspond to the three basic scattering functions. $I_1(Q)$ and $I_2(Q)$ represent the scattering of components 1 and 2, respectively, and $I_{12}(Q)$ is the cross term. A set of neutron scattering measurements with different D_2O:H_2O ratios in the solvent gives a set of equations in the form of Eq. (2), which can be solved to give the three basic scattering functions.

For a homogeneous scattering particle in solution, the inverse Fourier transform of the scattering profile gives the pair-distance, or vector distribution function, P(r):

$$P(r) = 1/2\pi^2 \int I(Q)(Q \bullet r) \sin(Q \bullet r)\, dQ \quad (3)$$

P(r) is the frequency of vectors connecting small-volume elements within the entire volume of the scattering particle (Moore, 1980) and therefore goes to zero at a value corresponding to the maximum dimension of the particle, d_{max}. P(r) is extremely sensitive to the symmetry of the scattering particle, and to the relationships between domains or repeating structures (Fig. 1). The radius of gyration, R_g, for a particle is defined as the root-mean-square distance of all elemental scattering volumes from their center-of-mass, weighted by their scattering densities, and is calculated as the second moment of P(r):

$$R_g^2 = \int P(r) r^2\, d^3r \,/\, 2 \int P(r)\, dr \quad (4)$$

The zeroth moment of P(r) gives the forward scatter, I(0), which is proportional to the molecular weight of the scattering particle and is an extremely sensitive indicator, therefore, of aggregation.

Ibel and Stuhrmann (1975) showed that R_g^2 dependence on the scattering contrast can be written as:

$$R_g^2 = R_m^2 + \alpha/\Delta\rho - \beta/\Delta\rho^2 \quad , \quad (5)$$

where R_m is the R_g at infinite contrast and $\Delta\rho = \bar{\rho} - \rho_s$, where $\bar{\rho}$ is the mean scattering density for the total particle. The coefficient α is related to the second moment of the scattering density fluctuations about the mean value for the scattering particle; β is related to the square of the first moment of the density

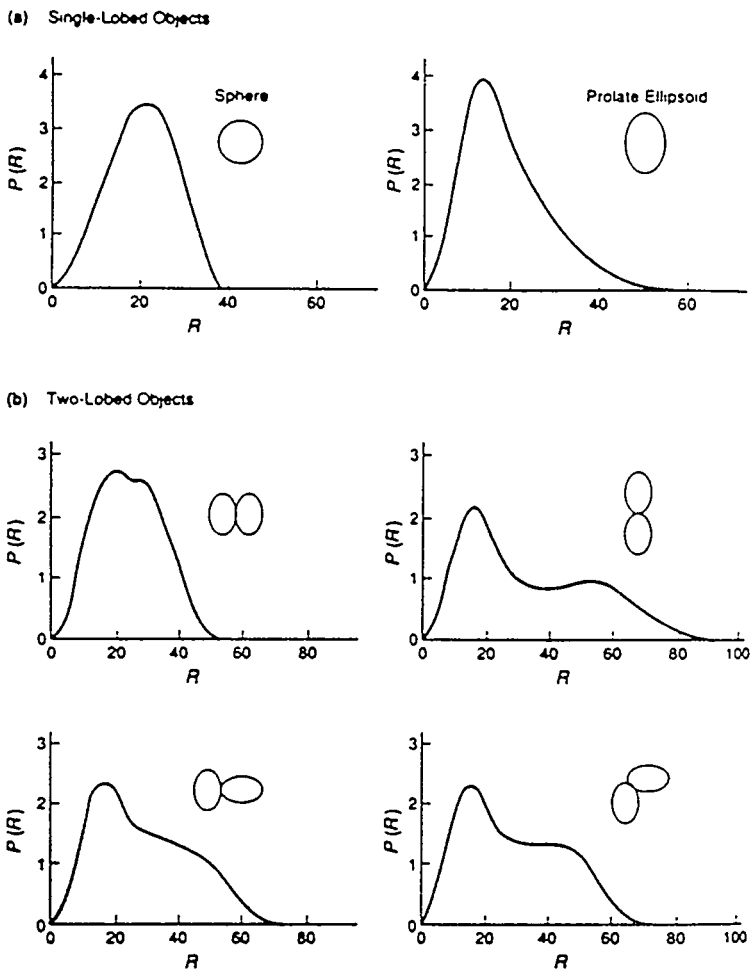


Fig. 1. The scattering profile I(Q) is related to the vector distribution function, P(r), by a Fourier transform. P(r) functions are shown for various one- and two-domain structures of uniform scattering density. The asymmetry of P(r) increases with the asymmetry of the scattering object.

fluctuations about the mean. If the sign of α is positive, then the lower scattering density component is located closer to the inside of the complex than the higher scattering density component is. A negative α indicates the reverse case. β is proportional to the square of the separation of the two components. If β is zero, then the centers-of-mass are coincident.

Finally, when one dimension of the particle is greater than the other two, Guinier (1939) showed that it is possible to approximate the scattering for certain small Q values as:

$$QI(Q) = I(0)e^{-QR_c^2/2} \quad , \qquad (6)$$

where R_c is the radius of gyration of cross-section.

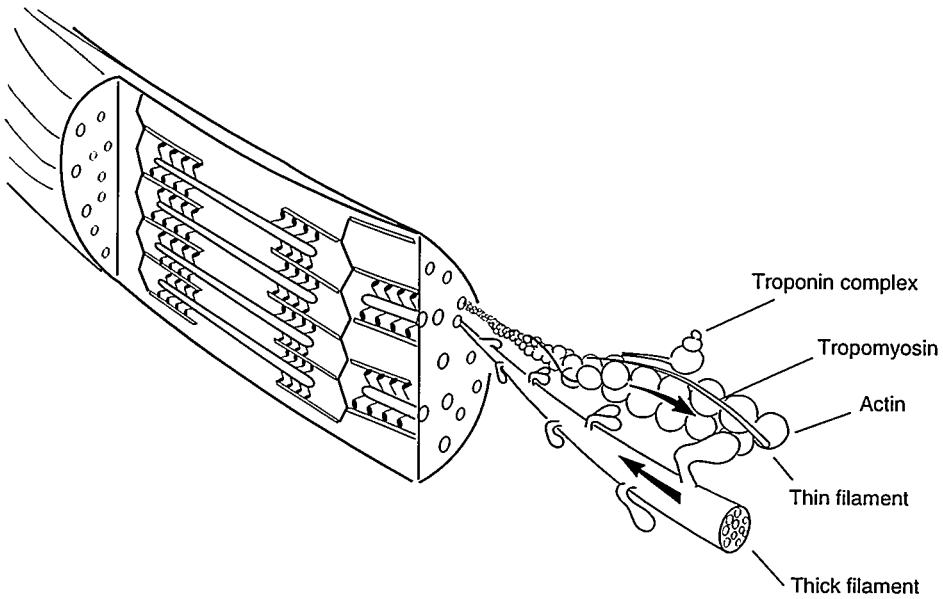


Fig. 2. Proteins compose thick and thin filaments that slide past each other, resulting in muscle contraction/relaxation. The troponin complex regulates the sliding of the thick and thin filaments via reversible binding of Ca^{2+} to TnC, which results in a signal being transmitted via TnI to modulate the interactions of the thick and thin filaments.

Troponin C and Troponin I

In the sliding-filament model of muscle action, thin and thick filaments move past each other, causing contraction or relaxation (Fig. 2). The contractile force is believed to be generated by the cyclic attachment and detachment of the myosin heads of the thick filaments to sites on the thin filaments (for review, see Leavis and Gergely, 1984; Zot and Potter, 1987). The thin filaments are composed of a double-stranded helical assembly of actin monomers. Tropomyosin is polymerized head-to-tail in the grooves of the thin filament actin helix (one tropomyosin to every 7 actin monomers), and each tropomyosin has one troponin complex bound to it. The troponin complex has three subunits: troponin C (TnC), which binds Ca^{2+}; troponin I (TnI); and troponin T (TnT). The regulation of the contraction/relaxation cycle is associated with an increase in Ca^{2+} concentration and is mediated through the TnC component of the troponin complex. When TnC binds $4Ca^{2+}$, a signal is transmitted via TnI, which releases its inhibition of the actin/myosin interaction through TnT and tropomyosin. Thus, the interactions of the thick and thin filaments are modulated to give rise to the sliding mechanism.

The only high-resolution structural data on the components of troponin is for TnC. The crystal structures of TnC (Herzberg and James, 1985; Sundaralingam *et al.*, 1985) show this component to have an unusual dumbbell shape (Fig. 3) similar to the evolutionarily related Ca^{2+}-binding protein calmodulin (Babu *et al.*, 1988). These dumbbell structures consist of two globular domains connected by a solvent-exposed α-helix of approximately 7 to 8 turns. The C-terminal domain of TnC contains two high-affinity Ca^{2+}/Mg^{2+}-binding sites, believed to be always occupied in muscle, whereas the N-terminal domain contains two lower affinity Ca^{2+}-specific binding sites. Each Ca^{2+}-binding site has the helix-loop-helix secondary structure or EF-hand motif (Kretsinger, 1980). It is Ca^{2+} binding to the N-terminal domain of TnC that regulates the contractile event.

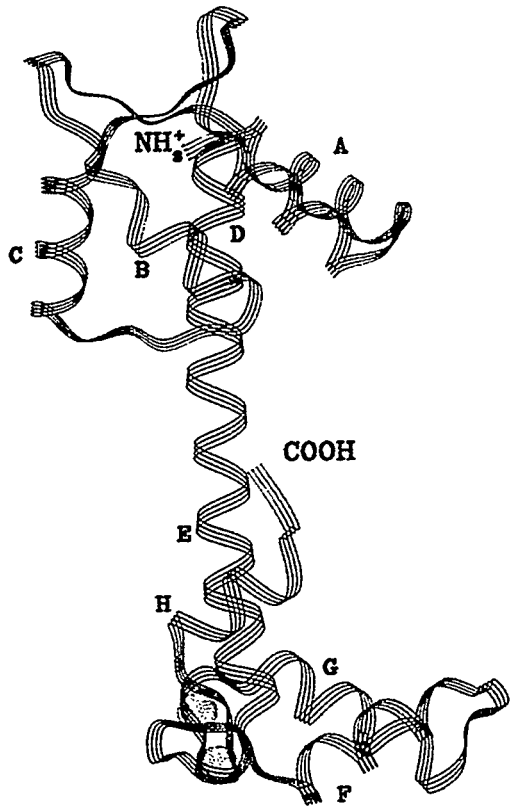


Fig. 3. Backbone trace of the crystal structure of $2Ca^{2+}$•TnC (Herzberg and James, 1985). The C-terminal domain is oriented to show the "cup-shape" of that domain. The pairs of helices (E/H, and F/G) forming the sides of the cup representing the C-terminal domain are labeled and the Ca^{2+} ions (yellow) are in the loop regions that form the base of that cup.

We completed neutron and x-ray scattering experiments on the $4Ca^{2+}$•TnC•TnI complex in solution (Olah *et al.*, 1994). For the neutron experiments, the TnC component was deuterated. Deuterated TnC was produced using an *E. coli* expression system grown on deuterated algal hydrolysate and D_2O. The average deuteration level in the TnC was 78% and was chosen so that neutron scattering data could be collected on either side of the solvent match point for the overall complex (62% D_2O) for optimal data analysis and interpretation. Neutron scattering data were measured for $4Ca^{2+}$•$TnC_{(78\%\ deuterated)}$•TnI in 0, 20, 40, 90, and 100% D_2O. X-ray scattering data were also collected for the complex, effectively giving another contrast point in the series for which the two components have equal mean scattering densities. Figure 4 shows the Stuhrmann plot

(Eq. 5) of the R_g data from the contrast series data. The data are well fit with a straight line (β zero) with a negative slope (α negative). The conclusion from this analysis is that the two components of the complex have centers-of-mass that are approximately coincident and the component with the higher scattering density—the deuterated TnC—is more toward the inside of the complex than the TnI.

A multiple linear regression routine was used to extract the basic scattering functions from the contrast series data using Eq. (2). Thus the scattering profiles (I(Q) *vs* Q) were obtained for the individual components of the complex. The scattering profile for the overall complex is given directly by the x-ray scattering measurements. Figure 5 shows the corresponding P(r) functions for the complex and its two components; Table II gives

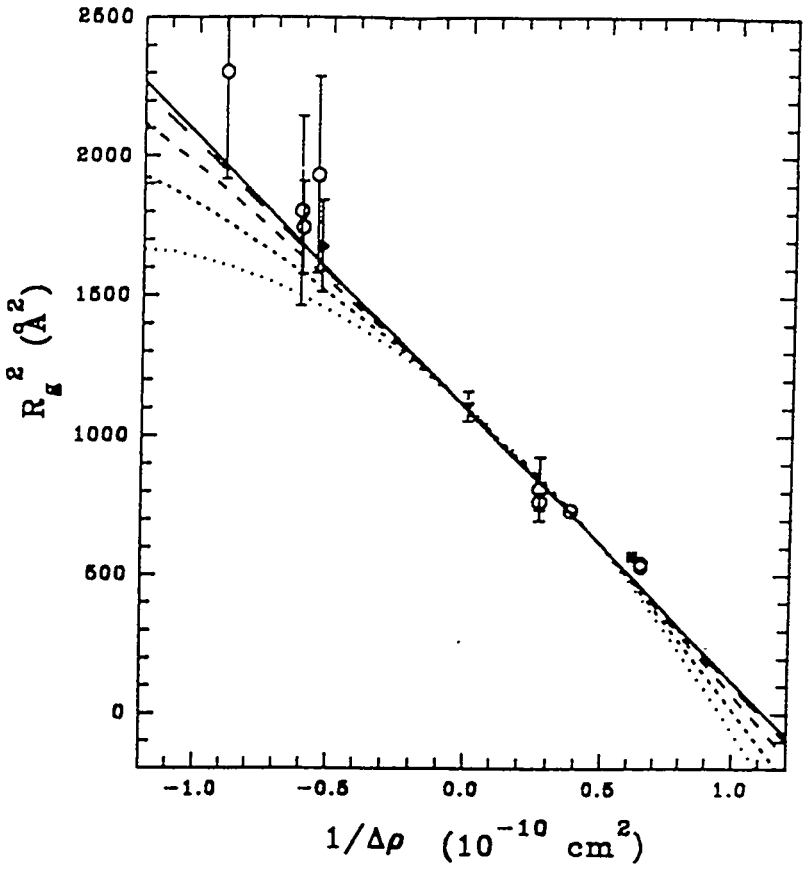


Fig. 4. Sturhmann plot (Eq. 5) of the R_g data from the contrast series on $4Ca^{2+}$•TnC•TnI. The line fits were determined from Eq. (5), assuming R_g values of 23.9 Å and 41 Å for $4Ca^{2+}$•TnC and TnI, respectively, and a separation distance for their centers-of-mass of 0 Å (—), 5 Å (— —), 10 Å (– – –), 15 Å (- - -), and 20 Å (....). neutron data (●), x-ray data (▼), TnC (■), TnI (◆).

the derived structural parameters. Each component, as well as the complex, has an asymmetric P(r) indicating an elongated structure. Guinier analysis (Eq. 6) gave R_c values for each component (Table II). A simple analysis of these R_c values shows the long axes of the complex, and its two components must be approximately coincident (Olah *et al.*, 1994).

The P(r) function for $4Ca^{2+}$•TnC is very similar to that predicted by the crystal structure. There is a small difference (20%) in the ratio of the two peaks at 18 and 45 Å, but this was mostly accounted for by the expected conformational change induced by Ca^{2+} binding to the N-terminal domain. The crystal structure shows no Ca^{2+} in this domain, whereas our neutron experiments were done in the presence of saturating Ca^{2+} concentrations. A comparison of the structurally similar N- and C-terminal domains of TnC suggests the Ca^{2+}-induced conformational change involves an opening of the cup-shaped domain to expose hydrophobic residues lining the inner surface of the cup (Herzberg *et al.*, 1986). The cup shapes are defined by two pairs of helices that form the sides of the cup with the Ca^{2+}-binding sites in the base of the cup (Fig. 3). We modified the crystal structure of TnC to open the N-terminal domain similarly to the Ca^{2+}-bound C-domain and optimized the fit of the P(r) function calculated from this modified structure with that calculated using the basic scattering function for $4Ca^{2+}$•TnC (Fig. 5) (Olah and Trewhella, 1994). The neutron data show that $4Ca^{2+}$•TnC is

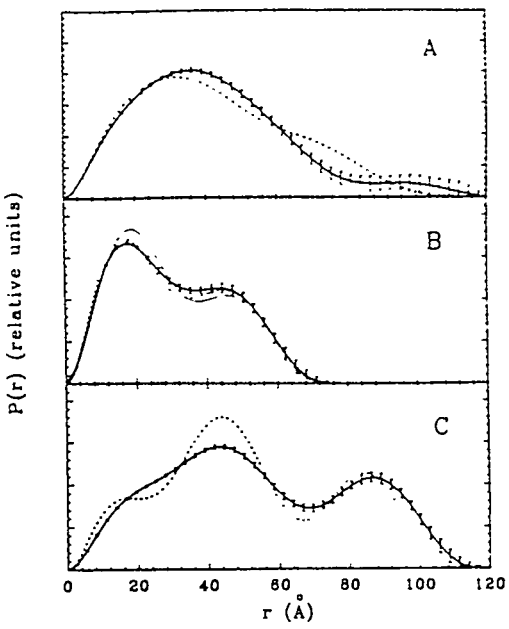


Fig. 5. P(r) functions calculated from the scattering profiles for (A) $4Ca^{2+}$•TnC•TnI, (B) $4Ca^{2+}$•TnC, and (C) TnI. The dashed lines show the P(r) profiles calculated from the model in Fig. 6. The thin solid line in the middle panel shows the P(r) calculated from the crystal structure of $2Ca^{2+}$•TnC.

fully extended in the complex with TnI. The P(r) function for TnI is even more extended than that for $4Ca^{2+}$•TnC, and its maximum dimension is the same as for the overall complex. The three peaks in the P(r) indicate TnI has domains, and/or repeating regular structural features.

The combined constraints of the basic structural parameters determined from the scattering experiments provide a basis to model the complex and its components.

Table II. Structural Parameters Derived from the Neutron and X-Ray Scattering Data and from the Model for $4Ca^{2+}$•TnC•TnI

Component		R_g (Å)	R_c (Å)	d_{max} (Å)
$4Ca^{2+}$•TnC	experiment	23.9 ± 0.5	10.7 ± 1.0	72 ± 2
	model	24.1	10.3	73
TnI	experiment	41.2 ± 2.0	20.5 ± 2.0	118 ± 4
	model	40.1	20.5	114
$4Ca^{2+}$•TnC•TnI	experiment	33.0 ± 0.5	16.2 ± 1.5	115 ± 4
	model	33.4	15.7	117

Reduced χ^2 Values for Refined Model Against Scattering Data

$4Ca^{2+}$•TnC	0.70	(Q-range 0.02 – 0.2 $Å^{-1}$)
TnI	0.92	(Q-range 0.02 – 0.15 $Å^{-1}$)
$4Ca^{2+}$•TnC•TnI	0.95	(Q-range 0.02 – 0.2 $Å^{-1}$)

We know TnC in the complex is very similar to the TnC crystal structure—both components and the complex are elongated structures; their long axes and their centers-of-mass are approximately coincident. TnI is more toward the outside of the complex than TnC, and TnI has the same maximum dimension as the overall complex. We also know the approximate molecular volumes for each component from their amino acid compositions and partial specific volumes. The volume constraint on TnI excludes models that have TnI uniformly wrapping TnC. We therefore developed a Monte Carlo modeling routine and tested possible ways TnI could wrap around TnC and satisfy all the known constraints from the scattering data (Olah and Trewhella, 1994). The result of this model search is the model shown in Fig. 6; Fig. 5 and Table II show the excellent agreement between this model and the scattering data.

The derived structure shows TnI forms a helical "slinky"-type structure around $4Ca^{2+}$•TnC. The diameter of the TnI central spiral is 12 Å—close to that expected for an α-helix—and it passes through the two hydrophobic cups on each globular domain of $4Ca^{2+}$•TnC. The $4Ca^{2+}$•TnC has a high affinity for peptides with sequences that have a propensity for forming basic amphipathic helices. Small-angle x-ray scattering

Fig. 6. The model for $4Ca^{2+}$•TnC•TnI determined from the neutron and x-ray scattering contrast series data. The TnC is shown as a space-filling model of the modified crystal structure with the N-terminal domain on top oriented to show TnI spiraling through its hydrophobic cup. The TnI is represented by random points distributed uniformly through the volume it occupies.

measurements (Blechner *et al.*, 1992) on $4Ca^{2+}$•TnC bound to short, helical peptides from this class show that TnC will sometimes—but not always—collapse around these structures such that the two globular domains come into close contact. Based on comparisons with the structurally similar calmodulin (see below), these

basic amphipathic helices are believed to form strong interactions with the hydrophobic cups of TnC's globular domains. A simple Chou-Fasman analysis (Chou and Fasman, 1978) of rabbit skeletal muscle TnI sequence predicts a very high proportion of α-helix: as much as 71% of the total sequence. Because the centers of mass of TnC and TnI are coincident, the residues forming the central spiral would correspond approximately to residues 55 through 125, assuming it was mostly α-helical. In rabbit skeletal TnI, this sequence contains two regions that have amino acid sequences ideal for forming amphipathic helical structures and are perfectly spaced with respect to each other to bind to the hydrophobic cup regions of $4Ca^{2+}$•TnC. Inspection of the structure shows that the TnI spiral interacts with the C, E, and G helices of TnC, which is expected based on other studies (reviewed in Zot and Potter, 1987). Finally, the TnI inhibitory region (TnI(96-115)) is positioned in our model to interact with helix E (the C-terminal part of the interconnecting helix region) of TnC, as predicted from proteolytic cleavage studies (Gabarek *et al.*, 1981). This region of TnI has alternate binding sites on actin and TnC and is the minimal TnI sequence segment that can inhibit actomyosin ATPase activity (Syska *et al.*, 1976; Talbot and Hodges, 1981; Van Eyk and Hodges, 1988).

This solution structure of $4Ca^{2+}$•TnC•TnI allows one to suggest a possible molecular basis for the Ca^{2+}-dependent regulation of muscle contraction. In the Ca^{2+}-bound complex, both ends of the TnI central spiral region are anchored by interactions with the hydrophobic cup regions of TnC; the TnI inhibitory sequence is constrained to be associated with the TnC; and the inhibitory function is switched off. Loss of Ca^{2+} from the N-terminal low-affinity Ca^{2+}-specific binding sites results in a closing of that domain, lowering its affinity for TnI at the N-terminal end, and allowing the TnI inhibitory sequence (96 to 115) the flexibility to shift from its binding site on TnC to that on actin—switching the inhibition on. The regulatory signal is thus transmitted *via* the central spiral region of TnI, which contains the inhibitory sequence.

Calmodulin and its Target Enzymes

Although calmodulin has structural and functional similarities to those of TnC, it is distinct in that it is a multifunctional Ca^{2+}-binding protein that modulates the activity of a diverse array of enzymes in a Ca^{2+}-dependent manner (Fig. 7). Calmodulin is typically only associated with its target enzyme in the presence of Ca^{2+}, and the calmodulin binding domains in various target enzymes are usually confined to a 17- to 20-residue sequence segment that has a high propensity for forming an amphipathic helix (O'Neil and DeGrado, 1989). There are a number of peptides that have no known functional relevance to calmodulin but share this structural motif and are known to form high-affinity, 1:1 complexes with calmodulin with similar structures (reviewed in Trewhella, 1992). Whereas

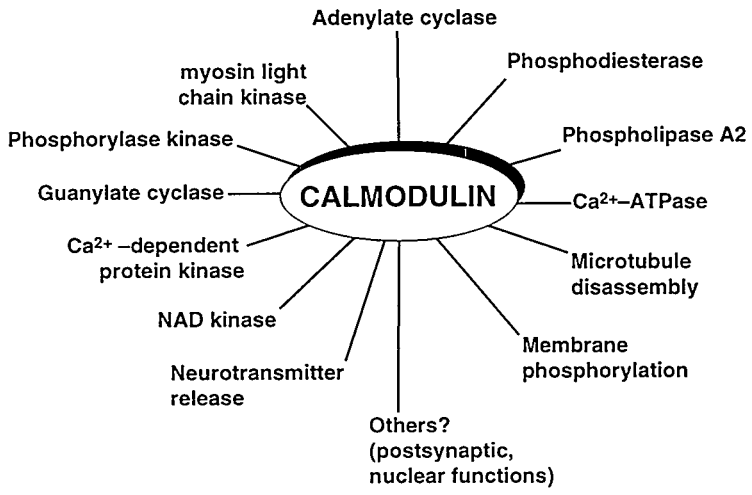


Fig. 7. Ca^{2+}•Calmodulin-dependent enzymes.

these peptides and binding domains share a common structural motif, they show considerable variation in the type and distribution of hydrophobic and charged residues, reflecting the functional diversity of calmodulin.

Early neutron scattering studies of calmodulin complexed with its binding domain from myosin light chain kinase (Heidorn *et al.*, 1989) show calmodulin forms a highly contracted structure surrounding that target peptide (Fig. 8). The contraction is achieved by means of flexibility in the interconnecting helix region of the molecule that links its two globular domains. More recently, multi-dimensional NMR and crystallographic data have revealed the details of these types of structures (Ikura *et al.*, 1992; Meador *et al.*, 1992, 1993)—in particular, the precise nature of the hydrophobic and charge interactions that give rise to the tight binding of the amphipathic helical target peptides. The target peptides form a bifurcated amphipathic helix structure in which the N- and C-terminal ends present one, two, or three large hydrophobic residues on opposite sides (Fig. 9). Calmodulin collapses around the peptide so that each end is "grasped" by one of the hydrophobic cup-shaped domains. Flexibility in the interconnecting helix region allows calmodulin to optimize its binding to different arrangements of hydrophobic and charged residues that are important in forming the complexes.

In contrast to the highly collapsed structures described above, calmodulin remains extended in its interaction with the catalytic subunit of phosphorylase kinase. Like TnC in troponin, calmodulin is an integral part of phosphorylase kinase and remains so in both the presence and absence of Ca^{2+}. Phosphorylase kinase is

a multisubunit enzyme that is made up of a tetramer of tetramers (reviewed by Pickett-Gies and Walsh, 1986). It has one catalytic subunit (γ) and three regulatory subunits (α,β,δ). The δ-subunit is calmodulin, and it regulates the activity of the γ-subunit in a Ca^{2+}-dependent manner. The calmodulin-binding region in the γ-subunit is distinct from other calmodulin-binding domains in that it is made up of two noncontiguous subdomains that bind to calmodulin simultaneously (Dasgupta *et al.*, 1989). Dasgupta *et al.*, further identified sequence similarities between these calmodulin-binding subdomains and TnI. The regions of sequence similarity include a 4 residue region of sequence identity that overlaps with the TnI inhibitory peptide sequence. Earlier small-angle x-ray and neutron scattering data on calmodulin complexed with the two calmodulin-binding subdomains identified in the γ-subunit of phosphorylase kinase showed calmodulin to be

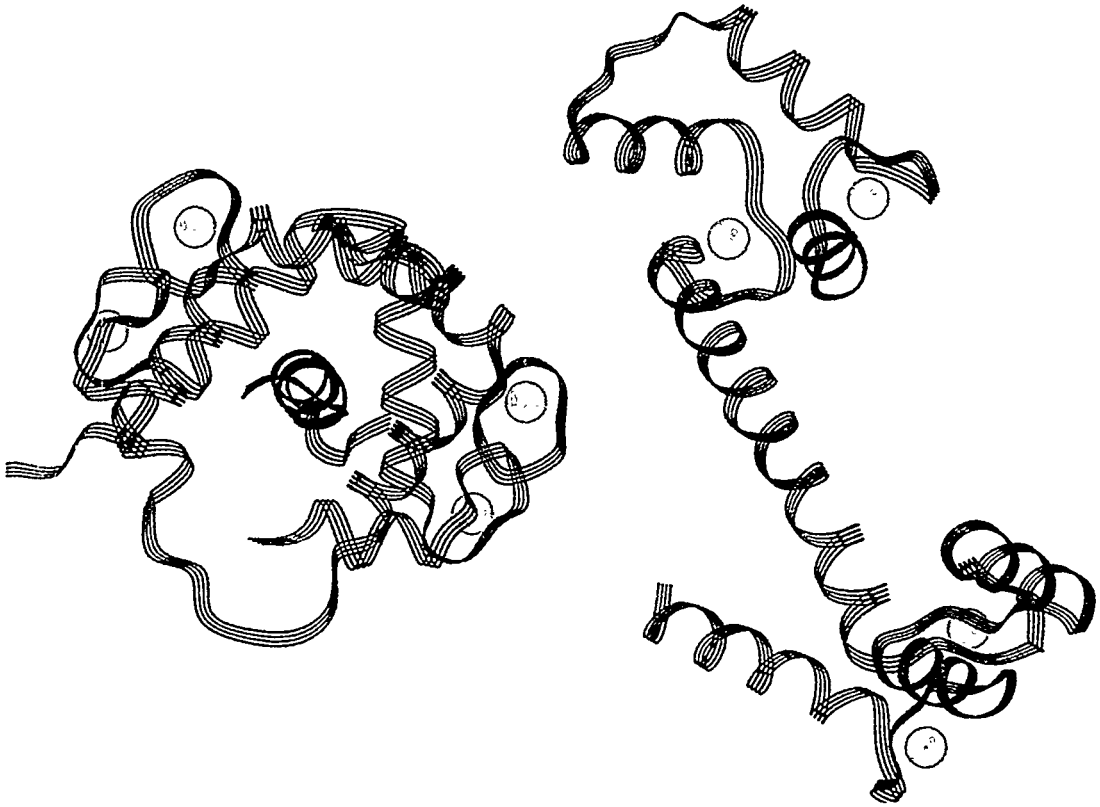
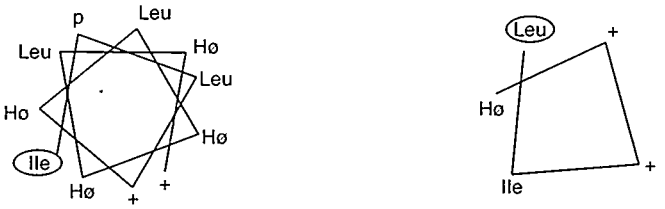


Fig. 8. Backbone trace of calmodulin [from the crystal structure coordinates(Babu* et al., *1988) on left] and of its complex with its binding peptide in myosin light chain kinase [from the NMR structure (Ikura* et al., *1993) on right]. Calmodulin is blue, Ca^{2+} is yellow, and the peptide is red.

Mastoparan



Mellitin

Calcineurin

skMLCK

smMLCK

Ca^{2+} Pump

extended (Trewhella *et al.*, 1990). Small-angle x-ray scattering experiments (Blechner *et al.*, 1992) on TnC complexed with a number of peptides showed that TnC also is extended in its interaction with the TnI inhibitory peptide sequence, as well as with another TnC-binding sequence (TnI, 1 through 30) that is thought to serve a regulatory role by acting as a negative regulator of the TnI inhibitory region (Ngai and Hodges, 1992). These observations led us to speculate (Trewhella *et al.*, 1990; Blechner *et al.*,, 1992) that the interaction of TnC with TnI may be similar to the interaction between calmodulin and the γ-subunit of phosphorylase kinase and, further, that the extended TnC and calmodulin structures may be important in maintaining Ca^{2+}-independent associations in specific systems. The $4Ca^{2+}$•TnC•TnI structure presented above supports this speculation and may provide a good model for the calmodulin/phosphorylase kinase interaction.

Conclusions

Neutron scattering experiments on complexes of calmodulin and TnC and their various targets have revealed important information about the function of these regulatory complexes. The ability of calmodulin to modulate its conformation, through flexibility in the interconnecting helix region to accommodate different target binding domains, is a remarkable example of nature building functional diversity as well as specificity into a compact, unusual shape. Nature uses this dumbbell-shape in other systems; for example, the essential and regulatory light chains of myosin show structural similarity to calmodulin and TnC and were recently solved crystallographically as part of the myosin S1 fragment (Rayment *et al.*,, 1994) and the regulatory domain of scallop myosin (Xie *et al.*, 1994). In both of these complexes, the essential and the regulatory light chains bind to long helical segments of the myosin heavy chain and show different degrees of collapse about their binding domains, intermediate between the highly collapsed calmodulin complexes and the fully extended TnC in its complex. The similarities between TnC in troponin and calmodulin in phosphorylase kinase lead to some interesting predictions that can now be tested with further scattering experiments as well as NMR and crystallographic studies. Stable isotope labeling for neutron and multidimensional NMR

Fig. 9. Helical wheel representations of a number of calmodulin-binding peptides. The ellipse-circled Trp residues in skeletal muscle MLCK, smooth muscle MLCK, and Ca^{2+} pump calmodulin-binding domains was used to align the sequences, which have also been divided into two regions based on the $4Ca^{2+}$•calmodulin•skMLCK NMR structure (Ikura et al., 1992): the one on the left, which interacts primarily with the carboxyl domain of calmodulin, and the one on the right, which interacts primarily with the N-terminal domain. p = polar residue, Hø= hydrophobic residues, +/- = positively/negatively charged residue.

experiments will continue to be crucial to our complete understanding of the molecular basis for these key regulatory processes.

Acknowledgements

The author is indebted to the many colleagues who contributed to the work reviewed in this article; in particular, the members and former members of our research group: Douglas B. Heidom, Glenn A. Olah, Steven L. Blechner, and Sue E. Rokop. I would also like to acknowledge the contributions of our collaborator Donald K. Blumenthal, who has been a constant source of ideas throughout this work. The work reported in this review was supported by the Department of Energy's Office of Health and Environmental Science Project KP-04-01-00-0 and NIH Grant GM40528. The work benefited from use of facilities at the Manuel Lujan, Jr., Neutron Scattering Center, a national user facility funded by the Department of Energy Office of Basic Energy Sciences. This work also benefited from the support of the National Science Foundation under Agreement DMR-9122444 and the National Institute of Standards Technology, US Department of Commerce, who provided neutron scattering facilities.

References

Babu, Y.S., Bugg, C.E., and Cook, W.J. (1988) *J. Mol. Biol. 204*, 191.

Blechner, S.L., Olah, G.A., Strynadka, N.C.J., Hodges, R.S., and Trewhella, J. (1992) *Biochemistry 31*, 11326.

Chou, P.Y. and Fasman, G.D. (1978) *Adv. Enzymes 47*.

Dasgupta, M., Honeycutt, T., and Blumenthal, D.K. (1989) *J. Biol. Chem. 265*, 17156.

Gabarek, Z., Drabikowski, W., Leavis, P.C., Rosenfeld, S.S., and Gergely, J. (1981) *J. Biol. Chem.. 256*, 13121.

Glatter, O. (1982) in *Small Angle X-Ray Scattering, Chap. 4*, O. Glatter and O. Kratky, eds., Academic Press, New York.

Guinier, A. (1939) *Ann. Phys. (Paris) 12*, 161.

Heidorn, D.B., Seeger, P.A., Rokop, S.E., Blumenthal, D.K., Means, A.R., Crespi, H., and Trewhella, J. (1989) *Biochemistry 28*, 6757.

Heidorn, D.B. and Trewhella, J. (1990) *Comments Mol. Cell. Biophys. 6*, 329.

Herzberg, O. and James, M.N.G. (1985) *Nature (London) 313*, 653.

Herzberg, O., Moult, J., and James, M.N.G. (1986) *J. Biol. Chem. 261*, 2638.

Ibel, K. and Stuhrmann, H.B. (1975) *J. Mol. Biol. 93*, 255.

Ikura, M., Clore, G.M, Gronenborn, A.M., Zhu, G., Klee, C.B., and Bax A. (1992) *Science 256*, 632.

Kretsinger, R.H. (1980) *CRC Crit. Rev. Biochemistry 8*, 119.

Leavis, P.C. and Gergely, J. (1984) *CRC Crit. Rev. Biochem. 16*, 235.

Meador, W.E., Means, A.R., and Quiocho, F.A. (1992) *Science 257*, 1251.

Meador, W.E., Means, A.R., and Quiocho, F.A. (1993) *Science 262*, 1718.

Moore, P.B. (1980) *J. Appl. Crystallogr. 13*, 168.

Moore, P.B. (1981) *J. Appl. Crystallogr. 14*, 237.

Ngai, S.-M. and Hodges, R.S. (1992) *J. Biol. Chem. 267*, 15715.

Olah, G.A. and Trewhella, J. (1994) *Biochemistry 33.*, 12800.

Olah, G.A., Rokop, S.E., Wang, C -L.A., Blechner S.L., and Trewhella, J. (1994) *Biochemistry 33*, 8233.

O'Neil, K.T. and DeGrado, W.F. (1989) *Prot. Struct. Func. Genet. 6*, 284.

Picket-Gies, C.A. and Walsh, D.A. (1986) *Enzymes (3rd Ed.) 264*, 396.

Rayment, I., Rypniewski, W.R., Schmidt-Bäse, F., Smith, R., Tomchick, D.R., Benning, M.M., Winkelmann, D.A., Wesenberg, G., and Holden, H.M. (1994) *Science 261*, 50.

Sundaralingam, M., Bergstrom, R., Strasburg, G., Rao, S.T., Roychowdhury, Greaser, M., and Wang, B.C. (1985) *Science 227*, 945.

Syska, H., Wilkinson, J.M., Grand, J.A., and Perry, S.V. (1976) *Biochem. J. 153*, 375.

Talbot, J.A. and Hodges, R.S. (1981) *J. Biol. Chem. 256*, 2798.

Trewhella, J. (1992) *Cell Calcium 13*, 377.

Trewhella, J., Blumenthal, D.K., Rokop, S.E., and Seeger P.A. (1990) *Biochemistry 29*, 9316.

Van Eyk, J.E. and Hodges, R.S. (1988) *J. Biol. Chem. 263*, 1726.

Xie, X., Harrison, D.H., Schlichting, I., Sweet, R.M., Kalabokis, V.N., Szeut-Györgyl, A.G., and Cohen, C. (1994) *Nature 368*, 306.

Zot, A.S. and Potter, J.D. (1987) *Annu. Rev. Biophys. Biophys. Chem. 16*, 535.

Stereoselective Synthesis of Stable-Isotope-Labeled Amino Acids

Clifford J. Unkefer,[1] Siegfried N. Lodwig,[2]
Rodolfo A. Martinez,[1] and Louis A. Silks, III[1]

[1]National Stable Isotope Resource
Chemical Science and Technology
Los Alamos National Laboratory
Los Alamos, NM 87545

[2] Science Division
Centralia College
Centralia, WA 98531

As a result of technological advances, magnetic resonance and vibrational spectroscopies have emerged as major players in biochemical structural and mechanistic studies. For example, time-resolved vibrational methods are used to probe the dynamics of fast biochemical reactions such as proton translocation and electron transport. Stable isotope labels are required for spectral assignment. In addition, stable-isotope-labeled amino acids are used in combination with NMR spectroscopy to study the structure and function of enzymes. Historically, proteins labeled with [^{13}C]- or [^{15}N]-amino acids were examined by direct detection NMR methods (London, 1980, 1984). Emphasis was on labeling only those amino acyl residues involved in binding or catalysis. Interest in labeling of macromolecules has been increased by the recent development of multiple quantum NMR methods (which allow indirect detection of a ^{13}C or ^{15}N nucleus based on its coupling interaction with the directly bonded protons (Griffey and Redfield, 1987)) and of three- and four-dimensional NMR methods that allow correlation of proton, carbon, and nitrogen resonances (Oschkinat *et al.*, 1988; Kay *et al.*, 1990, respectively). These NMR techniques require uniform labeling and have allowed structural biologists to solve complete three-dimensional solution structures of proteins. NMR and vibrational methods are critically dependent on our ability to introduce both site-

specific and uniform stable isotope labels into biomacromolecules in general and in proteins in particular. This manuscript describes stereospecific strategies for the synthesis of amino acids labeled with ^{13}C, ^{15}N, and ^{2}H. Although the methodologies are discussed in terms of the synthesis of specific labeling patterns, they can be extended to the synthesis of virtually any isotopomer of the amino acids.

Labeling Requirements for NMR

The size of biomolecules that can be studied by modern NMR methods is currently limited. Two factors contribute to this size limitation. First, larger molecules tumble slowly, which effectively decreases the spin-spin lattice relaxation rate (T_2) and increases the resonance linewidth. Rapid T_2 relaxation can greatly lower the sensitivity of multiple-pulse NMR experiments. In addition, the increased linewidth—coupled with the fact that the absolute number of protons increases in direct proportion to molecular weight—results in severe spectral overlap problems for most globular proteins greater than 15 KDa. If NMR techniques are to be extended to the study of larger and larger biomolecules, it is critical to develop deuterium labeling strategies that will minimize the ^{1}H dipolar contribution to T_2 relaxation of protons and carbons—effectively increasing the sensitivity and the resolution at a given field.

Another approach to overcoming the spectral overlap problem at higher molecular weights involves three- and four-dimensional heteronuclear *J* correlation NMR techniques that spread out proton resonances on a ^{13}C or ^{15}N chemical shift axis, which effectively increases the resolution at any given field (Oschkinat *et al.*, 1988; Kay *et al.*, 1990). This approach requires the production of protein samples labeled uniformly with ^{15}N and ^{13}C and has been used to solve the NMR solution structure of, for example, the biomolecular complex of calmodulin and its binding domain from myosin light chain kinase (21 kDa) (Ikura *et al.*, 1991). In addition to increasing the effective resolution of the proton resonance signals, uniform labeling of proteins with ^{13}C and ^{15}N introduces a spin 1/2-spin 1/2 coupling network along the peptide backbone that links the N-terminal ^{15}N-amino group to the^{13}C-carboxyl terminus. By providing a physical basis for making the sequential chemical shift assignments, this ^{15}N and ^{13}C-coupling network solves this difficult and time consuming task. Any labeling strategy aimed at obtaining structural information should retain this feature of uniform labeling by incorporating labels in the protein backbone ($^{13}C\alpha$, *carboxyl*-^{13}C, and $^{15}N\alpha$).

Effects of Deuterium Labeling on NMR Spectra

Spectral Simplification

Consider the β-methylene protons found in 15 of the 20 common amino acids (Fig. 1). These protons are diastereotopic, and therefore chemically inequivalent, which gives rise to a tightly coupled AB multiplet in their proton NMR spectrum. Thus, scalar coupled resonances are observed for each of the β-methylene

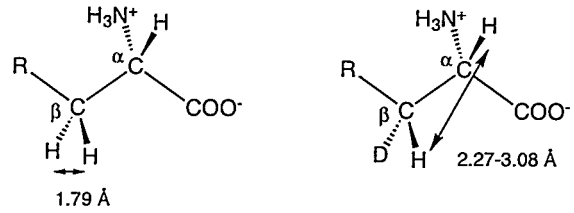


Fig. 1. An L-α-amino acid and its (3R)-[3-^{2}H] isotopomer.

protons exacerbating spectral crowding in this region. This problem can be solved by stereoselectively replacing the pro*S* (or pro*R*) proton with a deuteron, which eliminates half of the β-methylene proton resonances while retaining the coupling network required to obtain structural information. In addition, chiral substitution of the prochiral methylene protons would provide an unequivocal method for making stereochemical assignments to these proton resonances. Kushland and LeMaster demonstrated this approach to spectral simplification and chiral assignment by labeling thioredoxin with [2-^2H_R, 2-^{13}C]glycine (Kushland and LeMaster, 1993; LeMaster and Richards, 1988). In the chiral environment of the protein, the methylene protons on glycine give rise to separate resonances that are scalar coupled. Stereo-selective deuteration of glycine simplified the ^{1}H spectrum by removing one of the resonances as well as both the geminal scalar and dipolar coupling. Under conditions of deuterium decoupling, the ^{1}H resonance was significantly sharper, resulting in an increase in the signal-to-noise ratio.

Proton-Proton Relaxation

Protons are relaxed primarily by dipole-dipole interactions with other protons, which, at the long correlation times associated with large proteins, can yield broad resonances. Dipolar relaxation is strongly dependent on the internuclear distance ($1/r^6$). Because the methylene protons (Fig. 1) are close in space (1.8 Å), they are relaxed primarily by their dipolar interaction. This relaxation mechanism is removed by replacing one of the β-methylene protons with deuterium. The remaining proton will be relaxed by protons that are more distant (2.3 to 3.1 Å), which significantly decreases the relaxation rate and sharpens the resonance. The general utility of using deuterium to minimize proton dipolar broadening effects has been demonstrated by random fractional deuteration of proteins (LeMaster and Richards, 1988). By culturing bacteria in deuterated medium, proteins are randomly deuterated. This random deuteration dilutes the protons, thus increasing the effective proton-proton distances and minimizing the dipolar effect in the proton resonance linewidth. In addition to diluting the unwanted scalar coupling and dipolar interactions, the desired interactions with more distant protons are also diluted, lowering the sensitivity of these experiments. For deuteration to be fully effective, synthetic strategies for stereospecific deuterium labeling of amino acids must be developed.

Carbon Relaxation

The dipolar interaction of directly bonded protons dominates the T_2 relaxation of ^{13}C in protonated carbons. As Bax and coworkers have discussed (Grzesiek *et al.*, 1993), T_2 relaxation of protonated carbons during the long delays required in the heteronuclear *J* correlation NMR pulse sequences limits the sensitivity of these experiments in larger proteins. Again, because of its low γ, deuterium substitution significantly increases the T_2 of directly bonded carbons. The utility of this approach was tested by producing calcineurin B (19.7 kDa) samples that were uniformly enriched with ^{2}H, ^{15}N, and ^{13}C. Protons were then back-exchanged into the amide positions. This sample was backbone-labeled with ^{15}N and ^{13}C, protonated at the amides, and deuterated at αH. The increased T_2 of $C\alpha$ allowed magnetization to be relayed from one amide proton through the α-carbon to the sequentially adjacent amide proton. Chemical synthesis of amino acids with specific ^{15}N, ^{13}C, and ^{2}H labeling makes it possible to apply this approach to other spin systems in large proteins.

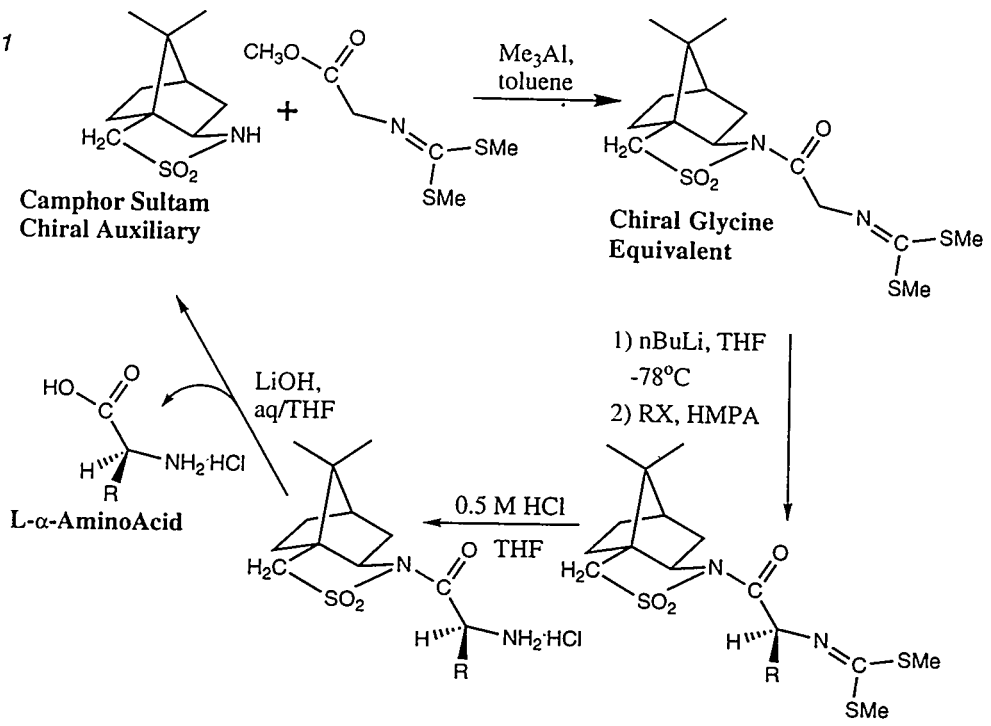
Synthesis of Labeled Amino Acids by Using the Camphor Sultam Chiral Auxiliary

Recently, Oppolzer and coworkers (Oppolzer et al., 1989; Oppolzer and Tamura, 1990) developed strategies for the synthesis of enantiomerically pure amino acids that involve the camphor sultam chiral auxiliary. As described below, this chemistry has potential for providing routes to virtually any isotopomer of all of the amino acids. The remarkable stereochemical selectivity of the camphor sultam auxiliary can be used to generate the proper stereochemistry at the α-carbon or to generate stereochemical centers by delivering isotopes specifically to prochiral sites. For example, the diastereotopic methyl groups in valine can be differentiated with isotopes by using this approach. In addition, the camphor sultam auxiliary has the potential for providing a method for stereospecific deuteration of the β-methylene group in many of the amino acids.

Alkylation of Chiral Glycine

Oppolzer first described a camphor-based chiral glycine equivalent that, as diagrammed in Scheme 1, is useful for stereoselective synthesis of the common α-amino acids (Oppolzer *et al.*, 1989). Oppolzer's chiral auxiliary contains a sultam ring fused to the camphor nucleus; N-protected glycine (N-[bis(methylthio)-methylene]-glycine) is linked as an amide to the nitrogen in the sultam ring. This chiral glycine equivalent is deprotonated by treatment with n-butyl lithium in THF at –78°C. Decomposition of the enolate with electrophiles is carried out in the presence of hexamethylphosphoramide. The N-blocking group is removed by acid hydrolysis. After treatment with dilute lithium hydroxide, the product amino acid and the camphorsultam auxiliary are separated and recovered.

Scheme 1

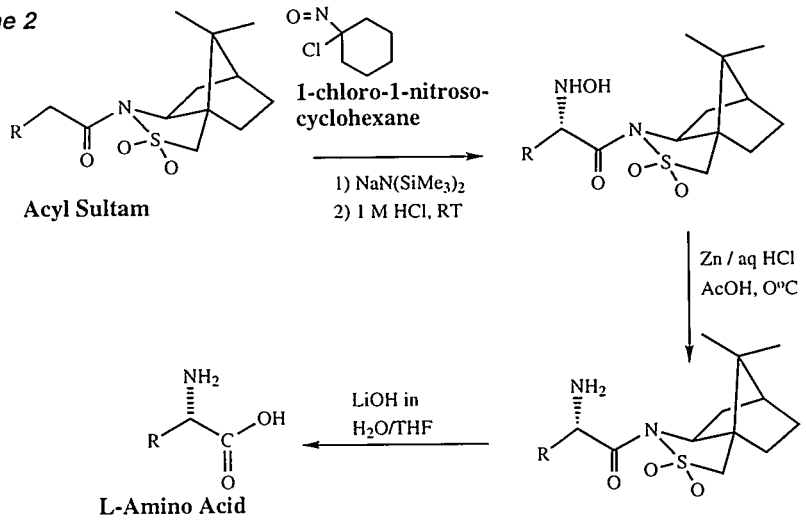


This process occurs with remarkable stereoselectivity. Starting with the (2*S*)-camphor sultam glycinate and using a series of alkyl iodides as electrophiles, we have prepared L-[3-^{13}C]alanine, L-[*methyl*-$^{2}H_6$]valine, L-leucine, L-[3-^{13}C]-phenylalanine, L-proline, L-lysine, L-aspartic acid, L-β-cyanoalanine, and L-ornithine. The alkylation reaction was uniformly efficient and had remarkable enantioselectivity (d.e. >99%). Efficient procedures for deblocking the product amino acid were carried out without racemization of the product. This method has the potential for producing all of the common amino acids. By starting with [1,2-$^{13}C_2$, ^{15}N]glycine, this route can be used to produce backbone labeled amino acids.

Electrophilic Amination of Acyl Sultams

Oppolzer and Tamura (1990) developed an "electrophilic amination" approach that also incorporates the camphorsultam chiral auxiliary (Scheme 2). Deprotonation of the N-acyl camphorsultam is effected by treatment with sodium hexamethyldisilazide. The enolate is treated with 1-chloro-1-nitrosocyclohexane followed by HCl to quench the reaction. The isolated sultam hydroxylamino acid is reduced by treatment with Zn to yield the sultam-linked amino acid. The sultam is removed by saponification with dilute LiOH to yield the free amino acid. This scheme produces L-amino acids with enantiomer ratios of 200 to 1. As outlined below, we developed a synthesis of ^{15}N-labeled

Scheme 2

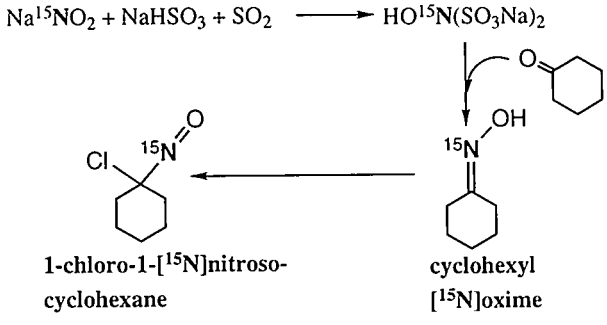


1-chloro-1-nitrosocyclohexane to make this a convenient method for the synthesis of L-[α-^{15}N]amino acids.

Preparation of 1-chloro-1-[^{15}N]-nitrosocyclohexane

Scheme 3 shows the preparation of 1-chloro-1-[^{15}N]nitrosocyclohexane. We prepared cyclohexanone [^{15}N]oxime using a modification of a procedure described by Eck and Marvel (1943). Potassium [^{15}N]nitrite was converted into the intermediate hydroxylamine disulfonate by treatment with sodium bisulfite and sulfur dioxide at 0°C. Addition of an excess of cylohexanone followed by heating at 50°C for 1 hr gave a mixture of the cyclohexanone [^{15}N]oxime and unreacted cyclohexanone. Cyclohexanone was removed *in vacuo* to yield pure [^{15}N]oxime in 60% yield, based on potassium [^{15}N]nitrite. The [^{15}N]oxime was then converted to 1-chloro-1-[^{15}N]nitrosocyclohexane by treatment with chlorine gas (Muller *et al.*, 1954) in quantitative yield.

Scheme 3

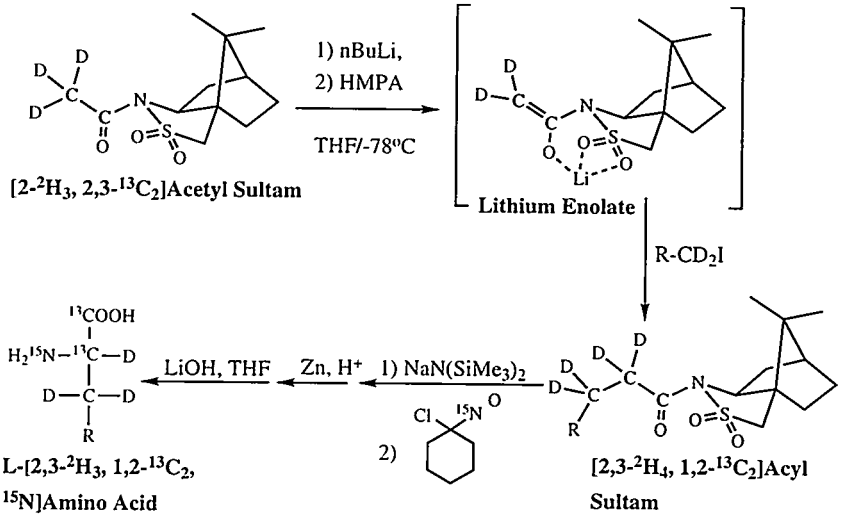


Treatment of the acyl sultams corresponding to L-alanine, L-valine, L-[1-^{13}C]valine, L-leucine, and L-phenylalanine with 1-chloro-1-[^{15}N]nitrosocyclohexane gave the related L-hydroxylamino acyl sultams in good yields. Zinc reduction of the hydroxylamino acyl sultams yielded the sultam-linked amino acids. NMR analysis of the crude compounds showed that, in these two steps, the ratio of the L-enantiomer to the D-isomer was between 100:1 and 200:1. Deblocking affords the L-[α-^{15}N]amino acids in enantiomeric excess of greater than 99%.

Synthesis of Backbone-Labeled Amino Acids

As discussed above, backbone labeling of amino acids at the α-carboxyl and Cα with ^{13}C and the α-amino group with ^{15}N is required for making sequential assignments in proteins. This labeling pattern could be derived via synthesis of [^{15}N, 1,2-$^{13}C_2$]glycine and incorporation into the sultam-based chiral glycine equivalent described above. The overall conversion of [1,2-$^{13}C_2$]acetate to [^{15}N, 1,2-$^{13}C_2$]-glycine and its incorporation into the sultam-derived N-[bis(methylthio)-methylene]-[^{15}N, 1,2-$^{13}C_2$]glycinate requires eight steps. Subsequent alkylation of the sultam glycinate and deblocking to the labeled L-α-amino acid requires three steps; therefore, the overall synthesis of backbone-labeled amino acids via [^{15}N, 1,2-$^{13}C_2$]glycine requires 11 steps. Many of the steps involve blocking and deblocking the amino group on glycine. We have developed a strategy for the synthesis of backbone-labeled amino acids that eliminates the necessity for blocking the amino group and shortens the overall synthesis of backbone-labeled

Scheme 4



amino acids to five steps. As diagrammed in Scheme 4, this route involves the alkylation of the [1,2-$^{13}C_2$]acetyl derivative of the camphorsultam to yield a [1,2-$^{13}C_2$]acyl sultam. The $^{15}N\alpha$ group is added by treatment of the [1,2-$^{13}C_2$]-acylsultam enolate with 1-chloro-1-[^{15}N]-nitrosocyclohexane to yield L-[^{15}N, 1,2-$^{13}C_2$]-hydroxyamino acid. Conditions described by Oppolzer are used to reduce the hydroxylamino group and to remove the sultam without racemization of the product L-[α-^{15}N, 1,2-$^{13}C_2$]amino acid. Using this scheme, we have prepared L-[^{15}N, 1,2-$^{13}C_2$]alanine and L-[^{15}N, 1,2-$^{13}C_2$]-phenylalanine.

A trivial extension of this route will allow backbone-labeled amino acids to be deuterated at the α and/or β positions. The α-proton is derived from the methyl protons in acetic acid, which can easily be exchanged with ^2H_2O. Alkylating the acetyl sultam with [1-$^2H_2$]alkyl halides results in deuterium substitution in the β-methylene position. The deuterated alkyl halides are prepared efficiently by lithium aluminum deuteride reduction of the corresponding carboxylates to yield the [α-$^2H_2$]alcohols followed by conversion to the [1-$^2H_2$]alkyl halides. As discussed above, Bax and coworkers have shown that proteins labeled in the backbone with ^{13}C and ^{15}N and substituted with deuterium at the α and β positions are useful for sequential assignments in large proteins (Grzesiek *et al.*, 1993).

Stereoselective Labeling of Prochiral Groups in Amino Acids

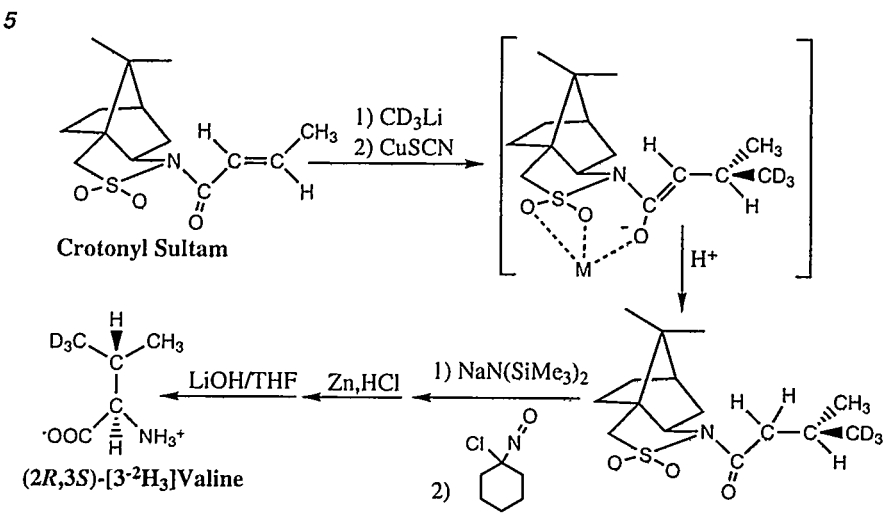
Camphor Sultam-Based Labeling of Prochiral Methyls in Valine

Oppolzer and coworkers have demonstrated the chiral 1,4 addition of nucleophiles to crotonic acid linked as an amide to the camphor sultam chiral auxiliary (Oppolzer and Kingma, 1989). We examined this route as a method for differential labeling of the methyl groups in valine. The crotonyl sultam was treated with a methyl copper reagent generated *in situ* by the addition of [$^2H_3$]methyl lithium and CuSCN (Scheme 5). This effected the stereospecific 1,4 addition of the [$^2H_3$]methyl group to the pro*S* face of the crotonyl (1*S*)-(-)-2,10-camphorsultam. Addition of the amino group to the (3*S*)-[3-$^2H_3$]valeroyl sultam as described above yielded (2*R*,3*S*)-[3-$^2H_3$]-valine. By starting with crotonyl (1*R*)-(+)-2,10-camphorsultam the same reactions would yield the natural enantiomer (2*S*) of valine labeled in the pro*R* methyl group. By removing the (3*R*)-[3-$^2H_3$]-valeric acid from the sultam, extending the chain by one carbon to form (4*R*)-4-[4-$^2H_3$]methylpentanoic acid, the methyl groups in leucine could be differentially labeled.

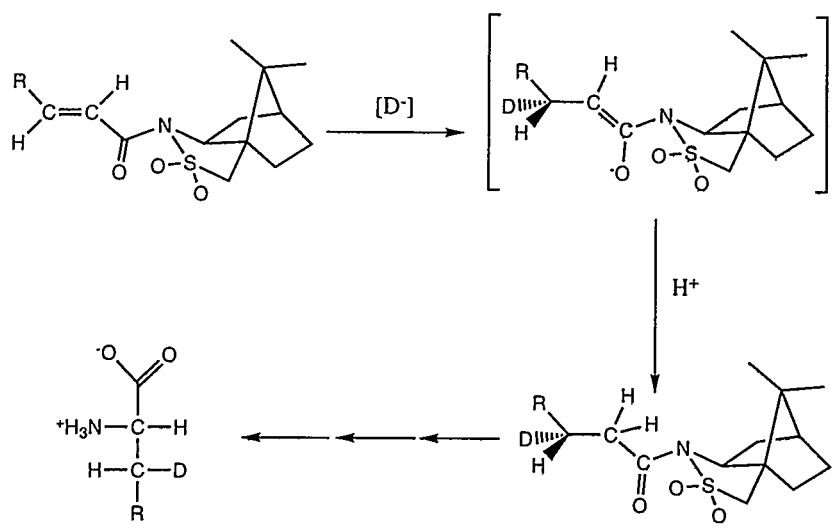
Chiral Deuteration of the β-Methylene Group

In a manner similar to that in the reaction described above, the 1,4 addition of deuterides to camphorsultam-linked acrylates has potential for the synthesis of amino acids chirally deuterated at the β-methylene position (Scheme 6). For example,

Scheme 5



Scheme 6



we treated cinnamoyl camphorsultam with sodium borodeuteride and isolated a low yield (20%) of [3-^{2}H]hydrocinnamoyl camphor sultam. Based on ^{1}H and ^{13}C NMR spectra of the isolated product, the addition of the deuteride was face-specific, yielding only one of the possible diastereomers. Using methods discussed above, it is straightforward to convert hydrocinnamoyl camphorsultam to L-phenylalanine. We have not yet assigned the absolute stereochemistry of the product and are currently investigating other hydride transfer reagents that will potentially increase the yield of this reaction. With the possible exceptions of serine and cysteine, this scheme has the potential of providing a route to differential labeling the β-methylene protons on all of the common amino acids.

Summary

For magnetic resonance and vibrational spectroscopies to reach their full potential, they must be used in combination with sophisticated site-specific stable isotope labeling of biological macromolecules. Labeled amino acids are required for the study of the structure and function of enzymes and proteins. Because there are 20 common amino acids, each with its own distinguishing chemistry, they remain a synthetic challenge. The Oppolzer chiral auxiliary provides a general tool with which to approach the synthesis of labeled amino acids. By using the Oppolzer auxiliary, amino acids can be constructed from several small molecules, which is ideal for stable isotope labeling. In addition to directing the stereochemistry at the α-carbon, the camphorsultam can be used for stereo-specific isotope labeling at prochiral centers in amino acids. By using the camphorsultam auxiliary we have the potential to synthesize virtually any isotopomer of all of the common amino acids.

References

Matthews, D.E. and Bier, D.M. (1983) *Ann. Rev. Nutr. 3*, 309.

Floss, H.G. and Beale, J.M. (1989) *Angew. Chem. Int. Ed. Engl. 28*, 146-177.

Houck, D.R., Hanners. J.L., and Unkefer, C.J. (1991) *J. Am. Chem. Soc. 113*, 3162-3166.

London, R.E. (1984) in *Topics in Carbon-13 NMR Spectroscopy*, Vol. 4, G.C. Levy, ed., John Wiley and Sons, New York, pp. 54-90.

London, R.E. (1980) in *Magnetic Resonance in Biology, Vol. 1*, J.S. Cohen, ed., John Wiley and Sons, New York, pp. 1-69.

Griffey, R.H. and Redfield, A.G. (1987) *Quart. Rev. Biophys. 19*, 51-82.

Oschkinat, H., Griesinger, C., Kraulis, P.J., Sorensen, O.W., Ernst, R.R., Gronenborn, A.M., and Clore, G.M. (1988) *Nature 332*, 374-376.

Kay, L., Clore, M., Bax, A., and Gronenborn, A. (1990) *Science 249*, 411-413.

Ikura, M., Kay, L.E., Krinks, M., and Bax, A. (1991) *Biochemistry 30*, 5498.

Kushland, D.M. and LeMaster, D.M. (1993) *J. Biomolec. NMR 3*, 710.

LeMaster, D.M. and Richards, F.M. (1988) *Biochemistry 27*, 142

Grzesiek, S. Anglister, J. Ren, H., and Bax, A. (1993) J. *Am. Chem. Soc. 115*, 4369-4370.

Oppolzer, W., Moretti, R., and Thomi, S. (1989) *Tetrahedron Lett. 30*, 6009-6010.

Oppolzer, W. and Tamura O. (1990) *Tetrahedron Lett. 31*, 991-994.

Eck, J.C. and Marvel, C.S. (1943) *Org. Syn. Coll. 2*, 76-78.

Muller, E., Metzger, H., and Fries, D. (1954) *Chem. Ber. 87*,1449-1453.

Oppolzer, W. and Kingma, A.J. (1989) *Helv. Chim. Acta 72*, 1337-1345.

Time-Resolved Infrared Studies of Protein Conformational Dynamics

William H. Woodruff,[1] Timothy P. Causgrove,[1] R. Brian Dyer,[1] and Robert H. Callender[2]

[1]Chemical Science and Technology Division
Los Alamos National Laboratory
Los Alamos, NM 87545

[2]Physics Department
City College of The City University of New York
New York, NY 10031

Understanding the ways in which the structure of a protein evolves with time during functional dynamics is the key to understanding the relationships between structure and function. A large number of structural and spectroscopic probes have been brought to bear on this problem. Unfortunately, however, the information gained is invariably incomplete. In general, the more structural information is derived from a measurement (for example, an x-ray crystal structure) the less information is gained about the structural dynamics and thus about the relationship between structure and function. On the other hand, experimental probes that give excellent dynamics information (for example, transient electronic spectroscopy) often give little or no direct information on the time evolution of structural changes.

Vibrational (infrared and Raman) spectroscopies offer a potential solution to this dilemma. Vibrational spectra are directly sensitive to the masses of atoms and groups bonded together, the strengths of the bonds themselves, and the geometric disposition of the bonded masses;

in short, to structure. Modern applications of vibrational spectroscopies in a time-resolved manner allow these spectroscopies to track changes in molecular structure that occur on ultrafast timescales (approaching the vibrational periods of chemical bonds: ~100 fs or less), as well as dynamics occurring on arbitrarily long timescales. In application to protein dynamics, time-resolved infrared (TRIR) spectroscopy is particularly valuable because the infrared spectrum contains observables that are sensitive to every structural feature of the polypeptide and its sidechains; thus the TRIR spectrum in principle offers quite complete information on the time evolution of biomolecular structure.

In practice, however, extracting the information on the behavior of specific structures from the infrared spectrum of a protein having ~10^5 vibrational modes is an extreme challenge. The task is made somewhat simpler by the fact that many of the structures show typical vibrational "group frequencies" such that one knows *a priori* what areas of the spectrum are relevant. This has been achieved by some 50 years of empirical observation of these spectra and their relationships with structure, by site-directed mutagenisis techniques, and by increasingly effective theoretical and computational approaches that aid rational interpretation of the spectra. Nevertheless, in a congested protein spectrum, there is only one way to identify unambiguously the vibrational behavior of a single, specific structure; that is to shift its vibrational frequency (without otherwise changing its chemical properties) by substitution of isotopes having different masses. Using this approach, it is possible to identify the infrared absorbance of one particular chemical bond in a protein having a molecular weight of ~100 kDa, and the TRIR experiment can reliably report the structural dynamics of that bond and associated groups.

In the following, we discuss the application of TRIR to two important classes of protein dynamics. The first concerns the functional dynamics of a protein in its native structural form. The system chosen here is the O_2 storage protein myoglobin. The reactions studied are the photodissociation and recombination reactions of CO at the heme site and the responses of the polypeptide structure to these reactions at the heme, as reflected by the TRIR spectra. The second type of reaction addresses the question of what events occur as an unfolded polypeptide chain assumes its native folded structure; that is, the protein folding problem, and in particular, the fast events in protein folding. To address these issues we observe the response of a synthetic α-helical polypeptide to a laser-induced temperature jump on timescales of 100 ps and longer. For each type of reaction, some qualitative insights can be gained from the results on the natural isotopic systems. However, the need for isotope studies in order to arrive at the desired, detailed structural interpretations of the results is evident.

Protein Response to Ligation Reactions in Myoglobin

The structure and dynamics of conformational changes in hemoglobin resulting in cooperativity in ligand binding has been a long-standing problem in biophysics. The use of myoglobin (Mb) as a simpler, noncooperative model protein has been established through spectroscopic and crystallographic examination. A central question in Mb ligand binding is the dynamics and energetics of motion of the protein relative to the heme. This problem has been illustrated by examining the carefully characterized x-ray crystal structure of Mb. If this compact structure remained fixed in solution, no pathway would exist for small ligands such as CO and O_2 to diffuse from solution into the heme binding site (Perutz *et al.*, 1965; Brooks *et al.*, 1988). Clearly, fluctuations of the protein in solution must play an important role in the function of the enzyme. Time-resolved infrared (TRIR) spectroscopy is uniquely suited as a probe for these processes, particularly the behavior of the protein, which generally is not observable by other spectroscopies (Dyer *et al.*, 1989). TRIR spectroscopy also has the capacity to study protein motions with minimal interference from the heme, in contrast to techniques based on visible light, such as resonance Raman and circular dichroism.

The use of time-resolved difference spectra in the amide I region provides a powerful tool for the study of protein dynamics. It has long been known that infrared spectra in the amide region are sensitive to protein secondary conformation (Elliott and Ambrose, 1950). Recent advances in equipment and techniques, including difference spectroscopy and various resolution enhancement techniques, have permitted researchers to quantitatively predict secondary structures from infrared spectra (Dong *et al.*, 1990, 1992; Dousseau and Pézolet, 1990; Kaiden *et al.*, 1987; Susi and Byler, 1986), particularly in the amide I region (Dong *et al.*,1992; Dousseau and Pézolet, 1990). In light of these results, it is now possible to study secondary structures in time-resolved experiments on protein dynamics and function.

The ligation reactions of small molecules such as CO with the heme site of Mb exemplify the mechanisms available to O_2—potentially revealing the molecular details of the enzyme function. CO is an ideal candidate for the initial TRIR experiments in the amide I region because it is easily photolyzed and the structure of both MbCO and unligated Mb have been extensively studied by crystallographic methods (Kuriyan *et al.*, 1986; Takano, 1977). In addition, CO ligation is often preferred in time-resolved spectroscopic studies because of the stability of MbCO in solution and because CO exhibits little geminate recombination (Henry *et al.*, 1983; Gibson *et al.*, 1986). Photolysis of CO from Mb and its recombination processes have been studied by a large array of techniques, including UV-Vis absorption (Henry *et al.*, 1983; Petrich et al, 1988, 1989), molecular dynamics (Kottalam

and Case, 1988), time-resolved CD (Xie and Simon, 1991), TRIR (Dixon *et al.*, 1988; Jedju *et al.*, 1988; Anfinrud *et al.*, 1989; Gerwert *et al.*,1985), time-resolved resonance Raman (Kim and Baldwin, 1990), and photoacoustic calorimetry (Dill and Shortle, 1991; Chan and Dill, 1993). Of these techniques, TRIR holds the most promise for providing detailed information on dynamics of the protein itself.

Infrared spectroscopy has already played a major role in investigating the ligand binding dynamics in Mb and Hb from the subpicosecond (Jedju *et al.*, 1988; Anfinrud *et al.*, 1989) to millisecond (Dixon *et al.*, 1988; Gerwent *et al.*, 1985) time scale. These studies probed the Fe-CO bleach at 1943 cm^{-1} and, in some cases, the 2135 cm^{-1} absorption of photolyzed CO trapped in the heme pocket. Dynamics of the protein, however, have yet to be probed by TRIR spectroscopy of the protein vibrations themselves.

Fast Events in Protein Folding/Unfolding

Native proteins are generally found in specific compact structures, and the nature of the particular structure is key to biological function. On the other hand, it is not understood how proteins arrive at the correct compact structure, which is just one of a small subset of states compared to the enormous number of possible biologically inactive structures. It is widely believed that most fold into their native structures in ways that are determined by primary sequence and the interactions between the component residues with themselves and with their surroundings. The molecular details of the folding process are under intense investigation, both theoretically and experimentally, because of the fundamental nature of this problem and because the future design of biologically active proteins and peptides must take into account how these biopolymers arrive at their final structure. (There are many excellent reviews on this problem: Kim and Baldwin, 1990; Chan and Dill, 1993; Privalov and Gill, 1988; Gething and Sambrook, 1992; Dill, 1990; Creighton, 1992; and Daggett and Levitt, 1994).

At present, most kinetic measurements of protein folding (or the reverse, protein unfolding) employ techniques based on the rapid mixing of solutions. For instance, the reversibly "denatured" unfolded protein in solution (for example, at high concentration of urea) is rapidly diluted or mixed with reagents that reverse the denaturing conditions in order to chemically initiate folding. Extremes of pH (either high or low) will unfold many proteins reversibly so that neutralization by mixing will cause them to refold. The best temporal response of studies based on mixing is typically 1 ms, the "dead time" of a stopped-flow apparatus. In addition, temperature may be used as a perturbation to fold and unfold proteins. The most commonly used probes of protein folding are NMR and circular dichroism (CD). NMR is quite structurally specific; in ideal cases 3D structure can be gained.

However, NMR as a dynamics probe is relatively slow. CD lacks the detailed structural specificity of NMR, but is sensitive to long-range chirality (namely secondary structure). It is clear from the foregoing that the short-time limit for the experimental observation of protein folding has, until now, been ~1 ms. However, it is experimentally clear that considerable changes in structure during folding occur in shorter times.

These faster events, which include the development of both secondary structure and aspects of the final folded tertiary structure, cannot be initiated or monitored using current experimental approaches. Resolution of the folding dynamics requires new approaches that retain structural specificity but have much faster time resolution than conventional stopped-flow and NMR techniques. As a recent case in point, chemical conditions were found whereby cytochrome *c* could be maintained in an unfolded state with CO bound to its heme site. Laser initiation of the folding reaction was accomplished by photodissociation of the bound ligand using nanosecond pulses, and specific kinetic changes on the microsecond time scale were monitored by probing the kinetics in the optical absorption of the visible chromophore bands (Jones *et al.*, 1993)

We have recently shown the feasibility of a rapid laser induced temperature jump (T-jump). A substantial temperature jump is accomplished using a short (picosecond time scale) infrared 'pump' pulse at a wavelength where water absorbs. The laser energy is absorbed by the water and the temperature of the volume of water in the laser beam is rapidly increased. Studies have shown that a temperature increase is transmitted across proteins on the picosecond time scale (Genberg *et al.*, 1987; Anfirud *et al.*, 1989). In our apparatus, a 30°C temperature jump can be achieved in 100 ps and is sustained for several milliseconds in a sample cell set up for TRIR absorption measurements (discussed below). Thus, for the first time, it is possible to initiate the kinetics of protein unfolding on a time scale with a resolution of 100 ps—some 10^7 times faster than current (rapid mixing) kinetic studies.

There are many probes of structure available that may be used to monitor the dynamical events on fast time scales: fluorescence, absorption, CD, and vibrational spectroscopy, to name a few. Our approach is to employ time-resolved vibrational spectroscopy to follow the time-course of these reactions. The advantages of time-resolved vibrational measurements as structure-specific probes of chemical dynamics are well established. The structural specificity derives from the connection of molecular vibrations to specific structures that determine the frequency, intensity, and linewidths of the absorptions. In principle, infrared and Raman spectroscopies yield information on every discrete structure and interaction in the protein because virtually all of the

normal vibrations of the protein are infrared and/or Raman active. The amide vibrations are sensitive to secondary structure, and vibrations of amino acid sidechains can monitor their environment and ionization state in certain important cases. The IR and Raman changes track molecular dynamics down to ultrashort time scales because vibrational transitions respond to changes in structure in times as short as the periods of the vibrations themselves (10 to 100 fs).

Laser and Spectroscopic Methods

Two TRIR spectrometers have been used in our studies to date, and the two cover different time scales. The first uses a real-time, single-wavelength technique developed for obtaining infrared transients on the nanosecond time scale. A widely tunable CW lead salt infrared diode laser functions as the probe, and a Q-switched Nd:YAG laser with a dye laser and doubling or mixing optics serve as the pump source. Transient changes in the transmission of the CW IR beam through the sample induced by a photochemical event are detected by a HgCdTe detector and are then digitized and signal-averaged. The response function of this instrument is ~300 ns and, with deconvolution of the response function with the measured transient, can resolve a response with a lifetime faster than 100 ns, as described by Causgrove and Dyer (1993). Given a weak signal ($\Delta A = 10^{-4}$) and the 10-Hz repetition rate of this system, ~15 min of signal averaging at each wavelength will yield a 50-to-1 signal-to-noise ratio.

The second TRIR spectrometer works on a timescale from 100 fs to tens of nanoseconds, depending on nonlinear optical techniques for production of short infrared pulses and optical delay for measurement of time. One version of this apparatus is shown in Fig. 1. The pump pulse is generated by a dye laser synchronously pumped by a frequency-doubled, mode-locked Nd:YAG laser and then is amplified in a three-stage dye amplifier pumped by a Nd:YAG regenerative amplifier. The infrared probe pulse is created by generating the difference frequency between the 532-nm pulse from the regenerative amplifier and part of the amplified dye pulse in a $AgGaS_2$ crystal. Tuning the dye laser yields infrared pulses tunable in the amide-I range. Before and after each protein measurement, the zero-of-time and the instrument temporal response (dependent solely on the pump/probe cross-correlation width) are determined by substituting for the sample a Si wafer in which the pump causes an instantaneous decrease in transmission of the probe. The temporal response varies with laser conditions but generally is well described by a $sech^2$ function of FWHM = 3 ps. This technique was described recently (Dyer *et al.*, 1991), except the wavelength coverage of the technique has been extended to below 1800 cm^{-1} ($AgGaS_2$ nonlinear material instead of $LiIO_3$) and the use of MCT infrared detectors has been added.

To rapidly heat a protein in solution for the folding/unfolding studies, a laser pulse at a wavelength where

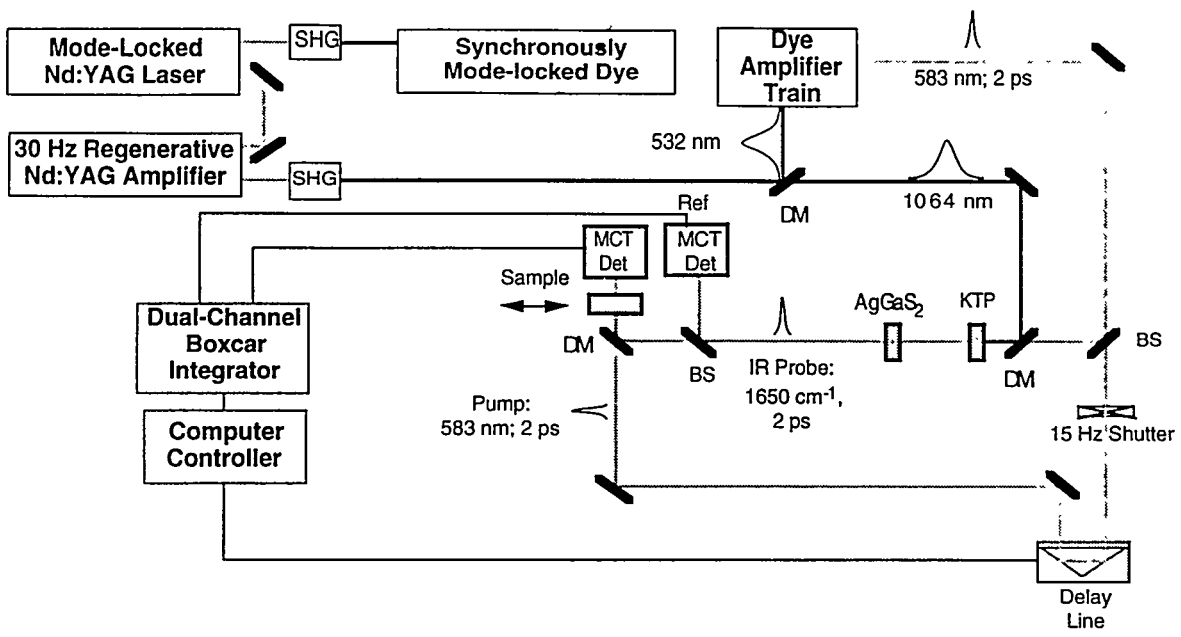


Fig. 1. Picosecond time-resolved infrared spectroscopy apparatus.

water has absorption has been used as the "pump" pulse in the T-jump experiments. Specifically, for the picosecond apparatus described above, the pump laser used is a regeneratively amplified Nd:Yag laser pumping an appropriate Raman shifter (one Stokes shift using hydrogen) to yield 100-ps pulses at 5000 cm^{-1}. This pulse width can be reduced to ~5 ps using a fiber and a grating to chirp and recompress the pulse. For the nanosecond apparatus, the pump 5000 cm^{-1} pulse is produced by difference-frequency generation in which the YAG fundamental is differenced with a dye laser. In either case, the laser energy is absorbed by the water and the temperature of the volume of water in the laser beam is increased within the time of the pump laser beam pulses because the temperature is transmitted across proteins on the 1- to 5-ps time scale (Genberg *et al.*, 1987; Anfirud *et al.*, 1989). The D_2O has weak absorption at 5000 cm^{-1} and this is ideal for our purpose because proteins do not absorb at 5000 cm^{-1}. The chosen frequency of the laser pump pulse is such that ~10% of the IR light is absorbed across the absorption cells. This ensures uniform heating of the sample within the interaction volume. Given the specific heat of water and an interaction volume of 1 $\mu\ell$ (100 x 100 x 100 μm^3), one can calculate that 0.2-mJ laser light absorbed will result in a 48°C temperature rise. A tighter focus results in high-power densities in the

interaction volume and a higher T-jump. Our typical runs employ 0.5 mJ of laser power, which produce an ~15°C T-jump (see below). This is quite sufficient for most experiments, but the equipment is capable of producing a T-jump of 30°C or more. The size of the T-jump is calibrated by using the differential change of D_2O absorbance *vs* temperature as an internal thermometer. The diffusion of heat to outside the interaction volume takes milliseconds, so that it is possible to generate a very fast temperature rise (as fast as the pump laser pulse, essentially) that lasts for milliseconds. The T-jump resolution of the real-time instrument is the laser pulse duration (~10 ns); thus the time resolution is limited by detector risetime, as noted above. The risetime of the T-jump for the picosecond instrument is 100 ps because of the width of the pump pulse; this can be improved if necessary.

Reaction initiation in the myoglobin experiments is effected by photodissociation of CO rather than by T-jump as in the folding experiments. In the realtime myoglobin experiments, the laser pump pulse was supplied by the 532-nm second harmonic of a Nd:YAG laser, with typical pulse energies of ~120 μJ. The measured decay was generated by subtracting traces taken with the Nd:YAG laser blocked from those with the laser unblocked. Ultrafast TRIR measurements on the myoglobin system were obtained by means of the pump-probe experiment (Fig. 1). The photodissociation pulse is supplied directly by the output of the amplified dye laser at ~595 nm. The temporal response varies with laser conditions but generally is well described by a $sech^2$ function of FWHM = 3.5 ps.

Results and Discussion on Myoglobin

Myoglobin dynamics from 100 ns to 10 ms

Typical decay traces of infrared absorption changes at 1943 and 1660 cm^{-1}, following laser photodissociation of CO, are shown in Fig. 2. The traces were taken under identical conditions (that is, using the same amplifier and HgCdTe detector) and were normalized to the same peak amplitude. Although the signal at 1660 cm^{-1} is smaller, the signal-to-noise ratios are similar because of greater laser power from the lower frequency diode and higher responsivity of the detector. Analysis of the traces was done according to a second-order kinetics rate law. The resulting fits gave second-order rate constants of 3.8 (1943 cm^{-1}) and 4.1 (1660 cm^{-1}) x 10^5 $M^{-1}s^{-1}$—within experimental error of each other but slightly lower than the value of 5.0 x 10^5 $M^{-1}s^{-1}$ determined by UV-Vis spectroscopy (Gibson *et al.*, 1986). Errors may be introduced in this analysis by the assumption of a saturated CO solution or by slightly different photolysis efficiencies between the two measurements.

Transients were also taken at several wavelengths using faster amplifiers in an attempt to resolve kinetics attributable to the motion of the protein in response to CO photodissociation. The rise of the transient at 1660 cm^{-1} is compared in Fig. 3 to the Fe-CO bleach at 1943 cm^{-1}, which is instantaneous on this time scale

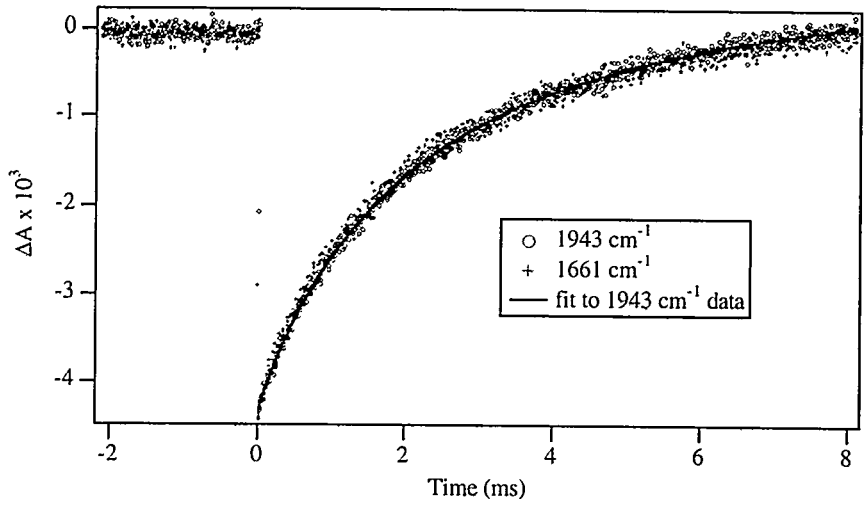


Fig. 2. TRIR absorption traces of 0.8-mM MbCO at 1943 and 1660 cm^{-1}. The data had peak absorbance changes of 4.3 x 10^{-3} and 1.6 x 10^{-3}, respectively. Each of the decay traces is an average of data from 1600 individual laser shots. The solid line is a three-parameter fit to the 1943 cm^{-1} data according to A+B second-order kinetics.

because the Fe-CO bond is broken within 300 fs (Anfinrud *et al.*, 1989). Clearly, the risetimes of these two transients are indistinguishable and therefore instrument-limited.

For all wavelengths probed in the amide I region, the rise of the transient absorption was limited by the response of the system, all showing a 10 to 90% rise time of 300 ns; with deconvolution, changes on a time scale down to ~100 ns would be observable with the present system. We conclude that protein motion, or at least the portion observable in the amide I IR band, is complete within 100 ns.

By tuning the frequency of the probe source, transients such as those shown in Fig. 3 were generated across the amide I band from 1608 to 1723 cm^{-1}. The initial amplitude of each of these transients was recorded and plotted against frequency to create a Mb-MbCO difference spectrum (Fig. 4B). The spectrum contains several features, the most prominent of which is peaked at 1664 cm^{-1} with an absorbance change of 1.5 x 10^{-3} (using 1.6 mM MbCO), compared to an absorbance of 0.8 for the total amide I peak. The spectrum shows several other features. The width of most peaks in the difference spectrum appears to be ~8 to 10 cm^{-1}. The TRIR difference spectrum was also examined for horse skeletal muscle Mb, which has a slightly different amino acid sequence than sperm whale Mb. The largest difference between the two samples is in the high-frequency region, where horse Mb shows a sharper 1680 cm^{-1} peak than sperm whale Mb with no signals observable in the 1700 to 1720 cm^{-1} range. Figure 4A shows the static difference spectrum (Mb-MbCO) obtained by subtracting Fourier trans-

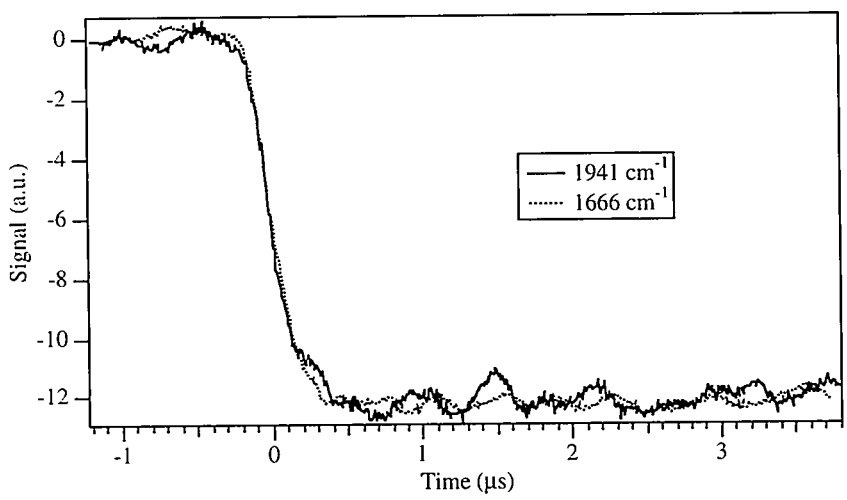


Fig. 3. Risetimes of TRIR transients at 1941 and 1666 cm^{-1}.

form infrared (FTIR) spectra of carbonmonoxy and deoxy Mb for sperm whale Mb. Although the baseline is clearly not flat in this region, all major features in the TRIR spectrum are reproduced in the static FTIR spectrum. This is convincing evidence that the spectrum obtained for MbCO 100 ns after photodissociation of CO is essentially identical in the amide I region as the static spectrum of deoxy Mb.

Ultrafast Myoglobin Dynamics

We have directly observed the photodissociation of CO from the heme binding site in picosecond TRIR experiments. The time-resolved bleaching of the Fe-CO absorption at 1943 cm^{-1} is indistinguishable from the instantaneous instrument response, determined before and after the protein measurement using a Si wafer. This observation is consistent with the femtosecond UV-Vis (Petrich *et al.*, 1988, 1989) and TRIR (Anfinrud *et al.*, 1989) measurements, which indicate that CO photodissociation occurs in less than 150 fs, probably on the time scale of one vibrational period of the Fe-CO stretch (520 cm^{-1}, 64 fs). The Fe-CO infrared transient was recorded for 100 ps following the pump pulse with no observable decrease in the bleach, which could be attributed to geminate recombination, indicating that the barrier to recombination must form rapidly. The absence of rapid geminate recombination of CO is typical for heme proteins and has been attributed to electronic (spin) barriers (Petrich *et al.*, 1988, 1989) and frictional forces resulting from iron displacement opposing rebinding (Dousseau and Péolet, 1990).

The protein response immediately following the photodissociation of CO was probed in additional picosecond TRIR experiments. The infrared transient

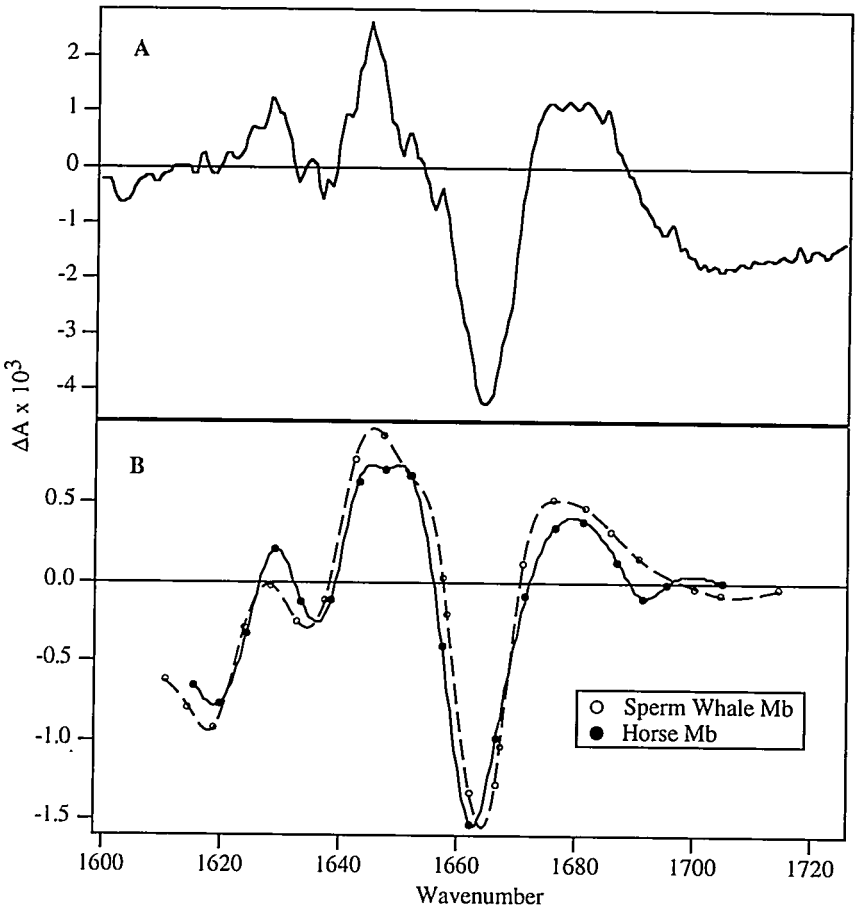


Fig. 4. (A) Difference spectrum (Mb-MbCO) generated from FTIR spectra of carbonmonoxy and deoxy Mb samples of 2-mM sperm whale Mb. (B) Difference spectra generated from initial amplitudes of IR transients of 0.8-mM sperm whale Mb (○) and horse skeletal muscle Mb (●) in the amide I region. The symbols are actual data; lines are cubic spline interpolations.

difference spectrum between 1610 and 1710 cm^{-1} is shown in Fig. 5 (top panel) and compared to the 100-ns spectrum shown previously. The intensity of the 1666 cm^{-1} bleach is exactly what was expected, based on the nanosecond difference spectrum. The experimentally determined instrument (Si) response is best fit with a $sech^2$ function of FWHM = 20 ps.

It is clear from Fig. 5 that the TRIR spectrum obtained 100 ns after photodissociation is experimentally indistinguishable from the static FTIR difference spectrum (MbCO - Mb), allowing for baseline uncertainties in the static FTIR spectrum (lower panel). There are, however, significant differences between the 50-ps transient difference spectrum (upper panel) and the 100-ns and static

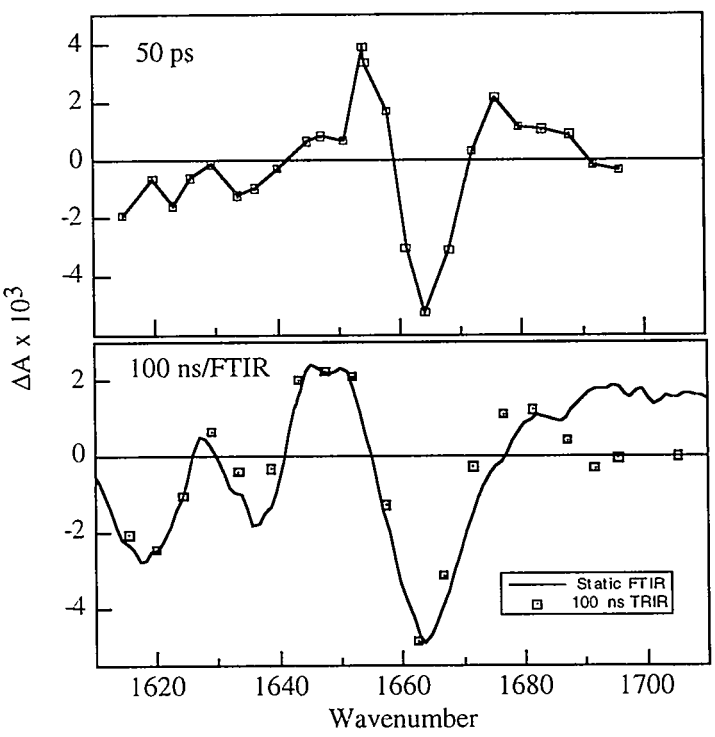


Fig. 5. Transient difference spectra of the horse heart Mb-CO photodissociation reaction. Upper panel, 50 ps after photodissociation; lower panel, 100 ns after photodissociation. The solid line in the upper panel is simply drawn through the experimental points. In the lower panel, the solid line is the static FTIR difference spectrum (Mb-CO-Mb).

spectra. Many general features are similar, but the double-peaked feature near 1650 cm^{-1} in the lower panel is a single peak at higher frequency in the 50-ps spectrum. This indicates (1) that the protein relaxations are complete within 100 ns and the structure at that time, by the amide I criterion, is indistinguishable from static, unliganded myoglobin; (2) that a substantial fraction of the relaxation has already occurred within 50 ps of photodissociation, most notably the major bleach is fully developed; and (3) there are significant relaxations that occur between 50 ps and 100 ns because the spectra at these two times are not identical. The precise nature of the structural changes on the various time scales has not been determined, and isotope labeling studies will be required before this can be done.

The assignment of infrared band shapes and subbands from steady-state FTIR spectra to specific secondary structures is currently an active area in the literature. The main techniques for resolving subbands in the amide I region have been resolution enhancement by Fourier self-deconvolution (Susi and Byler, 1986; Yen *et al.*, 1983;

Ma *et al.*, 1989) and generation of second derivative spectra (Dong *et al.*, 1990, 1992) It should be noted that both of these techniques require very high signal-to-noise and careful elimination of water vapor bands. Recently criteria for the subtraction of water and water vapor backgrounds have allowed consistent bands beneath the amide I envelope to be assigned to specific secondary structures.

There are many complications in the comparison of these assignments to the features seen in the Mb difference spectra shown in Figs. 4 and 5. First, the difference spectrum is the sum of all changes in the protein, which may involve a complex superposition of bands. Also, there may be contributions to the difference spectra from amino acid sidechain vibrations. These are generally considered small when measuring the overall amide I spectrum; however, the difference spectrum is sensitive to only those vibrations that change, which may increase the relative contribution of specific sidechain vibrations that undergo large conformational changes. Comparison of the area of the overall amide I peak to that of the 1664 cm^{-1} bleach in Fig. 4 (corrected for photolysis efficiency) gives a ratio of 350 to 1—corresponding to a bleach of less than one backbone CO vibration. Therefore, the contribution of sidechain vibrations in the difference spectrum must be considered. The role of Arg^{45} (CD3), in particular, has been the subject of several examinations (Kuriyan *et al.*, 1986; Westrick *et al.*,1990; Chance *et al.*, 1990). X-ray structures have shown that Arg^{45} forms a salt bridge with one of the propionate groups of the heme and that the breaking of the salt bridge is an important (and one of the largest) structural differences between the CO and deoxy forms of Mb (Kuriyan *et al.*, 1986). This may be important in the interpretation of the IR difference spectra as arginine is one of the amino acids likely to show sidechain vibrations in the 1600 to 1700 cm^{-1} region. Horse Mb does not have arginine; this may be reflected in the high-frequency end of the horse TRIR difference spectrum, which is very different from that of the sperm whale Mb (Fig. 4B).

Tables of amide I band assignments have been made by Dong, *et al.* (1990) for proteins in H_2O solution and by Susi and Byler (1986) for a standard protein set in D_2O. Although there is general agreement between the two assignments, there are some differences in the exact frequency assignments. The most straightforward assignment is that for α–helix, which shows only a single narrow band at 1653 to 1656 cm^{-1}. Mb, which is 80% α–helix, is dominated by a single peak at 1654 cm^{-1} (Dong *et al.*, 1990). Interestingly, the Mb difference spectra have an isosbestic point near this frequency, consistent with the finding that the helix structures of Mb are quite similar between the CO and deoxy forms, but only shift relative to each other (Kuriyan *et al.*, 1986). The bleach at 1664 cm^{-1} in Fig. 4 is in the region assigned to turn structures; the bands at 1676 cm^{-1} (a transient absorption

in Fig. 4) have been assigned to either turn or extended chain conformations. The other main absorption at 1645 cm^{-1} is in the area assigned to unordered structure. Bands at lower frequency have been assigned to both either β-sheet (Dong *et al.*, 1990) or general β-type (Susi and Byler, 1986) structures. In the case of Mb, the lower frequency signals must represent bleaching of the small regions of β-turns because the protein contains no β-sheet (Kuriyan *et al.*, 1986; Takano, 1977). To make specific assignments to each of the spectral features in Figs. 4 and 5, additional data is required. Further information on the interpretation of these spectra could be gained from perturbed systems. For example, the contribution of some ionizable sidechains is, in principle, available from pH-dependent studies; other sidechain vibrations can be addressed through selective isotopic labeling and site-directed mutagenesis.

The most important result of this study is the measurement of the protein dynamics associated with the photodissociation of CO from the heme. What is clear from the previous discussion is that the dynamics measured are associated with the changes in protein backbone conformation to which the amide I vibrations are most sensitive, and possibly the motion of specific sidechains which also have absorptions in this spectral region. The "realtime" TRIR measurements indicate that the conformational change from the carbonmonoxy to the deoxy form is over within 100 ns of the photodissociation. The picosecond TRIR experiments give information on the initial stages of protein motion and provide a connection between molecular dynamics simulations and actual protein motion. The picosecond absorbance transient for Mb at 1666 cm^{-1} suggests that a significant part of the conformational change occurs in 15 ps. These dynamics can be compared to indirect measurements which used resonance Raman spectroscopy to probe heme pocket relaxation (Findsen *et al.*, 1985). These studies suggest that the tertiary structural changes in the position of the proximal histidine are complete within less than 30 ps. In these experiments, the Fe-N(His) stretching frequency reaches its equilibrium deoxy value within the 30-ps pulsewidth of the resonance Raman probe pulse. Another indirect measure of the protein response to CO photodissociation was obtained by Miller *et. al.* from picosecond phase grating spectroscopy (Surewicz and Mantsch, 1988). These results suggest that a global change in the protein structure is occurring in less than 30 ps. In contrast, Simon *et. al.* report that relaxation of the transient circular dichroism spectrum probed at 355 nm of photodissociated MbCO to the equilibrium deoxy spectrum requires 300 ps (Xie and Simon, 1991). A phase of the conformational relaxation dynamics that is slower than 50 ps but faster than 100 ns is also evident in our data (Fig. 5) and this may correspond to Simon's 300-ps transient. Additional picosecond experiments are addressing this question. It is tempting to suggest that the faster relaxations are associated

with protein structures on the proximal side of the heme that are altered in response to the out-of-plane motion of the iron-histidine unit upon CO photodissociation, and the slower ones are associated with distal structures that change in response to migration of the photodissociated CO through the protein structure. These remain speculations, however, in the absence of isotope data.

Results and Discussions: Polypeptide Folding/Unfolding

We have obtained preliminary data on the "F_S peptide" (Suc-$Ala_5(Ala_3ArgAla)_3$-Ala-NH_2), which is 21 residues long (Suc = succinyl). This peptide was a gift from Peter Kim; he and his colleagues have shown that it forms >90% helical structure in water at low temperatures (Lockhart and Kim, 1992). The IR spectrum in the amide-I region of the F_S peptide in D_2O was taken at several temperatures (concentration of 0.5 mM, pathlength 100 μm, sample volume 200 μℓ, ~0.2 mg of peptide). A peak at 1675 cm^{-1} was found and is attributed to C=O stretches of the succinyl group (and not from amide-I bands). It undergoes small changes in intensity and position over the temperature range taken, likely indicating small changes in hydrogen bonding (in general, C=O stretches shift to lower frequencies with stronger hydrogen bonds; see, for example, Callendar and Deng, 1994; Callender *et al.*, 1989). At low temperature, the amide-I band (really the amide-I' band because the peptide is in D_2O) was found to lie at 1632 cm^{-1}. The 1632 cm^{-1} species, which is assigned to the helix structure, converts to a species whose amide-I frequency lies at 1651 cm^{-1} (presumably unfolded peptide), and this takes place over a broad temperature range. The difference spectrum between 30 and 20°C is shown as the solid line in Fig. 6.

The time-dependent changes in IR absorbance were measured at several wavelengths in the spectrum of the F_S peptide after the 5000 cm^{-1} "pump" light initiated the T-jump. The optical absorbance changes measured 5 μs after the T-jump are plotted in Fig. 6 as solid circles. The changes in the absorbance of the F_S peptide occur within the 300-ns resolution of the instrument. Thus, the helix unfolding transition of the F_S peptide occurs faster than 300 ns. The refolding kinetics of the peptide as the temperature decreases contains two time constants. Most of the spectrum tracks the ~4-ms decay of the temperature within the interaction area. This result is expected because the transition of the F_S peptide from folded to unfolded follows an extended range of temperature (see Fig. 6; this is behavior typical of small peptides). What is unexpected is a transient with a faster decay (~0.4 ms) than the temperature decay kinetics. The dynamics of the fast-decay component are also plotted in Fig. 6. We are not certain about the meaning of this result at present. Measurements employing the picosecond apparatus described above have been performed out to 500 ps. No changes in the spectrum were observed. Thus, the unfolding kinetics take place during

500 ps to 300 ns. We are now performing experiments to bridge the gap between 500 ps and 300 ns.

Clearly, the study of the early dynamics of the protein folding problem has barely begun. The studies performed so far in our laboratory just whet our appetite for more work. However, some important general issues are apparent even from these limited studies. One is that stopped-flow techniques are hopelessly inadequate to study protein folding, and the need for fast initiation protocols to study protein folding is obvious. The dynamics of protein folding clearly involve events on the submicrosecond time scale. A characteristic time scale of the helix-coil transition is somewhere between 500 ps and 300 ns, and the helix-coil transition is likely a major motif in the folding of a majority of proteins. It is clear from these results that ultrafast events such as the nucleation of secondary structure from a random conformation are possible along the protein folding pathway. Isotope studies are required to establish (1) which parts of the polypeptide structure are involved in the unfolding/refolding as observed by TRIR and (2) the degree of homogeneity or inhomogeneity of both the folded and unfolded forms.

Conclusions

We have demonstrated that TRIR in the amide I region gives structural information regarding protein conformational

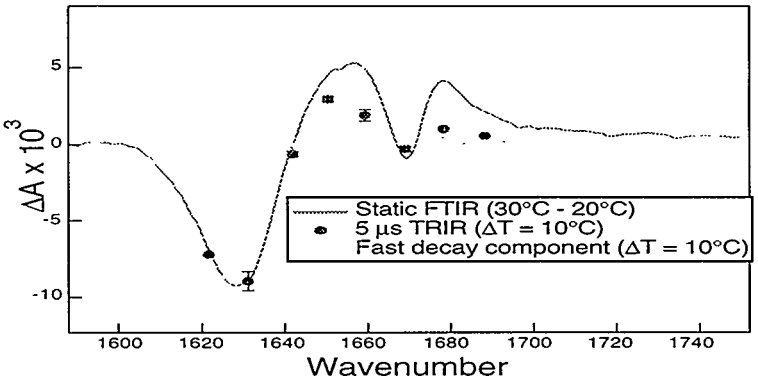


Fig. 6. A comparison of the static and transient (taken at 5 μs after the T-jump) IR difference spectra throughout the amide-I spectral region. Also shown is the "fast decay'" component. The transient data tracks well the static data in the 1620- to1650-cm^{-1} range, but the IR peaks identified with the succinyl group do not. This probably indicates that the helical nature of the F_S peptide has different dynamics than the succinyl group—which would not be unexpected. The fast decay component also tracks reasonably well the static FTIR spectrum, although it is difficult to see this on the graph because of its much smaller amplitude.

changes in realtime, both on processes involved in the development of the functional structure (protein folding) and on protein structural changes that accompany the functional dynamics of the native structure. Assignment of many of the amide I peaks to specific amide or sidechain structures will require much additional effort. Specifically, the congestion and complexity of the protein vibrational spectra dictate that isotope studies are an absolute requirement for more than a qualitative notion of the structural interpretation of these measurements. It is clear, however, that enormous potential exists for elucidating structural relaxation dynamics and energetics with a high degree of structural specificity using this approach.

Acknowledgements

The authors gratefully acknowledge support from The National Institutes of Health grants GM35183 (RHC) and GM45807 (WHW) as well as LDRD grants from Los Alamos National Laboratory, XL60 (WHW, RBD) and X15B (RBD). TPC acknowledges support from the Director's Postdoctoral Fellowship program at LANL. This work was performed at Los Alamos National Laboratory under the auspices of the US Department of Energy.

References

Anfinrud, P.A., Han, C., and Hochstrasser, R.M. (1989) *Proc. Natl. Acad. Sci. USA 86*, 8387-8391.

Brooks, C.L., Karplus, M., and Pettitt, B.M. (1988) *Adv. Chem. Phys. 71*, 111.

Callender, R., Deng, H., Sloan, D., Burgner, J., and Yue, T.K. (1989) in *Proceedings of International Society for Optical Engineering* , International Society for Optical Engineering, Los Angeles, California, pp. 154-160.

Callender, R. and Deng, H. (1994) *Ann. Rev. Biophys. Biomol. Struct. 23*, 215-245.

Causgrove, T.P. and Dyer, R.B. (1993) *Biochemistry 32*, 11985-11991.

Chan, H.S. and Dill, K.A. (1993) *Physics Today (February)*, 24-32.

Creighton, T.E., ed. (1992) *Protein Folding*, Freeman, New York.

Chance, M.R., Courtney, S.H., Chavez, M.D., Ondrias, M.R., and Friedman, J.M. (1990) *Biochemistry 29*, 5537-5545.

Daggett, V. and Levitt, M. (1994) *Cur. Opin. Struct. Biol. 4*, 291-295.

Dill, K.A. (1990) *Biochemistry 29*, 7132-7155.

Dill, K.A. and Shortle, D. (1991) *Ann. Rev. Biochemistry 60*, 795-825.

Dixon, A.J., Glyn, P., Healy, M.A., Hodges, P.M., Jenkins, T., Poliakoff, M., and Turner, J.J. (1988) *Spectrochim. Acta 44A*, 1309-1314.

Dong, A., Huang, P., and Caughey, W.S. (1990) *Biochemistry 29*, 3303-3308.

Dong, A., Huang, P., and Caughey, W.S. (1992) *Biochemistry 31*, 182-189.

Dousseau, F. and Pézolet, M. (1990) *Biochemistry 29*, 8771-8779.

Dyer, R.B., Einarsdóttir, Ó., Killough, P.M., López-Garriga, J.J., and Woodruff, W.H. (1989) *J. Am. Chem. Soc. 111*, 7657-7659.

Elliott, A. and Ambrose, E.J. (1950) *Nature 165*, 921-922.

Dyer, R.B., Peterson, K.A., Stoutland, P.O., and Woodruff, W.H. (1991) *J. Am. Chem. Soc. 113*, 6276-6277.

Findsen, E.W., Friedman, J.M., Ondrias, M.R., and Simon, S.R. (1985) *Science 229*, 661-665.

Genberg, L., Heisel, F., McLendon, G., and Miller, R.J.D. (1987) *J. Phys. Chem. 91*, 5521.

Gerwert, K., Rodriguez-Gonzalez, R., and Siebert, F. (1985) in *Time-Resolved Infrared Spectroscopy Applied to Photobiological Systems*, A. Laubereau and M. Stockburger, eds., Springer-Verlag, Berlin, pp. 263-268.

Gething, M.-J. and Sambrook, J. (1992) *Nature 355*, 33-45 .

Gibson, Q.H., Olson, J.S., McKinnie, R.E., and Rohlfs, R.J. (1986) *J. Biol. Chem. , 261*, 10228-10239.

Henry, E.R., Sommer, J.H., Hofrichter, J., and Eaton, W.A. (1983) *J. Molec. Biol. 166*, 443-451.

Jedju, T.M., Rothberg, L., and Labrie, A. (1988) *Opt. Lett. 13*, 961-963.

Jones, C.M., Henry, E.R., Hu, Y., Chan, C.-K., Luck, S.D., Bhuyan, A., Roder, H., Hofrichter, J., and Eaton, W.A. (1993) *Proc. Natl. Acad. Sci. USA 90*, 11860-11864.

Kaiden, K., Matsui, T., and Tanaka, S. (1987) *Appl. Spectros. 41*, 180-184.

Kim, P.S. and Baldwin, R.L. (1990) *Ann. Rev. Biochem. 59*, 631-660.

Kottalam, J. and Case, D.A. (1988) *J. Am. Chem. Soc. 110*, 7690-7697.

Kuriyan, J., Wilz, S., Karplus, M., and Petsko, G.A. (1986) *J. Mol. Biol. 192*, 133-154.

Lockhart, D.J. and Kim, P.S. (1992) *Science 257*, 947-951.

Ma, H.M., Liu, Y.X., Fei, Y., and Li, F.M. (1989) *J. Appl. Phys, 65*, 5031-5034.

Perutz, M.F., Kendrew, J.C., and Watson, H.C. (1965) *J. Mol. Biol. 13*, 669-678.

Petrich, J.W., Poyart, C., and Martin, J.L. (1988) *Biochemistry 27*, 4049-4060

Petrich, J.W. and Martin, J.L. (1989) *Chem. Phys. 131*, 31-47.

Privalov, P.L. and Gill, S.J. (1988) *Adv. Protein Chem. 39*, 191-233 .

Stoutland, P.O., Dyer, R.B., and Woodruff, W.H. (1992) *Science 257*, 1913-1917.

Surewicz, W.K. and Mantsch, H.H. (1988) *Biochim. Biophys. Acta 952*, 115-130.

Susi, H. and Byler, D.M. (1986) *Meth. Enzymol. 130*, 291-311.

Takano, T. (1977) *J. Mol. Biol. 110*, 569-584.

Westrick, J.A., Peters, K.S., Ropp, J.D., and Sligar, S. (1990) *Biochemistry 29*, 6741-6746.

Xie, X. and Simon, J.D. (1991) *Biochemistry 30*, 3682-3692.

Yen, R., Shank, C.V., and Hirlimann, C. (1983) *Mater. Res. Soc. Symp. Proc. 13*, 13-16.

Discussion Group Reports

GHz Nuclear Magnetic Resonance

Discussion Leaders

Timothy A. Cross, Gary Drobny, and Jill Trewhella

Participants

William W. Bachovchin, Adrian Bax, Walter L. Chazin, Betty Deingar, Michael Kennedy, David Live, John L. Markley, Nicholas A. Matwiyoff, Gregory J. Moore, Raymond S. Norton, Linda Okerlund, Donni Rajapopal, Anthony S. Serianni, Jacob Schaefer, Clifford J. Unkefer, Santhana Velupillai, Peter E. Wright, Erik R.P. Zuiderweg

The views expressed in this report do not necessarily represent the consensus of the participants, but rather a collection of thoughts from this discussion session.

Objective

This discussion group addressed three basic questions:

- What are the potential scientific benefits of GHz NMR in structural biology research?
- What problems are anticipated in obtaining GHz NMR spectra of biological molecules?
- What are the resolution and stability requirements for biomolecular structural work at 23.5 Tesla?

Introduction

For the past dozen years, 500- and 600-MHz spectrometers have become available in many laboratories. The first 600-MHz NMR spectrometer (at Carnegie Mellon University) was commissioned more than 15 years ago and, until 1994, represented the highest field available for high-resolution NMR. This year, we have witnessed unprecedented progress in the development of very high field magnets for NMR spectroscopy, including the delivery of the first commercial 750-MHz NMR spectrometers. In addition, NMR signals have been obtained from 20-Tesla magnets (850 MHz for ^{1}H's) at both Los Alamos National Laboratory and Florida State University in the NHMFL (National High Magnetic Field Laboratory). These preliminary

experiments have been performed in magnets with 100-ppm homogeneity, but a 20-Tesla magnet developed for the NHMFL will be brought to field this year with a projected homogeneity of 0.1 ppm over a 1-cm-diam spherical volume.

Oxford Instruments has signed a contract to produce a high-resolution 21.15-Tesla (900-MHz) magnet for the Environmental Molecular Sciences Laboratory at Pacific Northwest Laboratory. The NHMFL has completed its Benchmark design of an NMR-quality, widebore 21.15-Tesla (900-MHz) magnet that will be upgraded to 25 Tesla (1.06 GHz). This jump to 25 Tesla will require development of high-T_c superconducting materials. A 40-MW dc power supply with stability of 2 ppm installed at the NHMFL offers the possibility of developing resistive or hybrid magnets with field homogeneity and stability of 1 ppm or less. With the development of a long-pulse 60-Tesla magnet, NMR in pulsed magnetic fields is also being pursued at the NHMFL in Los Alamos. Although the impact on the macromolecular community may not be immediately apparent, these magnets are being designed to achieve the highest possible stability and homogeneity for magnetic resonance studies.

We have made considerable progress in magnet technology during the past 15 years, but there has been limited application of this technology to the development of very high field NMR-quality magnets. The lack of competition between NMR magnet manufacturers may have been partially responsible for this slow development; however, that situation has changed dramatically. Clearly, substantial advances in field strength are now possible. Optimization of these fields for structural biology must be planned now if we are to receive the potential benefits of this new technology.

Anticipated gains in sensitivity at higher fields will not be realized without the development of RF technology in general and probe technology in particular. For example, at these high frequencies, efficient probes may require blending NMR and low-frequency ESR technology. RF development will call for a significant commitment of effort and the requirements must be anticipated now.

Looking to the future of high-field NMR and the role of GHz magnet technology, we asked these specific questions:

- If 25-Tesla fields combined with a 10- to 10^3-fold increase in sensitivity were available, what new science would be accessible?
- What potential problems and additional advantages are associated with NMR spectroscopy at much higher frequencies than those currently available for structural, mechanistic, and dynamic characterizations of biomolecules? In what ways are these problems limiting?
- What magnet specifications are needed for structural biology applications? What improvements in RF and probe design are possible?

- What will be the role of isotope-labeling experiments at high-field strengths?

For the purposes of the discussion, we deliberately set aside questions about the cost and feasibility of the proposed technology in order to focus on the scientific potential. During the last 30 years, there have been remarkable improvements in NMR instrumentation, and each incremental improvement has been accompanied by an investment cost as well as a significant increase in the price of the resulting NMR spectrometer. To date, the benefits gained have consistently justified the investments. Indeed, the application of NMR spectroscopy to significant biomolecular structures would have been impossible had fields of 11.75-Tesla and greater not been available. Furthermore, there has been a very substantial technological development in both the RF console and the experiments used to extract structural information from these biological macromolecules. It is this combination of science and technology that has resulted in such an important new structure-determination technique.

The next incremental advance of magnetic field strength—to 23.5 Tesla (1 GHz)—will again require a significant investment in both advanced magnet technology and development of high-frequency RF electronics. The next sections discuss science that might be done if the investment is made.

If 23.5-Tesla fields, as well as a 10- to 10^3-fold increase in sensitivity, were available, what new science would be accessible?

Increased resolution

The increased resolution obtained with GHz fields may have its greatest impact in the fields of carbohydrate and polynucleotide structural studies. The ability to study polysaccharide and polynucleotide structures is increasingly important. Polysaccharides play key roles in cell surface recognition, which is fundamental to an array of health issues, including vaccine development and drug design. Understanding the mechanisms by which proteins interact with DNA and RNA and thus regulate the flow of genetic information is one of the most fundamental questions in biology and also is key to the development of molecular medicine. This understanding requires structural data for both polynucleotides and their complexes with proteins. With current limitations on resolution in NMR experiments, both polysaccharides and polynucleotides present problems of severe spectral overlap because the sugar resonances have so little frequency dispersion. GHz spectroscopy has the potential to provide enough resolution to make studies of larger carbohydrates and polynucleotides a reality.

Improved Sensitivity

With 1 to 3 orders of magnitude improvement in sensitivity, NMR experiments on smaller or less-concentrated samples are

possible immediately. This capability would greatly help with the problem of aggregation, which is common and is devastating for spectroscopic studies of many biological systems. Similarly, proteins with low solubility can be studied. Limitations currently imposed by the solubility and amount required for NMR studies (typically, 0.5 mℓ at millimolar concentrations) exclude the vast majority of proteins that have been identified to date. An increase in sensitivity also means that NMR data sets can be achieved in less time with a similar amount of material. Recording NMR spectra from single crystals of proteins becomes feasible, facilitating comparisons with NMR solution structures of the same protein and/or x-ray structure data from the same crystal form. Spectroscopy of natural-abundance ^{13}C and ^{15}N samples becomes easier. Different approaches for isotopic labeling and macromolecular purification can be used if micromolar instead of millimolar concentrations are needed. For instance, isotopic labeling by tissue culture methods would become substantially easier if much less protein were required.

A 3-orders-of-magnitude improvement in sensitivity at the currently available NMR field strengths would also permit greatly improved structural constraints. For example, NOE intensities could be determined at much greater signal-to-noise, making it possible to detect distances up to 15 Å. Nucleic acid structures, in particular, have been poorly constrained by the current NOE restriction—distances of 5 to 6 Å. With constraints up to 15 Å, nucleic acid structures could be very well defined and protein structures much more precisely described.

Many key biological and chemical processes take place in or on surfaces (for example, cell membranes). These processes are important not only to issues in health, but also to the development of biotechnologies, bioremediation, and chemical synthesis and degradation. Increased sensitivity means that NMR spectroscopy of two-dimensional crystals such as membrane proteins or biomolecular materials becomes feasible. The structure, dynamics, and solvent accessibility of biomineralized materials such as biosensors on supports are envisioned. The study of surfaces rather than bulk samples becomes possible. Consequently, chemistry occurring on catalytic surfaces could be studied as well as the chemistry of molecular separations occurring on modified silica supports.

What potential problems and additional advantages are associated with NMR spectroscopy at much higher frequencies than those currently available for structural, mechanistic, and dynamic characterizations of biomolecules? In what ways are these problems limiting?

As described above, the principal advantages of GHz NMR spectroscopy are the gains in resolution and sensitivity. There are other advantages associated with higher fields, as well as potential disad-

vantages, which in some circumstances may limit the range of applications for NMR spectroscopy at 25 Tesla.

Chemical Shift Dispersion

One of the advantages at high magnetic fields is that the strong coupling limit that occurs in aromatics and in sugar rings will be significantly alleviated at very high fields. Combined with increased dispersion, this will represent a major advance for studying nucleic acids and carbohydrates as well as aromatic residues in proteins.

Relaxation

Some individuals expect that relaxation problems (T_2) will be a limitation with very high field spectroscopy. Although there is little doubt that the static chemical shift anisotropy (CSA) for nonprotonated carbons will become a major relaxation mechanism, it is not clear what will happen to the spectral density function. It was argued that substantial correlated motions will be present in some macromolecular systems and that this will affect the spectral density function in ways that are not intuitively obvious. Therefore, it is unlikely that generalized statements about linewidths will be valid. Consequently, although CSA may be a disadvantage for high-field NMR spectroscopy, it may not be as great a problem as many believe.

T_1 relaxation appears to be an increasing problem at higher field strengths. This problem could be partially offset if anticipated gains in sensitivity are realized.

RF penetration

RF penetration depth will be a problem at very high frequencies, and this difficulty will be exacerbated by high salt concentrations in some biological samples. Like the RF heating problem, this results from electric fields generated by the electromagnetic radiation. Coil design, geometry, and shielding can be optimized to minimize electric fields in samples.

Magnetic alignment

It is still unlikely that magnetic alignment, which increases with the B_o^2, will be a major problem. This effect will continue to be primarily a bulk phenomenon rather than a single-molecule effect, and consequently, it should be concentration-dependent. At multi-GHz frequencies, the effects on single molecules may become more significant. If so, the techniques pioneered by Prestegard and Sanders to develop structural constraints from partially ordered systems could be used to provide a new constraint in nonliquid-crystalline samples.

Probe Design

One of the challenges at high fields will be to tune and match probe frequencies over a wide range from 100 MHz to GHz. Every advantage possible will be taken of ESR technology and new device technology such as that being developed in the low GHz range for the cellular phone industry. Developing probes with excellent B_1 field homogeneity will be essential but difficult to accomplish; however, working with horizontal solenoid coils may have advantages.

What magnet specifications are necessary? What improvements in RF and probe design are possible?

Magnet Design

This meeting is one of the first times that the spectroscopy community has had an opportunity to provide input into the development of the next-generation of magnets. These specifications are dependent upon the spectroscopy to be performed. The expertise of the individuals participating in this discussion encompasses high-resolution NMR as well as biological and chemical solid state NMR. Because several NMR disciplines were not present at this meeting, this discussion cannot include any comments from those areas; for instance, macrocyclic high-resolution NMR, where very high homogeneity is required because of the extremely complicated spin systems; microimaging, which can take advantage of sensitivity improvements by decreasing voxel size; and NMR in condensed matter physics (traditionally, this has required low homogeneity, but high fields may simplify the spectroscopy considerably).

This report represents the NMR disciplines that primarily impact structural biology. For chemical and biological solid state NMR, homogeneity between 1 and 10 ppm can be tolerated for most applications. This community's chief concern is sensitivity, and the group's priorities for magnet engineering lie in the development of very high fields with modest homogeneity and a 5- to 12-cm range of bore diameters.

For the high-resolution community, sensitivity is also the top priority, but the requirements for homogeneity vary with the specific experiment.

For carbohydrate spectroscopy, where much of the spectroscopy will continue to be performed with ^{1}Hs, resolution remains an important issue, although even here 10^{-3} ppm (over a 1-cm dsv) is probably more than necessary. For samples that have been uniformly ^{13}C and ^{15}N isotopically labeled, the linewidths in the ^{13}C and ^{15}N dimensions of pulse-field gradient multidimensional spectra are typically >0.05 ppm. The resolution in the ^{1}H dimension(s) are better, but possibly only slightly better than 0.05 ppm. Although this suggests the possibility of multidimensional NMR in lower homogeneity magnets, if solvent suppression is needed, line shapes of at least 20/40 (non-spin linewidth in Hz at 0.55 and 0.11% of the signal height) will be necessary.

RF and Probe Design

It was suggested that the probe coil could be placed in a sideways geometry, thereby achieving a sensitivity enhancement of 1.4. The arguments for vertical orientations of the sample tube no longer hold because in high-field experiments, the samples are not spun nor do they need to be changed frequently.

Today it is possible to consider the use of superconducting resonators for detecting NMR signals. The very high Q factors available with such devices will result

in both sensitivity advantages and Q-switching problems. The GaAs technology for pin diodes may be a mechanism for solving the Q-switching problem. Cryogenic cooling of the preamplifier and the leads to the probe will also have significant benefits. There are, in fact, many new electronic devices on the market that could significantly improve NMR consoles. The development of high-performance 10-mm NMR probes would theoretically improve sensitivity by a factor of 4. Conversely, microprobes—although they sacrifice sensitivity—use small samples and represent a significant improvement over samples for which large-scale preparations are prohibitive. It is possible that a combination of magnetic field strength and console/probe developments will lead to factor-of-10 sensitivity improvements.

What will be the role for isotope-labeling experiments at high-field strengths?

Quadrupole nuclei will be used to a much greater extent in high fields. Considerable biological action takes place at the oxygens in macromolecules, and therefore, it would be useful to directly observe these sites. Not only is it expensive, ^{17}O is available only in low isotopic enrichment. Numerous spectroscopists and enzymologists are interested in using high-enrichment ^{17}O. However, from the medical field there has been a far greater demand for ^{18}O and, hence, the fractionation columns in industry have been optimized for production of ^{18}O rather than ^{17}O. The columns at Los Alamos have not been used for years and would require a major overhaul. High enrichment is a very large cost for the companies. Laser-based methods for isotopic separation may help alleviate this problem. Direct observation of ^{17}O at the NHMFL will be a very attractive possibility on the widebore 900-MHz magnet and possibly on the 20-Tesla Magnex magnet—thus providing additional pressure to solve the current ^{17}O supply problem.

Several different deuteration schemes have been described (see Unkefer *et al.* in this volume), each of which has the potential to increase the molecular weight range of proteins that can be studied by NMR. Random partial deuteration has already proven a very effective tool for this purpose. In one example, detailed studies of a membrane protein in a micelle were made possible through partial deuteration. Deuteration would significantly increase Cα T_2 relaxation times, which would be very beneficial. Deuteration of the Cβ sites would simplify the spin systems and the spectra—making it possible to study larger proteins. Finally, chiral-specific deuteration would be a major help in generating stereospecific assignments, without which structures are not as well defined. These labeling strategies are also described in the paper by Unkefer *et al.* in this proceedings and discussed at more length in the report of the Peptides and Proteins Discussion Group.

Conclusions

Increases in sensitivity and resolution are the two primary driving forces for higher magnetic fields. Even sensitivity improvements of as little as 10% would be considered very significant by this NMR community. If major advances in sensitivity were possible, numerous scientific doors would be opened to new molecular systems. The quality of the structural constraints achieved by solution NMR would be considerably improved through an improvement in sensitivity. Resolution would also have a significant impact on studies of nucleic acids and polysaccharides. In addition, there are advantages in simplifying the spectra by going to higher fields. Strong couplings in solution NMR and odd-halves quadrupolar nuclei in solid state NMR are both simplified at higher fields.

Although solid state NMR spectroscopy does not require a magnet with an homogeneity of 10^{-3} ppm, it appears that solution NMR still requires homogeneity close to this specification. It is not yet clear what impact pulse-field gradients will have on this specification, but it is possible that solvent suppression by gradient techniques may lead to a relaxation of the homogeneity specification.

With the advent of multidimensional NMR, the use of stable isotopes has increased substantially. Now, with more sophisticated labeling schemes, it is clear that isotopes will continue to be a very useful tool for simplifying spin systems and providing enhanced sensitivity and selectivity. Since its inception, NMR has suffered from low sensitivity. Even with the speculation, discussed here, about substantial improvements in sensitivity, there will continue to be a need for additional sensitivity. Stereospecific assignments are becoming increasingly important for achieving high-resolution structural characterizations, and with chiral deuteration another avenue is opened for such assignments.

A year and one-half ago, we sent a letter to the NMR community asking for responses to some of the questions that have been the focus of this discussion. One of themes that appeared in the responses was that 600 MHz had substantial resolution and sensitivity advantages over 500 MHz, and in general, before their installation, the advantages of the higher fields had been underestimated and the disadvantages had been overestimated. Another theme was that not all high-resolution experiments would need high fields; in fact, certain types of experiments may be optimally performed at less than maximal field strength. In solid state NMR, this fact has been evident for many years. These considerations do not diminish the importance of high fields in solid state NMR—it simply means that it is possible to think of experiments that can be optimally done at less than maximal fields. We anticipate that in high-resolution macromolecular spectroscopy, the same situation will arise. Furthermore, the experience from solid state NMR also showed that some of the rationale for performing experiments at

lower fields evaporated as additional technologies were developed. For instance, magic angle spinning speeds increased in response to a need for greater speeds at higher fields. Before these spinners were developed, there was considerable discussion about whether faster spinners could be developed and how detrimental slow spinning would be for the higher field spectroscopy. Today faster spinners are readily available.

This group discussion took place partially in response to the letters from the community. We wish to develop as realistic an evaluation of the potential for higher fields as possible. The NSF-supported National High Magnetic Field Laboratory and the NIH-supported Los Alamos Stable Isotope Resource and the DOE-supported Environmental Molecular Sciences Laboratory at the Pacific Northwest Laboratory provide national facilities for our community. However, for these facilities to best serve the community, input is needed to help set priorities and provide a strong and coherent statement about their importance to the community. This discussion report documents that there are many advantages in pursuing higher magnetic fields, advanced technology for sensitivity enhancement and isotope labeling technology for a variety of purposes in macromolecular NMR.

Peptides and Proteins

Discussion Leaders

William W. Bachovchin and Clifford J. Unkefer

Participants

Jennifer Ashurst, Gerald T. Babcock, H. Barrett, John Bean, Walter J. Chazin, David P. Cistola, Timothy A. Cross, Bruce Dunlap, John S. Gounarides, Andrew P. Hansen, Griselda Hernandez, James Hicks, Janice Huwe, David M. LeMaster, David Live, Rodolfo A. Martinez, Gary McClosky, Raymond S. Norton, Van Qui Qian, Ponni Rajagopal, Joe Retford, Hanns Senn, Chou Tak Tan, Venkataraman Thanabal, Harry Vacek, Robert L. Van Etten, William H. Woodruff, Peter E. Wright, Erik R.P. Zuiderweg

Objective

Protein and peptide structural determination through modern spectroscopic techniques is critically dependent on our ability to label proteins and peptides with ^{2}H, ^{13}C, and ^{15}N. The potential to solve the next generation of structural biology problems will depend upon advances in labeling techniques and spectroscopic tools. This discussion group's objective was to identify labeling technologies that are crucial to future advances in structural biology.

Introduction

As a result of technological advances, magnetic resonance and vibrational spectroscopies have emerged as major players in biochemical mechanism and structural studies. These methods are dependent on our ability to introduce stable isotope labels into proteins of interest. For these techniques to realize their full potential, there must be significant advances in the methods used both to synthesize labeled amino acids and to incorporate them into proteins of interest. For example, modern strategies for making one-to-one NMR chemical shift assignments in macromolecules require the production of uniformly ^{13}C- and/or ^{15}N-labeled proteins. These proteins have generally been produced by culturing and overexpressing bacteria on D-[U-6-^{13}C]glucose and [^{15}N]ammonium salts. Because of the extensive post-translation processing of many eukaryotic proteins, they must be expressed using

cultured cells that have complex nutritional requirements, which makes labeling difficult at best. We must identify the most efficient cell culture expression systems and define their labeling requirements. In addition, if we are to extend NMR spectroscopy to the study of larger and larger biomolecules, it is critical to develop specific deuterium labeling strategies that will minimize the ^{1}H-^{1}H dipolar contribution to the resonance linewidth. More efficient methods for specifically incorporating deuterium, ^{13}C, and/or ^{15}N must also be developed. Because only limited resources are now devoted to the development of these synthetic and biosynthetic strategies, we must focus our efforts on the most important targets.

The intent of this discussion group was to help define important targets for today and for the future. The questions below were used to simulate discussion.

- What amino acid labeling patterns do we need that are not currently available? When is stereochemistry important?
- What labeled amino acids are so expensive that cost is a barrier to performing experiments? (Specifically discuss uniform as well as specific labeling.)
- Will specific ^{2}H labeling, used in combination with ^{13}C and/or ^{15}N labeling, be generally useful for studying larger biomolecular structures?
- If cheaper methods that employ bacterial expression systems were developed to produce uniformly ^{13}C- and ^{15}N-labeled proteins, could more structures be studied?
- What are the problems associated with expressing labeled proteins in cultured cells?
- What are the most important methodology problems that must be solved in the short term? What labeled compounds will be required?
- What are the most important scientific questions in peptide and protein chemistry? How could isotopes be used to help answer these questions?

The report that follows is divided into four sections, based on the general type of experiment for which the labeled amino acid will be most useful. This report emphasizes the importance of continued development of methods for synthesis of labeled amino acids. Because they all were considered very important, no attempt was made to prioritize specific recommendations.

Specifically Labeled Amino Acids for Vibrational Spectroscopy

The heavier isotopes (^{2}H, ^{13}C, ^{15}N, and ^{18}O) are critically important when applying vibrational spectroscopy to complex biological problems. Because they shift vibrational frequencies, isotopes facilitate spectral assignment. In addition, isotopes can be used to shift a particular vibrational transition away from the background in the spectrum of a complex biological sample. For example, *carboxyl*-^{13}C and

$\alpha^{15}N$ substitution in a single amide will shift its signal outside the envelope that contains the signals from the rest of the amides in a peptide or protein. Vibrational spectroscopy plays a particularly important role in characterizing the structure and function of metallo proteins. For these studies, it is crucial to label amino acids involved in metal-ligand bonds—particularly histidine and cysteine—as well as enzyme prosthetic groups involved in metal binding, such as heme. Finally, labeled tyrosine has been useful in characterizing a number of radical species in the active site of enzymes. Four specific needs were identified.

- Develop *carboxyl*-^{13}C- and $\alpha^{15}N$–labeled amino acids and their amine-blocked derivatives for peptide synthesis. In addition, methods of the specific incorporation of these amino acids into single sites in proteins will be important in the advancement of vibrational methods to study protein structure. Finally, methods for the specific ^{18}O enrichment of amide carboxyls in peptides and proteins are needed.
- Develop more efficient methods for synthesizing all ^{15}N-isotopomers of histidine.
- Develop methods for synthesizing all the ^{13}C and ^{2}H isotopomers of cysteine. In addition, cysteine highly enriched with ^{33}S or ^{34}S would be very useful. This goal may be limited by the availability of the sulfur isotopes.
- Use known methods to make all ^{13}C, ^{2}H, ^{15}N, and ^{17}O isotopomers of L-tyrosine available.

Producing Uniformly Labeled Proteins for Structural NMR Spectroscopy

As discussed above, uniformly ^{13}C- and ^{15}N-labeled proteins are required for structural NMR studies and are produced by using bacterial or cultured cell expression systems. The efficiency of protein expression is the greatest cost-variable in producing uniformly labeled proteins. There was general consensus that anything that could be done to standardize these expression systems would be important. Although the experienced members of the group felt that optimizing the expression of each protein was necessary and generalities were difficult, they agreed that the following steps would help a variety of researchers.

- Collect and distribute strains and vectors useful for expressing labeled proteins.
- Define minimal growth requirements for the bacterial- and cultured-cell expression systems by using a variety of potentially useful labeled carbon sources such as acetate, succinate, glucose, and glycerol.
- L-cysteine, L-glutamine, and L-tryptophan are essential for the growth of cells used for expressing mammalian proteins. These amino acids are limited in commercial medium supplements that are derived from hydrolysis of algal cells. Synthesizing these particular

amino acids labeled with ^{13}C and ^{15}N would greatly enhance the production of proteins in cultured cells. It would be most useful to enrich these amino acids uniformly with ^{13}C and ^{15}N. In the absence of uniform labeling, backbone-labeled L-[1,2-$^{13}C_2$, ^{15}N]cysteine, glutamine, and tryptophan would be helpful. Although backbone labeling would allow sequential assignment, the assignments could not be relayed to the sidechains.

- Of the possible carbon sources for growth of bacteria and cultured cells, only [U-2-^{13}C]acetate and D-[U-6-^{13}C]glucose are readily available. More economical methods for preparing other potential growth substrates such as glycerol would be useful.

Stereospecific Labeling Diastereotopic Groups in Amino Acids

Fifteen of the twenty common amino acids have a pair of β-methylene protons, which are diastereotopic and yield separate resonances that are difficult to assign. The problem of assignment can be overcome by stereoselectively substituting deuterium for these diastereotopic protons. Assignment of the β-methylene protons is particularly important because the scalar coupling between the α- and β-protons can be used to constrain the amino acid sidechain conformation.

Valine and leucine also contain diastereotopic methyl groups that yield separate resonances that are difficult to assign. Again, stereospecific isotopic substitution in the methyl groups would facilitate assignment of these resonances. The group participants felt strongly that "chirally deuterated" amino acids would help the NMR determination of protein structure but that these amino acids are not currently available to the community. They made several specific recommendations.

- Develop new methods for "chiral deuteration" at the β-methylene position of amino acids. Emphasis should be placed on synthetic routes that would allow incorporation of ^{13}C and ^{15}N amino into the amino acid in combination with stereospecific deuterium at the β-methylene position.
- Develop methods to differentially label the diastereotopic methyl groups in valine and leucine with ^{13}C as well as ^{2}H.

Specifically Labeled Amino Acids for Magnetic Resonance Techniques

Specifically labeled amino acids have been and will remain important for characterizing the role of active sites in proteins. In addition, specific labels are useful for characterizing sidechain dynamics and are required for solid state NMR studies of membrane peptides and proteins. Specific labeling will be needed to elucidate larger structures through NMR spectroscopy. In particular, specific labeling may be useful in characterizing interresidue interactions that stabilize the protein-protein and nucleic acid-protein complexes. Because there is potential need

for all ^{13}C, ^{15}N, and ^{2}H isotopomers for all of the amino acids, it is difficult to pinpoint specific targets. However, the group's specific recommendations are listed here.

- Efficient incorporation of many specifically labeled amino acids requires expression of proteins in amino-acid-requiring auxotrophs of *E. coli*. Because the literature describing auxotrophs is difficult to access, it would be useful if the Stable Isotope Resource at Los Alamos would collect available auxotrophs and define their minimal growth requirements.
- The Stable Isotope Resource should develop general strategies to synthesize labeled amino acids that could provide any isotopomer. Many of the required labeling patterns will be very specific to particular applications.
- Because many specifically labeled amino acids have limited applications, the isotope industry is unlikely to find them profitable. These amino acids and other labeled compounds form a class of compounds similar to that of "orphan drugs." Applications for these orphan-labeled compounds could provide answers to very important biomedical problems. The Stable Isotope Resource at Los Alamos can assume another important role by providing these orphan-labeled compounds.

Conclusion

Advances in magnetic resonance and vibrational spectroscopy make it possible to derive detailed structural information about biomolecular structures in solution. These techniques are critically dependent on the availability of labeled compounds. For example, NMR techniques used today to derive peptide and protein structures require uniformly ^{13}C- and ^{15}N-labeled samples that are derived biosynthetically from [U-6-^{13}C]glucose. These experiments are possible now because, during the 1970s, the National Stable Isotope Resource developed algal methods for producing [U-6-^{13}C]glucose. If NMR techniques are to be used to study larger proteins, we will need sophisticated labeling patterns in amino acids that employ a combination of ^{2}H, ^{13}C, and ^{15}N labeling. The availability of these specifically labeled amino acids requires a renewed investment in new methods for chemical synthesis of labeled amino acids. The development of new magnetic resonance or vibrational techniques to elucidate biomolecular structure will be seriously impeded if we do not see rapid progress in labeling technology. Investment in labeling chemistry is as important as investment in the development of advanced spectroscopic tools.

Carbohydrates/Nucleosides/RNA-DNA-Ligand Interactions

Discussion Leaders

R. Kaptein, B. McConnell, A.S. Serianni, and L. A. (Pete) Silks, III

Participants

David Ashburn, Kirsten Berghnmans, Patrick Bortmann, Zhouping Cai, Walter J. Chazin, Angel Garcia, P. T. Gilham, Roger A. Jones, Masatsune Kainosho, Ongkar Khalsa, Garry C. King, Jon Lapham, David Live, Suraj Manrao, Arthur Pardi, Yan Qiu Qian, Ponni Rajagopal, John SantaLucia, Jr., Jacob Schaefer, Hans Senn, Teresa Strzelecka, Lisa Theisen, Ignacio Tinoco, Thomas Walker, Peter E. Wright, Rullian Wu, Daniel P. Zimmer

Objective

Carbohydrate and nucleotide structural determination using modern spectroscopic techniques is dependent on our ability to label oligonucleotides and oligosaccharides with stable isotopes. Uniform ^{13}C and ^{15}N labeling of oligonucleotides is important to present-day efforts, which are focused on determining the structure of relatively small oligosaccharides and oligonucleotides, which form the elements of larger structures. Because of the relatively recent interest in three-dimensional structure of polynucleotides and polysaccharides, the development of techniques used to label them has lagged behind parallel techniques used to label peptides and proteins. Therefore, this group's discussion focused primarily on problems faced today in obtaining oligonucleotides labeled uniformly with ^{13}C and ^{15}N. The next generation of problems, including structures of larger nucleotide fragments and protein-oligonucleotide complexes, were not examined in detail. As is the case with proteins, the potential of solution NMR methods to solve larger structures will depend critically upon advanced labeling techniques, including site-specific labeling.

Introduction

It has become abundantly clear that the greater NMR spectral dispersion of ^{15}N, ^{13}C, and ^{2}H incorporated into molecules

considerably larger than 10 kDa has resulted in astonishing progress in the research of structure and dynamics—not only for proteins and nucleic acids but also for their complexes. Therefore, time was set aside during this meeting to examine the prospects for counseling and supporting the treasured sources of these isotopically labeled macromolecules.

The discussions for Carbohydrates/ Nucleosides and RNA-DNA Ligand Interactions were combined because the same participants were interested in both molecular systems. Twenty-four conference attendees came to the discussion fresh from that morning's powerful reminders, by Drs. Wright, Pardi, Tinoco, Jones, Kaptein, and Serianni, that no progress in molecular structure and dynamics research is conceivable without a reliable supply resource for isotopically labeled nucleic acids, carbohydrates, and proteins.

Discussion began with the question: Did any researcher have specific urgent needs that might lead to a possible consensus of desiderata? (This report does not attribute the many cogent and useful comments to individual contributors, with the exception of Dr. Tinoco, who expressed his wish to obtain oxygen at spin 1/2 and at high natural abundance.)

The discussion quickly turned to the acquisition of a cheap, uniformly labeled DNA or RNA from labeled nucleotide triphosphates. This in turn led to the question of whether there could be agreement on a common requirement for a given molecular type—so that a consortium of researchers with a common, focused interest would encourage the supplier to make large quantities of a limited offering at a feasible cost. A researcher operating within this agreed-upon need would "sign up" for a gram or two of labeled compound; this would create a total market of hundreds of grams for the source to supply on a cost-effective basis.

The next task was to select the molecular entity most generally needed. This issue was resolved by unanimous agreement on uniformly labeled monomers of RNA and DNA because the technology for making triphosphates and DNA and RNA polymers is well established and the latter would be prohibitively expensive on a custom basis. For DNA, phosphoramidites were preferred to NTPs because—new developments in Plenow DNA synthesis notwithstanding—the cost of enzymatic synthesis must include the expense of making unlabeled primers, which are often found to be unworkable. Such costly uncertainties would be eliminated with the chemical route, and therefore, the monomer form of choice for DNA was the phosphoramidite.

Alternatively, the uniformly labeled nucleoside could be obtained from the supplier for subsequent conversion to the triphosphates to be used for *in vitro* polymerization by the researcher, or "contracted out" for conversion to the phosphoramidite by an experienced outside laboratory. For polymeric RNA

precursors, the triphosphate (or the uniformly labeled ribonucleoside to be phosphorylated by the researcher) was the choice because polymerization with the T7 enzyme is well established and produces good yields.

At this point, we were reminded by our colleagues at ISOTEC that although the "economics of scale" dictated that multiple requests for a single labeled molecule might be cost-effective (for example, the cost of producing large amounts of uniformly labeled ribose might be low), subsequent steps (for example, adding a uniformly labeled nucleobase) would increase the cost geometrically, if not exponentially, and would ultimately result in a much more expensive product. Based on this consideration, it was generally agreed that an acceptable focus for collectively ordered isotopic compounds was the nucleoside level and, if possible, the phosphoramidites for deoxynucleosides and the triphosphates for the ribonucleosides.

A second, lower priority for labeled monomers was isotopic substitution at the ribose alone. The chemical coupling of ^{13}C-labeled ribose to pureness is possible, but difficult, and can be done routinely for the pyrimidines in small, but reasonable yields. The problem of spectral crowding of the uniformly labeled ribose carbons was raised and the desirability of placing the ^{13}C only on the spectrally disperse C1′ or C2′ provided a possible solution. A more attractive solution, which would both conserve isotope and reduce spectral crowding, is to provide ribose labeled at the 1′-3′-5′ or 2′-4′ positions. The development of this "alternate labeling" could prove a desirable option for large-scale requirements. For the present, labeling ribose carbons only is more costly than obtaining uniformly labeled nucleosides from bacterial cultures. The group's choice was, again, the uniformly labeled nucleoside with a mandate to develop special NMR techniques to deal with the problem of resolution. Moreover, a strategy was offered for producing the polymeric DNA or RNA as well as facilitating synthesis and reducing the problem of spectral overlap: by incorporating only one labeled monomer at a time. It would be just as easy to synthesize three or four DNAs or RNAs that have one labeled nucleotide as it would be to synthesize the polymer with all four.

The issue of deuterium substitution was discussed as a less attractive option for a commercial source of labeled compounds because its advantages would be offset by the same problem of broad resonances.

Other peripheral, but important ways of reducing costs for academic researchers were discussed, such as exploring the increasing collaboration between academic and industrial research. One possibility was the concept of recycling what was termed as "leftovers" from the bacterial production of labeled compounds. For example, after labeling nucleotides from isotopically fed bacteria, one researcher saves the labeled proteins for another

researcher, and *visa versa*. Our participating commercial representatives offered to coordinate this effort and to supply all interested parties with a list of counterparts nationwide who could supply these materials. However, it was pointed out that the isotopic enrichment of "leftovers" would probably be insufficient for obtaining good NOE data; an isotopically rich media is needed.

Alternatives to bacteria as sources of labeled compounds were considered. The attraction of using algae, with its requirements for simple, cheap, labeled nutrients (that is, CO_2 and ammonia), was offset by the perceived low quality of the products and the difficulty of extraction. The novel suggestion of using plants as sources of specific labeled compounds was raised, and bacteria were viewed as the logical choice. In terms of cost, using ^{13}C glycerol as a better alternative nutrient to ^{13}C glucose would depend on the supplier's commitment to a specific technology.

The issue of labeled oligosaccharides was an implicit beneficiary of the group's discussion of the problems of obtaining labeled DNA and RNA. It was also not explored in depth because "the field of oligosaccharide research is crying for good spectroscopists" and the market is not presently sufficient for a massive commitment by a commercial supplier. At the conclusion of the meeting, it was noted that the potential market for labeled RNA and DNA will soon rival and surpass that for labeled proteins.

An equally important consideration in generating a large demand from a group of scientists is the cooperation of the resource and the researcher in developing new techniques for isotope substitution. One example was the ^{15}N amino labeling of cytosine and the ^{13}C methyl labeling of thymine developed in one researcher's laboratory. Participants pointed out that presently, the biomedical community is more aggressive in seeking this cooperative development with the commercial supplier than the general science community is, and as financial beneficiary of this interaction, the commercial supplier could be helpful in initiating this cooperation.

Abstracts for Posters

Synthesis And Biosynthesis of ^{13}C-, ^{15}N-Labeled Deoxynucleosides Useful for Biomolecular Structural Determinations

David A. Ashburn, Katherine Garcia, John L. Hanners, Louis A. (Pete) Silks III, and Clifford J. Unkefer

Stable Isotope Resource
Chemical Science and Technology Division
Los Alamos National Laboratory
Los Alamos, NM 87545

Currently, there is a great emphasis on elucidating the structure, function, and dynamics of DNA. Much of the research involved in this study uses nuclear magnetic resonance (NMR) spectroscopy. Effective use of NMR spectroscopy for DNA molecules with mw > 10,000 requires stable isotope enrichment. We present strategies for site-specific isotopic labeling of the purine bases adenosine and guanosine and the biosynthesis of [U-^{13}C, ^{15}N]DNA from methylotrophic bacteria. With commercially available 6-chloropurine, an effective two-step route leads to 2′-deoxy-[*amino*-^{15}N]adenosine (dA). The resulting d[*amino*-^{15}N]A is used in a series of reactions to synthesize 2′-deoxy-[2-^{13}C, 1,*amino*-$^{15}N_2$]guanosine or any combination thereof. An improved biosynthesis of labeled DNA has been accomplished using *Methylobacterium extorquens* AS1. Each liter of growth medium contains 4 g of methanol to yield 1 g of lyophilized cells. As much as 200 mg of RNA per liter of culture has been obtained. We are currently developing large-scale isolation protocols. General synthetic pathways to oligomeric DNA will be presented.

Conformational Study of C8 Diazocine Turn Mimics Using $^{3}J_{CH}$ Coupling Constants with ^{13}C in Natural Abundance

John W. Bean, Jacques Briand, Joelle L. Burgess, and James F. Callahan

SmithKline Beecham Pharmaceuticals
Mail Code UW2940
P. O. Box 1539
King of Prussia, PA 19406-0939

The conformations of two diazocine turn mimics, which were later incorporated into GPIIb/IIIa peptide antagonists, were investigated using nuclear magnetic resonance techniques. The two compounds, methyl [2,5 -dioxo-3-(*S*)-(3-ω-tosyl-guanidino-propyl)-4-methyl-octahydro-1,4-diazocin-1-yl]acetate (**1**) and methyl [2,5 -dioxo-3-(*S*)-(3-ω–tosyl-guanidino-propyl)-octahydro-1,4-diazocin-1-yl]acetate (**2**), differ only in their substituent at the diazocine position 4 nitrogen, yet this substitution results in a marked difference in the affinity of the resulting analogs for the GPIIb/IIIa receptor. It was of interest to determine if the difference observed in the antagonistic potency between these analogs was related to constitutional or, perhaps, conformational differences.

The backbone conformations of these two molecules can be determined by measuring vicinal coupling constants along the trimethylene portion of the C8 ring backbone and by measuring interproton NOE intensities between the diazocine methine proton and the protons of the trimethylene group. For compound **1**, $^{3}J_{HH}$ values measured from a P.E.COSY spectrum and interproton distances calculated from ROESY buildup curves indicated the presence of a single C8 ring backbone conformation where the trimethylene bridge adopted a staggered conformation and the Hα1 and Hγ1 protons of the trimethylene group were 2.2 Å from the methine proton. For compound **2**, however, partial overlap of the central Hβ1 and Hβ2 protons made it impossible to measure $^{3}J_{HH}$ values from the P.E.COSY spectrum. We therefore used a ^{13}C-filtered TOCSY experiment to measure the $^{3}J_{CH}$ values in both compounds **1** and **2**. These heteronuclear vicinal coupling constants measured with ^{13}C in natural abundance in conjunction with measured interproton NOE intensities indicate that these compounds share a common C8 ring backbone conformation.

Structural Studies on an Internal Loop from a Hairpin Ribozyme

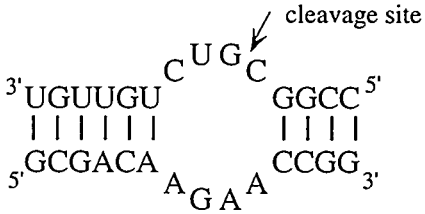
Zhouping Cai, John SantaLucia, Jr., and Ignacio Tinoco, Jr.

Department of Chemistry
and
Laboratory of Chemical Biodynamics
University of California
Berkeley, CA 94720

Ribozymes, RNA enzymes, catalyze site-specific RNA cleavage and ligation reactions. We are studying the three-dimensional structure of a hairpin ribozyme derived from the minus strand of tobacco ring spot virus satellite RNA [(-)sTRSV], which has been engineered to specifically cleave the HIV-1 RNA. The minimum structure for the catalytic reaction involves a 50-nucleotide ribozyme and a 14-nucleotide substrate. The proposed secondary structure of the ribozyme-substrate complex consists of four short helices separated by two internal loops. [J. O. Ojwang *et al.* (1992) *Proc. Natl. Acad. Sci. USA 89*, 10802].

The relatively large size (64-nucleotide) of the ribozyme-substrate complex presents formidable problems in solving the structure using NMR. Therefore we are studying smaller structural subunits of the complex. We are determining the high resolution structure of the symmetric internal loop involving the cleavage site and the flanking helices (Fig. 1). One strand of the internal loop was selectively ^{13}C-labeled at C8 of each purine and C6 of each pyrimidine. By using ^{13}C-edited two-dimensional NMR, the proton NOESY spectrum was greatly simplified. This allowed unambiguous sequential proton resonance assignments along each strand. Three-dimensional ^{1}H-^{13}C HMQC-NOESY was used to further facilitate resonance assignments.

We are also enzymatically synthesizing the entire 50-nucleotide ribozyme and will combine it with the ^{13}C-labeled substrate. Through comparison of the NOE connectivities of the labeled nucleotides from the internal loop alone with those from the entire complex, the differences between the two structures can be elucidated.



Symmetric internal loop from a hairpin ribozyme. The bottom strand was selectively ^{13}C-labeled.

Sequence-Specific ^{1}H, ^{13}C, and ^{15}N Resonance Assignments for Intestinal Fatty-Acid-Binding Protein Complexed with Palmitate (15.4 kDa)

Michael E. Hodsdon, James J. Toner, and *David P. Cistola*

Department of Biochemistry
and Molecular Biophysics
Washington University School of Medicine
St. Louis, MO 63110

Intestinal fatty-acid-binding protein (I-FABP) belongs to a family of soluble, cytoplasmic proteins that are thought to function in the intracellular transport and trafficking of polar lipids. Individual members of this protein family have distinct specificities and affinities for fatty acids, cholesterol, bile salts, and retinoids. We are comparing several retinol- and fatty-acid-binding proteins from intestine in order to define the factors that control molecular recognition in this family of proteins.

We have established sequential resonance assignments for uniformly $^{13}C/^{15}N$-enriched I-FABP complexed with perdeuterated palmitate at pH 7.2 and 37°C. The assignment strategy was similar to that introduced for calmodulin [Ikura, Kay and Bax (1990) *Biochemistry 29*, 4659]. We employed seven three-dimensional NMR experiments to establish scalar couplings between backbone and sidechain atoms. Backbone atoms were correlated using triple-resonance HNCO, HNCA, TOCSY-HMQC, HCACO, and HCA(CO)N experiments. Sidechain atoms were correlated using CC-TOCSY, HCCH-TOCSY, and TOCSY-HMQC. The correlations of peaks between three-dimensional spectra were established in a computer-assisted manner using NMR COMPASS (Molecular Simulations, Inc.).

Using this approach, ^{1}H, ^{13}C, and ^{15}N resonance assignments have been established for 120 of the 131 residues of I-FABP. For 18 residues, amide ^{1}H and ^{15}N resonances were unobservable, apparently because of the rapid exchange of amide protons with bulk water at pH 7.2. The missing amide protons correspond to distinct amino acid patterns in the protein sequence, which will be discussed.

During the assignment process, several sources of ambiguity in spin correlations were observed. To overcome this ambiguity, the additional *inter*-residue correlations often observed in the HNCA experiment were used as cross-checks for the sequential backbone assignments. In addition, the use of multiple spectra for identifying sidechain spin-systems was found critically important for the unambiguous assignment of this complex.

The sequential resonance assignments obtained for I-FABP are being used to study the structure and dynamics of the protein in solution—particularly its interaction with ligands.

Biosynthetic Incorporation of Telluromethionine into Dihydrofolate Reductase and Crystallographic Analysis of the Distribution of Tellurium Atoms in the Protein Molecule

Marci G. Kunkle, Krzysztof Lewinski, Jeffrey O. Boles, *R. Bruce Dunlap*, Jerome D. Odom, and Lukasz Lebioda

Department of Chemistry and Biochemistry
University of South Carolina
Columbia, SC 29208

Recent successes in crystallographic studies of proteins with methionine (Met) residues replaced with SeMet, pioneered by Hendrickson and coworkers, inspired us to replace Met with TeMet in *Escherichia coli* dihydrofolate reductase (DHFR). *E. coli* DHFR, which catalyzes the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, consists of 159 residues, 5 of which are Met. TeMet was incorporated into DHFR using the Met auxotroph, *E. coli* DL41, carrying the expression vector pWT8 with an IPTG inducible promoter and ampicillin resistance gene. The enzyme was purified by successive chromatography on Q-Sepharose and Phenyl Sepharose resins, yielding milligram quantities of homogeneous enzyme with a specific activity of 40 units/mg. TeMet DHFR exhibits kinetic properties similar to those of wt DHFR. Amino acid analysis indicated 3 authentic Met residues in TeMet DHFR, whereas atomic absorption spectroscopy detected 2 Te per protein molecule. Amino acid sequence analysis results suggested that only authentic Met was present in the first three Met positions (1, 16, and 20).

Crystals of Te-DHFR were grown in the presence of methotrexate from PEG 4000 and were isomorphous with wt-DHFR crystals grown from ethanol. Difference Fourier maps and restrained least-squares refinement show very little, if any, Te in the first three Met positions: Met^{1}, Met^{16}, and Met^{20}, whereas the occupancy of Te in positions 42 and 92 is 0.64. Apparently, the process of folding, subsequent purification, and crystallization select DHFR molecules with Te in Met^{42} and Met^{92}. Replacing Met with TeMet provides an internal probe that should facilitate structural and mechanistic studies of proteins.

This work was supported by NIH Grant GM 42907.

Stereospecific Assignments of Glycine in Proteins by Stereospecific Deuteration and ^{15}N Labeling

Andrew P. Hansen, Robert W. Curley, Jr., Michael J. Panigot, and Stephen W. Fesik

Pharmaceutical Discovery Division
Abbott Laboratories
Abbott Park, IL 60064
and
College of Pharmacy
Ohio State University
Columbus, OH 43210

Stereospecific assignments are important for accurately determining the three-dimensional structures of proteins through the use of multidimensional NMR techniques. It is especially important to stereospecifically assign the glycine α-protons in proteins because of the potential for different backbone conformations of this residue. These stereospecific assignments are critical for interpreting the $^3J_{NH,\alpha H}$ coupling constants and NOEs involving the glycine α-protons that determine the conformation of this part of the protein. However, it is often difficult to unambiguously obtain the stereospecific assignments for glycine residues by using only NOE data. In this poster we present a method for unambiguous, stereospecific assignment of the α-protons of glycine residues. This method involves synthesis of stereospecifically deuterated and ^{15}N-labeled Gly using a slightly modified procedure originally described by Woodard and coworkers for the stereoselective deuteration of glycine [Ramalingam *et al.*, (1988) *Tetrahedron 44*, 5597-5604]. The stereospecifically deuterated and ^{15}N-labeled Gly has been incorporated into recombinant proteins expressed in both bacterial systems (FKBP) and mammalian cells (u-PA). Two- and three-dimensional isotope-filtered and isotope-edited NMR experiments were used to obtain the stereospecific assignments of the glycine α-protons for these proteins.

Uniform ^{15}N- and $^{15}N/^{13}C$-Labeling of Proteins in Mammalian Cells and Solution Structure of the Amino Terminal Fragment of u-PA

Andrew P. Hansen, Andrew M. Petros, Robert P. Meadows, Andrew P. Mazar, David G. Nettesheim, Terry M. Pederson, and Stephen W. Fesik

Pharmaceutical Discovery Division
Abbott Laboratories
Abbott Park, IL 60064

Urokinase-type plasminogen activator (u-PA) is a 54-kDa glycoprotein that catalyzes the conversion of plasminogen to plasmin, a broad-specificity protease responsible for the degradation of fibrin clots and extracellular matrix components. The u-PA protein consists of three individual modules: a growth factor domain (GFD), a kringle, and a serine protease domain. The amino terminal fragment (ATF) includes the GFD—responsible for u-PA binding to its receptor—and the kringle domains. This protein was expressed and uniformly ^{15}N- and $^{15}N/^{13}C$- labeled in mammalian cells by methods that will be described. In addition, we present the three-dimensional structure of ATF that was derived from 1299 NOE-derived distance restraints along with ϕ angle and hydrogen bonding restraints. Although the individual domains in the structures were highly converged, the two domains are structurally independent. The overall structures of the individual domains are very similar to the structures of homologous proteins. However, important structural differences between the growth factor domain of u-PA and other homologous proteins were observed in the region that has been implicated in binding the urokinase receptor. These results may explain, in part, why other growth factors show no appreciable affinity for the urokinase receptor.

NMR Studies of Two Spliced Leader RNAs Using Isotope Labeling

Jon Lapham and Donald M. Crothers

Chemistry Department
Yale University
New Haven, CT 06520

Spliced leader RNAs are a class of RNA molecules (<200 nts) involved in the *trans* splicing of messenger RNA found in trypanosomes, nematodes, and other lower eukaryotes. The spliced leader RNA from the trypanosome *Leptomonas Collosoma* exists in two alternate structural forms with similar thermal stabilities [Lecuyer and Crothers (1993) *Biochem.*]. The 54 nucleotides on the 5′ end of the SL molecule is structurally independent from the 3′ half of the RNA, and displays the two structural forms. Furthermore, the favored of the two structures was shown to contain anomalous nuclease sensitivity and thermal stability features, which suggests that there may be tertiary interactions between the splice site and other nucleotides in the 5′ end.

Multidimensional NMR studies are underway to elucidate the structural elements present in the SL RNAs that give rise to their physical properties. Two spliced leader sequences have been studied. The first, the 54 nucleotides on the 5′ end of the *L. Collosoma* sequence, was selected because of earlier studies in our laboratory. The second sequence is the 5′ end of the trypanosome *Crithidia Fasciculata,* which was chosen because of its greater sequence homology to other SL sequences. Given the complexity of the NMR spectra for RNA molecules of this size, we have incorporated $^{15}N/^{13}C$-labeled nucleotides into the RNA.

One of the techniques we have developed to simplify the spectra of these RNA molecules is isotope labeling of specific regions of the RNA. This has been especially helpful in assigning the secondary structure of molecules that may be able to adopt multiple conformations. Using this technique one can examine a part of the molecule without spectral interference from the unlabeled portion. We hope this approach will promote an avenue for studying the structure of larger RNAs in their native surroundings.

Application of Heteronuclear Couplings to Conformational Analysis of Oligonucleotides

Guang Zhu,[1] *David Live,*[2] and Adrian Bax[3]

[1]Department of Electrical Engineering
University of Maryland
College Park, MD 20874
[2]Cellular Biochemistry and Biophysics Program
Memorial Sloan-Kettering Cancer Center
New York, NY 10021
[3]Laboratory of Chemical Physics
NIDDK National Institutes of Health
Bethesda, MD 20892

The value of vicinal coupling constants extracted from NMR spectra in deducing torsion angles for conformational analysis is well recognized. Due to the abundance of protons, their couplings have been most widely used. In many instances, couplings between protons and other nuclei may be a valuable complement to proton-proton couplings or, in some instances, may be the only coupling available to characterize the torsion angle about a bond. Recently, heteronuclear couplings have been used to great benefit in studies of isotopically enriched proteins, and this general approach has been extended to peptides at natural abundance. The possibility of using this approach to study oligonucleotides is also attractive but has not as yet been widely exploited. With the development of strategies for labeling such molecules, particularly RNAs, this may become an important component in conformational analysis. For DNA, labeling is less accessible, but sufficient quantities of unlabeled material are readily available for measuring these couplings at natural abundance. We chose several DNA systems to explore the usefulness of heteronuclear couplings in addressing the sugar conformation and the glycosidic torsion angle. Intensities of cross peaks in long-range HMQC experiments can be related to the couplings. Crosspeaks involving H1′ and C1′ atoms have been emphasized because of the superior shift dispersion at these positions between sugar protons and carbon atoms. Results will be shown for the self-complementary Dickerson duplex dodecamer sequence d(CGCGAATTCGCG) and for d(GGTCGG), which dimerizes to form a G-tetrad structure incorporating both *syn* and *anti* base orientations. The couplings provide a clear discrimination between presence of C3′-endo and C2′-endo conformations of the sugars and *syn* and *anti* bases arrangements. Intrasugar coupling values are consistent with a heteronuclear Karplus relation proposed by van Beuzekom *et al.* [(1990) *Magn. Reson. Chem. 28,* 68-74] that includes consideration of electronegativity effects. In addition, the relationship derived by Davies *et al.* [(1985) *Magn. Reson. Chem. 23,* 72-77] for couplings across the glycosidic bond of pyrimidines is applicable to purine bases found in the G-tetrad. Analysis of these couplings should enable more accurate characterization of oligonucleotide structures. A more complete discussion of these will appear soon [(1994) *J. Am. Chem. Soc. 116,* 8370-71].

Structural Studies on Leukaemia Inhibitory Factor

Raymond S. Norton,[1] **Till Maurer**,[1] **David K. Smith**,[1] **and Nicos A. Nicola**[2]

[1]NMR Laboratory
Biomolecular Research Institute
381 Royal Parade
Parkville 3052, Australia
[2]Walter and Eliza Hall
Institute of Medical Research
P.O. Royal Melbourne Hospital
Melbourne 3050, Australia

Leukaemia Inhibitory Factor (LIF) is a pleiotropic cytokine that acts on a wide range of target cells, including megakaryocytes, osteoblasts, hepatocytes, adipocytes, neurons, embryonic stem cells, and primordial germ cells. Many of its activities are shared with other cytokines, particularly interleukin-6, oncostatin-M, ciliary neurotrophic factor, and granulocyte colony-stimulating factor (G-CSF). Although secreted *in vivo* as a glycoprotein, nonglycosylated recombinant protein expressed in *E. coli* is fully active and has been used in our nuclear magnetic resonance (NMR) studies of the three-dimensional structure and structure-function relationships of LIF.

With 180 amino acids and a molecular mass of about 20 kDa, LIF is too large for direct structure determination by two-dimensional and three-dimensional ^{1}H NMR. It is necessary to label the protein with the stable isotopes ^{15}N and ^{13}C and employ heteronuclear three-dimensional NMR in order to resolve and interpret the spectral information required for three-dimensional structure determination [Smith, J.J., Redfield, C., Boyd, J., Lawrence, G.M.P., Edwards, R.G., Smith, R.A.G., and Dobson, C.M. (1992) *J. Mol. Biol. 224*, 899-904; Powers, R., Garrett, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M., and Clore, G.M. (1993) *Biochemistry 32*, 6744-6762].

This work has been undertaken with both human LIF and a mouse-human chimaera that binds to the human LIF receptor with the same affinity as the human protein and yet expresses in *E. coli* at much higher levels. Sequence-specific resonance assignments and secondary structure elements for these proteins will be presented and progress towards determination of their three-dimensional structures described.

This is a contribution from the Cooperative Research Centre for Cellular Growth Factors

Measurement of CO_2 and N_2O at Nanomolar Amounts Using Continuous-Flow Isotope-Ratio Mass Spectrometry (CF-IRMS)

Ayyub Patel,[2] **Shaun Downie,**[2] **Elizabeth Webster,**[1] **David W. Hopkins,**[1] **and Michael J. Rennie**[2]

[1]Department of Biological Sciences
[2]Department of Anatomy and Physiology
University of Dundee
Dundee, DD1 4HN, Scotland, U.K.

We are currently developing methods using Continuous Flow Isotope Ratio Mass Spectrometry (CF-IRMS) in conjunction with a thermal desorption purification unit to measure nanomolar levels of CO_2 and N_2O. Samples of the pure gases diluted in He/air and transferred to septum capped Exetainers (Labco) provided a simple means to investigate the technique. We analyzed CO_2 at natural abundance in the concentration range 50 to 5 nmoles and N_2O at two concentrations between 25 and 5 nmoles. The technique was then used to measure CO_2 (natural abundance and ^{13}C-labeled) generated from the ninhydrin reaction [Read, W.W., Read, M.A., Rennie, M.J., Griggs, R.C., and Halliday, D. (1984) *Biomed. Mass Spectrom. 11,* 348-352]. The results are summarized in the table below; values are expressed in delta ^{13}C notation relative to Pee Dee Belemnite.

The data show that, provided care is taken to minimize or eliminate sources of contamination (air leaks, etc.), CF-IRMS coupled with a thermal desorption unit permits measurement of ^{13}C enrichment in much smaller amounts of isolated amino acids than has been possible until now. The new methodology, including thermal desorption, should allow stable-isotope investigations on much smaller samples than are possible with other currently available techniques—while maintaining high precision.

We are grateful to The Wellcome Trust and Europa Scientific Ltd for their continued support. EAW and DWH acknowledge financial support from the NERC under the TIGER programme.

	50 nmol C/N		20 nmol C		10 nmol C/N		5 nmol C	
Sample	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CO_2 gas	-40.5	1.9	-37.8	2.0	-	-	-26.0	1.7
^{13}C-leucine	26.2	1.25	20.7	1.18	19.25	1.5	19.5	2.47
N_2O gas	-33.4	0.3	-	-	-27.7	2.7	-	-
N_2O from soil	-52.6	1.43	-	-	-57.1	3.2	-	-

Mechanism of Phosphoryl Transfer and Protein-Protein Interaction in the PTS System—An NMR Study

Ponni Rajagopal and Rachel E. Klevit

Department of Biochemistry
University of Washington
Seattle, WA 98195

HPr and Enzyme IIA^{Glc} are two of the components of the bacterial PTS (phosphoenolpyruvate: sugar phosphotranferase system) and are involved in the phosphorylation and concomitant translocation of sugars across the membrane. These PTS protein complexes also regulate sugar transport. HPr, phosphorylated at a histidine N1 site by Enzyme I and phosphoenol pyruvate, transfers the phosphoryl group to a histidine N3 position in Enzyme IIA^{Glc}. HPrs from Gram-positive bacteria undergo regulatory phosphorylation at Ser^{46}, whereby phosphorylation of the histidine residue is inhibited. Conversely, histidine phosphorylation inhibits phosphorylation at Ser^{46}. HPrs from Gram-negative bacteria possess a serine residue at position 46, but do not undergo regulatory phosphorylation.

Several questions have formed the major focus of our work:
Why does histidine phosphorylation inhibit phosphorylation at Ser^{46}?
What are the specific factors that play a role in HPr-Enzyme IIA recognition?

HPr forms an open-faced sandwich structure with a four-strand ß-sheet and 2 to 3 helices lying on top of the sheet. The active-site histidine and Ser^{46} occur in conformationally flexible regions. P-His-HPr from the Gram-positive bacterium *Bacillus subtilus* has been investigated by both homonuclear and heteronuclear two-dimensional and three-dimensional NMR experiments using an *in-situ* enzymatic regeneration system to maintain a constant level of P-His-HPr. The results show that localized conformational changes occur in the vicinity of the active-site histidine and also near Ser^{46}. HPr-Enzyme IIA^{Glc} complexes from both *Bacillus subtilis* and Gram-negative *Escherichia coli* were also studied by a variety of ^{15}N-edited two-dimensional NMR experiments, which were performed on uniformly ^{15}N-labeled HPr complexed to unlabeled Enzyme IIA^{Glc}. The complex is in fast exchange with a molecular weight of about 27 kDa. The focus of our work is to assess the changes undergone by HPr (the smaller of the two components), and so all the experiments were performed with excess Enzyme IIA present in the system. The binding site of HPr in complex with Enzyme IIA^{Glc} was mapped by monitoring chemical shift changes of the amide protons in HPr. To further characterize the changes undergone by HPr on binding, ^{15}N T_1s and T_2s, ^{1}H T_1s, amide exchange rates and NOE intensities were determined for both free and bound HPr.

Synthesis and Applications of Selectively ^{13}C-labeled RNA

John SantaLucia, Jr., L.X. Shen, H. Lewis, Zhouping Cai, and Ignacio Tinoco, Jr.

Department of Chemistry and
Laboratory of Chemical Biodynamics
University of California
Berkeley, CA 94720

Spectral overlap is a substantial problem in NMR studies of RNA molecules >30 nucleotides. To overcome this difficulty, we synthesized selectively ^{13}C-labeled RNAs and adapted several isotope-edited two- and three-dimensional NMR experiments originally developed for protein studies.

We optimized protocols for synthesis of multi-gram quantities of [6-^{13}C] CTP, [6-^{13}C] UTP, [8-^{13}C] ATP, and [8-^{13}C] GTP using a combination of synthetic organic and enzymatic methods. [6-^{13}C]Uracil is prepared in 40 to 50% yield from ^{13}C-cyanide in two steps. Using acetyl-tribenzoyl-ribose and standard chemistry, uracil is then attached to the sugar (90% yield). The tribenzoyl-uridine intermediate is converted into uridine or cytidine quantitatively, depending on the deblocking protocol. Labeled purines are synthesized using simple pyrimidine precursors and reacting with ^{13}C-formic acid (80% yield). Purine nucleosides are then synthesized using uridine phosphorylase and purine nucleoside phosphorylase. The nucleosides were converted to NMPs by treatment with $POCl_3$ in triethylphosphate.

We converted NMPs to NTPs by standard enzymatic methods. Selectively labeled RNAs were synthesized by run-off transcription using ^{13}C-labeled NTPs.

Several different strategies help solve overlap problems in larger RNAs. Isotope-edited two-dimensional NMR experiments such as ω1-1/2 X-filtered NOESY simplify NMR spectra by dividing the normal NOESY spectrum into two subspectra—one involving NOEs from protons bound to ^{12}C and one from protons bound to ^{13}C. For example, we labeled A and U residues of a 34-nucleotide pseudoknot, and the ^{12}C subspectrum of the 1/2 X-filtered NOESY contained NOEs only from G and C residues (along with adenine 2H); the ^{13}C subspectrum contained NOEs only from A and U residues. Each subspectrum has less overlap than the NOESY of an unlabeled sample; the editing strategy allows each resonance to be identified by residue type (A, C, G, or U). Another strategy takes advantage of chemical shift dispersion of ^{13}C resonances. In two- and three-dimensional HMQC-NOESY experiments, the first spreads out NOE information according to ^{13}C chemical shift in the first dimension (unlike the usual NOESY experiment, which uses two proton frequency axes); the second edits the normal NOESY spectrum according to ^{13}C chemical shift in a third dimension, which greatly resolves overlap, but at $>\sqrt{2}$ loss in sensitivity.

In addition, it is possible to obtain new structural information such as *J*-couplings involving ^{13}C. In nondecoupled NMR experiments, E. COSY-type crosspeak fine structure permits measurement of small heteronuclear *J*-couplings.

Synthesis and Applications of ^{13}C Glycerol

Emily Stocking, **Ongkar Khalsa, Rodolfo A. Martinez, and Louis A. (Pete) Silks, III**

NIH Stable Isotope Resource
Chemical Science and Technology Division
Los Alamos National Laboratory
Los Alamos, NM 87545

Due in part to the use of labeled glycerol for the ^{13}C enrichment of biomolecules, we are currently developing new synthetic routes to various isotopomers of glycerol. Judging from our experience, traditional methods of glycerol synthesis are not easily adapted for isotopic enrichment and/or have poor overall yields (12 to 15%). Furthermore, the use of glycerol for enrichment can be prohibitively expensive and its availability is limited by the level of demand. We are presently developing a short *de novo* synthesis of [U-$^{13}C_3$]glycerol from carbon dioxide (~53% overall yield for four steps) and are examining the feasibility of synthesizing site-specific ^{13}C-labeled glycerol and dihydroxyacetone (DHA) from labeled methanol and carbon dioxide.

One application of ^{13}C glycerol we have examined is enzymatic conversion of [U-$^{13}C_3$]glycerol to glyceraldehyde-3-monophosphate or dihydroxyacetone monophosphate (DHAP) with yields ranging from 25 to 50% (as determined by NMR spectroscopy). We are also pursuing the chemical conversion of ^{13}C-labeled DHA to DHAP. We are especially interested in ^{13}C-labeled DHAP because we are investigating its use as a chemoenzymatic precursor for both labeled 2-deoxyribose and 2-deoxyribonucleic acids.

NER PROTEIN OF PHAGE MU: ASSIGNMENTS USING $^{13}C/^{15}N$-LABELED PROTEIN

Teresa Strzelecka, **Angela M. Gronenborn, and G. Marius Clore**

Laboratory of Chemical Physics
NIDDK, National Institutes of Health
Bethesda, MD 20892

The Ner protein is a small (74-amino acid) DNA-binding protein that regulates a switch between the lysogenic and lytic stages of phage Mu. It inhibits expression of the C repressor gene and down-regulates its own expression.

Two-dimensional NMR experiments on uniformly ^{15}N-labeled protein provided most of the backbone and some of the sidechain proton assignments [A. Gronenborn *et al.* (1989) *Biochemistry 28,* 5081]. The secondary structure determination using two-dimensional NOESY experiments showed that Ner consists of five α-helices. However, because most of the sidechain protons could not be assigned, the full structure was not determined.

Using uniformly $^{13}C/^{15}N$-labeled Ner and a set of three-dimensional experiments [A. Bax and S. Grzesiek (1993) *Acc. Chem. Res. 26,* 131], we were able to assign all of the backbone and 98% of the sidechain protons. In particular, the CBCANH and CBCA(CO)NH experiments were used to sequentially assign the Cα and Cβ resonances; the HCCH-CTOCSY and HCCH-COSY were used to assign sidechain carbon and proton resonances.

Mechanistic Studies of 3-Deoxy-D-*manno*-2-octulosonic Acid 8-Phosphate Synthase

Garry D. Dotson and *Ronald W. Woodard*

Interdepartmental Program in Medicinal Chemistry
University of Michigan
College of Pharmacy
Ann Arbor, MI 48109-1065

The enzyme 3-deoxy-D-*manno*-octulosonic acid 8-phosphate synthase [EC 4.1.2.16] (KDO 8-P synthase) catalyses the condensation of arabinose 5-phosphate (A 5-P) with phosphoenolpyruvate (PEP) to give the unique eight-carbon acidic sugar 3-deoxy-D-*manno*-octulosonic acid 8-phosphate (KDO 8-P) found only in gram-negative bacteria and required for lipid A maturation and cellular growth. The *E. coli* gene *kds*A that encodes KDO 8-P synthase has been amplified by standard PCR methodologies. The synthetic gene, subcloned into the expression vector pT7-7 was used to infect *E. coli* BL 21 (DE 3). Purification of crude supernatant from this transformant on Q Sepharose yields >200 mg of near-homogeneous KDO 8-P synthase per liter of cell culture. To explore the mechanism of KDO 8-P synthase, we prepared (*E*)- and (*Z*)-[3-^{2}H]PEP, [2-^{13}C]PEP, and [2-^{13}C,^{18}O]PEP chemically from the appropriately labeled 3-bromopyruvates by reaction with trimethylphosphite under Perkow reaction conditions. Our ^{1}H-NMR analysis of the stereochemistry at C3 of the KDO 8-Ps, obtained by separate incubation of (*E*)- and (*Z*)-[3-^{2}H]PEP with A 5-P in the presence of KDO 8-P synthase, demonstrated that the reaction is stereospecific with respect to both the C3 of PEP and the C1 carbonyl of A 5-P. (*Z*)-[3-^{2}H]PEP gave predominantly (3*S*)-[3-^{2}H]KDO 8-P and (*E*)-[3-^{2}H]PEP gave predominantly (3*R*)-[3-^{2}H]KDO-8P, which indicates condensation of the *si* face of PEP upon the *re* face of A 5-P—an orientation analogous to that seen with the similar aldehyde lyase DAH 7-P synthase. The fate of the enolic oxygen of [2-^{13}C,^{18}O]PEP, during the course of the KDO 8-P synthase-catalyzed reaction as monitored by both ^{13}C- and ^{31}P-NMR spectroscopy demonstrated that the inorganic phosphate (Pi) and not the KDO 8-P contained the ^{18}O. We also analyzed the trimethyl ester derivative of the Pi by mass spectral analysis. Incorporation of ^{18}O into the C2 of KDO 8-P, from solvent ($H_2{}^{18}O$) was observed by ^{13}C-NMR during the incubation of [2-^{13}C]PEP with A 5-P in $H_2{}^{18}O$ in the presence of KDO 8-P synthase. [2-^{13}C]KDO 8-P, prepared enzymatically, was used to determine the extent of nonenzymatic incorporation of ^{18}O into the C2 position of KDO 8-P. On the basis of these findings, we will present a new mechanism for this unusual aldol condensation.

NMR Studies of Bent DNA Using ^{13}C-Enriched Samples

Daniel P. Zimmer and Donald M. Crothers

Department of Molecular Biophysics
and Biochemistry
Department of Chemistry
Yale University
New Haven, CT 06511

Bending of the DNA double helix can be brought about by introducing runs of adenines (A-tracts) in phase with the helical repeat of the DNA. The requirements for bending of DNA by A-tracts are that the length of the A-tract be greater than 3 base pairs and that the A-tracts must be in phase with the helical repeat (every 10 or 11 bp). Other factors, such as the number of adenines in the run, flanking sequences, and whether the A-tracts are phased with respect to the 5′ A or the 3′A, have effects upon the degree of bending as assayed by electrophoretic mobility on native polyacrylamide gels.

There are a number of models for bending A-tract DNA. The junction-bending model postulates that the structure of A-tracts is similar to the fiber diffraction structure of poly A, in which there is a significant degree of base pair tilt with respect to the helix axis. In this model, bending occurs at the junction between the A-tract and the B-form helix to allow favorable stacking interactions to occur. The bend of the helix could arise as a result of some other perturbation of B-form DNA by A-tracts, such as propeller twist; bending also could be due to a combination of factors. Our goal is to find the structural features of A-tracts responsible for bending of the helix by performing NMR on oligonucleotides containing A-tracts to obtain higher resolution structural data.

One of the problems encountered in NMR structure determination of nucleic acids and other macromolecules is the assignment of resonances to nuclei. This procedure can be greatly facilitated through the use of ^{13}C-enriched nucleic acid samples. We are developing a technique for the enzymatic synthesis of labeled DNA for NMR. The technique we are developing is similar to RNA labeling techniques already in use. The technique involves growth of methylotrophic bacteria on $^{13}CH_3OH$, digestion of the DNA to deoxynucleotides, enzymatically charging the monophosphates to triphosphates, and use of the triphosphates in a novel DNA synthesis reaction. We have developed this technique to make high resolution structure determination of apparently bent DNAs easier. In addition to facilitating assignment, we hope that labeling of oligonucleotides will provide structural information—specifically some of the dihedral angles in the deoxyribose-phosphate backbone. We also plan to use ^{13}C-enriched DNA in NMR studies of the DNA in protein-DNA complexes.

Heteronuclear Cross-Polarization in Multinuclear Multidimensional NMR: Prospects for Triple-Resonance CP

Ananya Majumdar and *Erik R.P. Zuiderweg*

Biophysics Research Division
University of Michigan
Ann Arbor, MI 48109

Heteronuclear multiple-pulse-based Cross Polarization (HECP) [Zuiderweg (1990) *J. Magnet. Reson. 89*, 533) between scalar coupled spins is gaining an important role in high-resolution multidimensional NMR of isotopically labeled biomolecules, especially in experiments involving net magnetization transfer. It has generally been observed that in these situations, the performance of HECP is superior to that of INEPT-based sequences. In particular, HECP-based three-dimensional HCCH spectroscopy is more efficient than the INEPT version of the same experiment [Majumdar *et al.*, (1993) *J. Biomol. NMR 3*, 387]. Differences in sensitivity have been intuitively attributed to relaxation effects and technical factors such as radiofrequency (rf) inhomogeneity.

We present theoretical analyses and computer simulations to probe the effects of these factors. Relaxation effects were treated phenomenologically; we found that relaxation differences are relatively small (up to 25%) between pulsed-free-precession (INEPT) and HECP—although always in favor of HECP. We explored the rf effects by employing a Gaussian distribution of rf amplitude over sample volume. We found that inhomogeneity effects significantly favor HECP over INEPT, especially under conditions of "matched" inhomogeneity in the two rf coils.

The differences in favor of HECP indicate that an extension of HECP to triple resonance experiments (TRCP) in I -> S -> Q net transfers might yield better results relative to analogous INEPT-based net transfers. We theoretically analyze the possibilities of TRCP and find that transfer functions are critically dependent on the ratio J_{IS}/J_{SQ}. When J_{IS} equals J_{SQ}, we find that 100% transfer is possible for truly simultaneous TRCP and this transfer is obtained in a time $1.41/J$. The TRCP time requirement compares favorably with optimally concatenated INEPT-transfers, where net transfer I -> S -> Q is complete at $1.5/J$. When J_{IS} and J_{SQ} are unequal, the efficiency of truly simultaneous TRCP drops dramatically. However, essentially full I -> S -> Q CP transfer can be obtained when it is performed as HECP[IS](τ_1) - TRCP[ISQ](τ_2) - HECP[SQ](τ_3), where τ_1, τ_2, and τ_3 are the different durations of the respective CP steps. We explored these parameters as a function of J_{IS} and J_{SQ} and found that, for most practically viable cases, τ_1 or τ_3 must be set to zero. In all cases, we find that $\tau_1+\tau_2+\tau_3$ is equal to or up to 20% shorter than the time required for optimally concatenated INEPT-based net transfers. In addition to the theoretical considerations, we present experimental verifications of TRCP.

List of Conference Participants

Ashburn, David A.
East Tennessee State University
2822 West Walnut Cottage 15
Johnson City, TN 37604

Ashurst, Jennifer
European Molecular Biology Laboratory
EMBL, Memerhofstrabe 1 69117
Heidelberg GERMANY
49 622 138 7548

Avona, Vincent
ISOTEC, Inc.
3858 Benner Road
Miamisburg, OH 45342
(513) 859-1808

Babcock, Gerald T.
Michigan State University
320 Chemistry Building East Lansing, MI 48824
(517) 355-9715

Bachovchin, William W.
Tufts University
Department of Biochemistry
136 Harrison Avenue
Boston, MA 02111
(617) 956-6881

Barr, Mary
Los Alamos National Laboratory
Material Science and Technology Division
MS E501
Los Alamos, NM 87545
(505) 667-7991

Barrett, Hugh
University of Washington
Center for Bioengineering FL-20
Seattle, WA 98195
(206) 685-2009

Batchelder, Lynne
Cambridge Isotope Laboratories
50 Frontage Road
Andover, MA 01810
(508) 749-8000

Bax, Adrian
National Institutes of Health
Chemical Physics Bldg. 2, Rm. 109
Bethesda, MD 20892
(301) 496-2848

Bean, John
SmithKline Beecham Pharmaceuticals
UW2940
P. O. Box 1539
King of Prussia, PA 19406
(610) 270-6670

Bear, David
University of New Mexico
School of Medicine
Department of Cell Biology
900 Camino de Salud, N. E.
Albuquerque, NM 87131-5226
(505) 277-8520

Berghmans, Kirsten
University of New Mexico
1201 Yale N. E.
Albuquerque, NM 87131
(505) 277-2062

Bortmann, Patrick
University of California
Molecular Biology Institute
405 Hilgard Avenue
Los Angeles, CA 90024
(310) 825-9232

Bradbury, E. Morton
Los Alamos National Laboratory
Life Sciences Division, MS M881
Los Alamos, NM 87545
(505) 667-2690

Britt, Mark
Los Alamos National Laboratory
Chemical Science and Technology Division
MS C345
Los Alamos, NM 87545
(505) 667-6782

Cai, Zhouping
University of California, Berkeley
c/o Prof. I. Tinoco
Chemistry Department
Berkeley, CA 94720
(510) 642-1440

Calvet, Alain
Institute De Recherche Jouveina
3 A 9 rue de la loge-BP 100
94265 Fresnes Cedex, FRANCE
31 1 40 96 74 66

Carney, Martin
Ontario Hydro
2700 Lakeshore Road
West Mississauge, Ontario L5J 1K3
CANADA
(905) 855-4726

Castellino, Stephen
Rhone-Poulenc Ag
P. O. Box 1204 2
T. W. Alexander Drive
RTP, NC 27709
(919) 549-2580

Catasti,Paolo
Los Alamos National Laboratory
Theoretical Division, MS K710
Los Alamos, NM 87545
(505) 665-3807

Chazin, Walter J.
The Scripps Research Institute
10666 North Torrey Pines Road
Mail Drop MB2 LaJolla, CA 92037
(619) 554-9860

Cistola, David P.
Washington University
School of Medicine
Campus Box 8231
660 South Euclid Avenue
(St. Louis, MO 63110
314) 362-4382

Cross, Timothy
Florida State University
Department of Chemistry
Tallahassee, FL 32306
(904) 644-0917

Demgar, Betty
ISOTEC, Inc.
3858 Benner Road
Miamisburg, OH 45342
(513) 859-1808

Drobny, Gary
University of Washington
Department of Chemistry BG-10
Seattle, WA 98195
(206) 685-2052

Dunlap, Bruce
University of South Carolina
Department of Chemistry and Biochemistry
Columbia, SC 29208
(803) 777-6414

Dyer, R. Brian
Los Alamos National Laboratory
Chemical Science and Technology Division
MS J567
Los Alamos, NM 87545
(505) 667-4194

Edwards, Carol
Los Alamos National Laboratory
Life Sciences Division, MS M880
Los Alamos, NM 87545
(505) 665-2690

Floss, Heinz G.
University of Washington
Department of Chemistry BG-10
Seattle, WA 98195
(206) 543-0310

Gancarz, Alexander J.
Los Alamos National Laboratory
Chemical Science and Technology Division
MS J515
Los Alamos, NM 87545
(505) 667-4457

Garcia, Angel
Los Alamos National Laboratory
Theoretical Division, MS K710
Los Alamos, NM 87545
(505) 665-5341

Gilbert-Houghton, Terri
University of New Mexico
UNM Cancer Center, Cytometry Department
900 Camino de Salud, NE
Albuquerque, NM 87131
(505) 277-7249

Gilham, P. T.
Purdue University
Department of Biological Sciences
West Lafayette, IN 47907-139
(317) 494-4962

Gounarides, John
Sandoz Pharmaceuticals Corp.
59 Route 10
East Hanover, NJ
(201) 503-6394

Gronenborn, Angela M.
National Institutes of Health
Bldg. 5, Room 130
Bethesda, MD 20892
(301) 496-5414

Gupta, Goutam
Los Alamos National Laboratory
Theoretical Division, MS K710
Los Alamos, NM 87545
(505) 665-6463

Hansen, Andrew P.
Abbott Laboratories
D47G, AP9
Abbott Park, IL 60064
(708) 938-2477

Hernandez, Griselda
Los Alamos National Laboratory
Chemical Science and Technology Division
MS C345
Los Alamos, NM 87545
(505) 665-6976

Hicks, James
Parke-Davis Pharmaceutical Rs
2800 Plymouth Road
Ann Arbor, MI 48105
(313) 996-7509

Hoffman,Brian
Northwestern University
Department of Chemistry
2145 Sheridan Road
Evanston, IL 60208
(708) 491-7713 fax

Houck, David
Sterling Winthrop Inc.
1250 S.
Collegeville, PA 19426
(610) 983-5381

Huwe, Janice
USDA Biosciences Research Laboratory
P. O. Box 5674 University Station
Fargo, ND 58015
(701) 239-1236

Jett, Stephen
University of New Mexico
School of Medicine
Department of Cell Biology
900 Camino de Salud, N.E.
Albuquerque, NM 87131-5226
(505) 277-3443

Jones, Roger A.
Rutgers University
Department of Chemistry
P. O. Box 939
Piscataway, NJ 08855-0939
(908) 445-4900

Kainosho, Masatsune
Tokyo Metropolitan University
Faculty of Science
1-1 Minamiohsawa
Hachioji Tokyo, 192-03
JAPAN
81 426 77 2544

Kaptein, Robert
Utrecht University
Padualaan 8 3584 CH
Utrecht
HOLLAND
30 533787

Kennedy, Michael
Battelle
Box 999
Richland, WA 99352
(509) 372-2168

Kergil, Deanna
Los Alamos National Laboratory
Chemical Science and Technology Division
MS C345
Los Alamos, NM 87545
(505) 665-7282

Khalsa, Ongkar
Los Alamos National Laboratory
Chemical Science and Technology Division
MS C345
Los Alamos, NM 87545
(505) 665-2555

Khalsa, Guru Rattan
Los Alamos National Laboratory
Chemical Science and Technology Division
Stable Isotope Resource, MS C346
Los Alamos, NM 87545
(505) 667-6045

King, Garry C.
University of New South Wales
P. O. Box 1
Kensington, NSW 2033
AUSTRALIA
61 2 697 2021

Kurochkin, Alexander
University of Michigan
930 North University Ave.
Ann Arbor, MI 48109
(313) 763-0329

Lapham, Jon
Yale University
225 Prospect St.
New Haven, CT 06511
(203) 432-3991

LeMaster, David M.
Northwestern University
2-100 Hogan Hall
2153 Sheridan Road
Evanston, IL 60208
(708) 491-7329

Live, David
Sloan Kettering Institute
1275 York Ave.
Box 557
New York, NY 10021
(212) 639-2791

Manrao, Suraj
ISOTEC, Inc.
3858 Benner Road
Miamisburg, OH 45342
(513) 859-1808 (800) 448-9760

Markley, John L.
University of Wisconsin
420 Madison Mall
Madison, WI 53706
(608) 263-9349

Martinez, Rodolfo A.
Los Alamos National Laboratory
Chemical Science and Technology Division
MS C345
Los Alamos, NM 87545
(505) 667-1000

Martinez, Mary Ann D.
Los Alamos National Laboratory
Chemical Science and Technology Division
Stable Isotope Resource, MS C345
Los Alamos, NM 87545
(505) 667-5324

Matwiyoff, Nicholas A.
University of New Mexico
Center for Non-Invasive Diagnosis
1201 Yale N. E.
Albuquerque, NM 87131
(505) 277-0761

McClusky, Gary
Parke-Davis Pharmaceutical Research
2800 Plymouth Road
Ann Arbor, MI 48105
(313) 996-7525

McConnell, Bruce
University of New Mexico
1201 Yale N. E.
Albuquerque, NM 87131
(505) 277-8512

McInteer, B. B.
ICON Services Inc.
329 Potrillo Drive
Los Alamos, NM 87544
(505) 672-3642

Moore, Gregory J.
Los Alamos National Laboratory
Chemical Science and Technology Division
MS C345
Los Alamos, NM 87545
(505) 667-4006

Nadelson, Jeffrey
Sandoz Pharmaceuticals
59 Route 10
East Hanover, NJ 07936
(201) 503-7621

Newsome, Peter
Rhone-Povlene Ag. Co.
2 Alexander Drive
Research Triangle Park, NC 27709
(919) 549-2045

Norton, Raymond S.
Biomolecular Research Institute
381 Royal Parade
Parkville 3052
AUSTRALIA
61 3 903 9650

Okerlund, Linda
Bruker Instruments Inc.
47697 Westinghouse Drive
Fremont, CA 94539
(510) 683-4303

Olah, Glenn A.
Los Alamos National Laboratory
Chemical Science and Technology Division
MS C345
Los Alamos, NM 87545
(505) 667-4286

Ott, Donald
Los Alamos National Laboratory
Chemical Science and Technology Division
MS C920
Los Alamos, NM 87545
(505) 667-5637

Pardi, Arthur
University of Colorado
Department of Chemistry
P. O. Box 215 Boulder, CO 80309
(303) 492-6263

Patel, Ayyub
University of Dundee
Department of Anatomy and Physiology
Dundee, DD1, 4MN
SCOTLAND
0382 23181 x4814

Pettigrew, Jay
University of Pittsburgh
630 South Linden Avenue
Pittsburgh, PA 15208
(412) 648-8640

Qian, Yan Qiu
Memorial Sloan-Kettering Center
Box 557
1275 York Avenue
New York, NY 10021
(212) 639-7225

Rajagopal, Ponni
University of Washington
Center for Bioengineering SJ-70
Seattle, WA 98195
(206) 543-6391

Retford, Joe
ISOTEC, Inc.
3858 Benner Road
Miamisburg, OH 45342
(513) 859-1808

Rokop, Sue E.
Los Alamos National Laboratory
Chemical Science and Technology Division
MS C345
Los Alamos, NM 87545
(505) 667-2777

Santa Lucia, Jr., John
University of California, Berkeley
Department of Chemistry
Berkeley, CA 94720
(510) 642-1440

Schaefer, Jacob
Washington University
Department of Chemistry
One Brookings Drive
St. Louis, MO 63130
(314) 935-6844

Senn, Hans
Hoffmann-La Roche Ltd.
Bau 65/512
CH-4002 Basel
SWITZERLAND
0041 61 688 2028

Serianni, Anthony S.
University of Notre Dame
Department of Chemistry and Biochemistry
Notre Dame, IN 46556
(219) 631-7807

Silks, Louis A. "Pete"
Los Alamos National Laboratory
Chemical Science and Technology Division
Stable Isotope Resource, MS C345
Los Alamos, NM 87545
(505) 667-0151

Sklar, Larry
University of New Mexico
School of Medicine
Flow Cytometry Cancer Center
Albuquerque, NM 87131
(505) 277-7249

Spingola, Marc
University of New Mexico
School of Medicine
Department of Cell Biology
900 Camino de Salud, N.E.
Albuquerque, NM 87131-5226
(505) 277-8021

Springer, Penny A.
Los Alamos National Laboratory
Chemical Science and Technology Division
MS C345
Los Alamos, NM 87545
(505) 665-2555

Stark, Jean
Los Alamos National Laboratory
Protocol Office PA-2, MS P366
Los Alamos, NM 87545
(505) 667-6574

Stocking, Emily
Los Alamos National Laboratory
Chemical Science and Technology Division
Stable Isotope Resource, MS C345
Los Alamos, NM 87545
(505) 665-2555

Strzelecka, Teresa
Laboratory of Chemical Physics
9000 Rockville Pike Bldg. 5, Rm. B2-31
Bethesda, MD 20892
9301) 496-2815

Tan, Chou Tak
ISOTEC, Inc.
3858 Benner Road
Miamisburg, OH 45342
(513) 859-1808

Thanabal, Venkataraman
Parke-Davis Pharmaceutical Research
2800 Plymouth Road
Ann Arbor, MI 48105
(313) 996-7871

Theisen, Lisa
University of New Mexico
1201 Yale N. E.
Albuquerque, NM 87131
(505) 277-8512

Tinoco, Ignacio
University of California
Chemistry Department
Berkeley, CA 94720
(510) 6422-3038

Trewhella, Jill
Los Alamos National Laboratory
Chemical Science and Technology Division
MS C345
Los Alamos, NM 87545
(505) 667-2031

Tsetsis, Angela
Martek Biosciences Corporation
6480 Dobbin Road
Columbia, MD 21045
(410) 740-0081

Unkefer, Clifford J.
Chemical Science and Technology Division
Stable Isotope Resource Director, MS C345
Los Alamos National Laboratory
Los Alamos, NM 87545
(505) 667-5324

Unkefer, Pat J.
Chemical Science and Technology Division
MS J363
Los Alamos National Laboratory
Los Alamos, NM 87545
(505) 665-2557

Vacek, Harry
Cambridge Isotope Laboratories
50 Frontage Road
Andover, MA 01810
(508) 749-8000

Van Etten, Robert L.
Purdue University
Department of Chemistry
West Lafayette, IN 47907-1393
(317) 494-5276

Varela, Manuel
University of New Mexico
School of Medicine
Department of Cell Biology
900 Camino de Salud, N.E.
Albuquerque, NM 87131-5226
(505) 277-5773

Velupillai, Santhana
Theoretical Division, MS K710
Los Alamos National Laboratory
Los Alamos, NM 87545
(505) 665-3804

Wageman, William
Chemical Science and Technology Division
MS C345
Los Alamos National Laboratory
Los Alamos, NM 87545
(505) 667-5046

Walker, Thomas
ISOTEC, Inc.
3858 Benner Road
Miamisburg, OH 45342
(513) 859-1808

Woodard, Ronald
University of Michigan
930 North University Avenue
Ann Arbor, MI 48109
(313) 764-7366

Woodruff, William H.
Chemical Science and Technology Division
MS C345
Los Alamos National Laboratory
Los Alamos, NM 87545
(505) 665-2557

Wright, Peter E.
The Scripps Research Institute
Department of Molecular Biology
10666 North Torrey Pines Road
La Jolla, CA 29037
(619) 554-9721

Wu, Rullian
Chemical Science and Technology Division
MS C345
Los Alamos National Laboratory
Los Alamos, NM 87545
(505) 665-2555

Zimmer, Daniel
Yale University
11132 Town Walk Drive
Hamden, CT 06518
(203) 230-2207

Zuiderweg, Erik R.P.
University of Michigan
930 North University Avenue
Ann Arbor, MI 48109
(313) 936-3850