THE UNIVERSITY OF KANSAS

KANSAS INSTITUTE FOR THEORETICAL AND COMPUTATIONAL SCIENCE

Workshop on Algorithms for Macromolecular Modeling

Lawrence, Kansas

September 30 - October 2, 1994

WORKSHOP ON ALGORITHMS FOR MACROMOLECULAR MODELING

In the past, simplicity has often been the prime determining factor in the choice of algorithms in macromolecular modeling, but with more ambitious computations and higher expectations for realism, the demands on software and numerical algorithms have changed. The trend now is to combine novel numerical modeling techniques with sophisticated algorithms to bring big improvements in computational efficiency.

We hope that this workshop will benefit applied mathematicians and computer scientists from having direct interaction with the modelers—in formulating numerical and algorithmic problems and focusing on the most critical open problems. At the same time, we hope that biochemical modelers will profit by directing more attention to issues in numerics and algorithms.

We are grateful for generous financial support from

Kansas Institute for Theoretical and Computational Science

Department of Energy (DE-FG02-94ER25212.A000)

National Science Foundation (DMS-9412473)

and for tremendous assistance from Larisa Martin with the organization of the meeting.

Jan Hermans
Krzysztof Kuczera
Ben Leimkuhler
Bob Skeel

WORKSHOP ON ALGORITHMS FOR MACROMOLECULAR MODELING

Lawrence, Kansas, Sept.30-Oct. 2, 1994

All talks in Jayhawk room, except as noted

Friday, September 30, 1994

8:00 - 8:30	Registration
8:30 - 9:10	**Jan Hermans:** "Molecular Dynamics of Proteins: What has been achieved, and what problems stand in our way?"
9:15 - 9:55	**Tamar Schlick:** "Numerical Integration Schemes for Molecular Dynamics"
10:00 - 10:30	Coffee break
10:30 - 10:55	**Peter Wolynes:** "Navigating the Energy Landscape of Protein"
11:00 - 11:25	**John Straub:** "Simulated Annealing Using the Classical Density Distribution"
11:30 - 11:55	**Michael Levitt:** "Can the Amino Acid Sequence Distinguish the Correct Fold at Low Resolution?"
12:00 - 12:25	**Zhijun Wu:** "Continuation-Based Global Optimization for Molecular Conformation and Protein Folding"
12:30 - 2:00	Lunch
2:00 - 2:40	**Barry Honig:** "Recent Developments in the Application of Classical Electrostatics to Proteins and Nucleic Acids"
2:45 - 3:25	**Montgomery Pettitt:** "Modeling Water in and around Proteins"
3:30 - 4:15	Poster Session I: **Britt Park, Ilya Logunov, Witold Rudnicki, Shu-Yun Le, Adrian Roitberg, Hongmei Jian, Marios Philippopoulos, Gomathi Ramachandran, Lu Wang**
4:15 - 4:55	**Carol Post:** "Molecular Dynamics of Rhino Virus: Effects of an Antiviral Compound"
5:00 - 5:25	**Zan Luthey-Schulten:** "Protein Structure Prediction Using Optimized Hamiltonians"
5:30 - 5:55	**Carmay Lim:** "Discrete, Dynamic Polymer Modeling Using Cellular Automata Techniques"
6:00 - 7:30	Reception in Centennial room.

Saturday, October 1, 1994

8:00 - 8:30	Registration
8:30 - 9:10	**Peter Kollman:** "Force Fields and Free Energy Calculations for Complex Systems"
9:15 - 9:55	**Ron Elber:** "Molecular Dynamics with the Locally Enhanced sampling Method: Applications to Structure and Dynamics of Biomolecules"
10:00 - 10:30	Coffee break
10:30 - 10:55	**Arieh Warshel:** "Calculation of Electrostatic Energies in Macromolecules"
11:00 - 11:25	**Michael Holst:** "The Poisson-Boltzmann Equation"
11:35 - 12:00 Session 1 (Auditorium)	**Eric Jakobsson:** "Simulation of a Fluid Phase Lipid Bilayer Membrane: Incorporation of the Surface Tension into System Boundary Conditions"
Session 2 (Jayhawk room)	**David Case:** "Generation of Models for 'Unusual' DNA and RNA: A Computer Language for Structural Exploration"
12:05 - 12:30 Session 1 (Auditorium)	**Hon Chun:** "A Substructuring Approach for Reduced Variable Macromolecular Modeling"
Session 2 (Jayhawk room)	**Mike Prisant:** "Application of the Ray-Rpresentation and a Massively Parallel Special Pupose Computer to Problems of Protein Structure and Function"
12:30 - 2:00	Lunch
2:00 - 2:40	**Bernard Brooks:** "Advanced Methods for Macromolecular Simulation"
2:45 - 3:25	**Ridgway Scott:** "Parallel Algorithms for Biomolecular Modeling"
3:30 - 4:15	Poster Session II: **K. Srinivas, Yosi Shibberu, Robert Skeel, Marcia Fenley, Jianpeng Ma, Margaret Mandziuk, Hasim Saber, Qindsheng Zhao, Tai-Sung Lee**
4:15 - 4:55	**John A. Board:** "Multipole-Accelerated Evaluation of Force Fields in Macromolecular Modeling"

5:05 - 5:30 Session 1 (Kansas room)	**Dinesh Manocha:** "Conformational Analysis of Molecular Chains Using Nano-Kinematics"
Session 2 (Jayhawk room)	**Ramzi Kutteh:** "A Fast Multipole Algorithm for Molecular Simulations of Very Large Dipolar and Charged Dipolar Systems"
5:35 - 6:00 Session 1 (Kansas room)	**Eric Barth:** "Algorithms for Constrained Molecular Dynamics"
Session 2 (Jayhawk room)	**Lars Nyland:** "Algorithms for Shared Memory Parallel Molecular Dynamics in Sigma"

Sunday, October 2, 1994

8:15 - 8:55	**David Chandler:** "Modeling Quantum Processes in Complex Systems"
9:00 - 9:40	**Paul Bash:** "Modeling Enzyme Reactions Using Combined Quantum/Classical Methods"
9:40 - 10:00	Coffee Break
10:00:10:25	**Bogdan Lesyng:** "Quantum-Classical Molecular Dynamics, Theory and Algorithms"
10:30 - 10:55	**Weitao Yang:** "A Divide-and-Conquer Method and Its Application to Large Molecules"
11:00 - 11:25	**Alain St. Amant:** "Gaussian Density Functional Calculations on Large Systems Using Yang's Divide-and-Conquer Approach"

TIME TABLE AT A GLANCE

Friday, September 30, 1994 - morning

Time	Name	Room	Abstract Page
8:30 - 9:10	Jan Hermans	Jayhawk	39
9:15 - 9:55	Tamar Schlick	Jayhawk	73
10:00 - 10:30	Coffee break		
10:30 - 10:55	Peter Wolynes	Jayhawk	94
11:00 - 11:25	John Straub	Jayhawk	86
11:30 - 11:55	Michael Levitt	Jayhawk	56
12:00 - 12:25	Zhijun Wu	Jayhawk	95
12:30 - 2:00	Lunch break		

Friday, September 30, 1994 - afternoon

2:00 - 2:40	Barry Honig	Jayhawk	41
2:45 - 3:25	Montgomery Pettitt	Jayhawk	65
3:30 - 4:15	Poster Session I	Parlour next to Jayhawk	
	Britt Park		64
	Ilya Logunov		58
	Witold Rudnicki		71
	Shu-Yun Le		51
	Adrian Roitberg		70
	Hongmei Jian		48
	Marios Philippopoulos		66
	Gomathi Ramachandran		69
	Lu Wang		91
4:15 - 4:55	Carol Post	Jayhawk	67
5:00 - 5:25	Zan Luthey-Schulten	Jayhawk	59
5:30 - 5:55	Carmay Lim	Jayhawk	57
6:00 - 7:30	Reception	Centennial	

Saturday, October 1, 1994 - morning

Time	Name	Room	Abstract Page
8:30 - 9:10	Peter Kollman	Jayhawk	49
9:15 - 9:55	Ron Elber	Jayhawk	33
10:00 - 10:30	Coffee break		
10:30 - 10:55	Arieh Warshel	Jayhawk	92
11:00 - 11:25	Michael Holst	Jayhawk	40
PARALLEL SESSIONS			
11:30 - 11:55	Eric Jakobsson	Auditorium	43
	David Case	Jayhawk	24
12:00 - 12:25	Hon Chun	Auditorium	30
	Mike Prisant	Jayhawk	68
12:30 - 2:00	Lunch break		

Saturday, October 1, 1994 - afternoon

2:00 - 2:40	Bernard Brooks	Jayhawk	23
2:45 - 3:25	Ridgway Scott	Jayhawk	77
3:30 - 4:15	Poster Session II	Parlour next to Jayhawk	
	K. Srinivas		84
	Yosi Shibberu		83
	Robert Skeel		85
	Marcia Fenley		38
	Jianpeng Ma		60
	Margaret Mandziuk		61
	Hasim Saber		72
	Qindsheng Zhao		100
	Tai-Sung Lee		53
4:15 - 4:55	John A. Board	Jayhawk	21
PARALLEL SESSIONS			
5:00 - 5:25	Dinesh Manocha	Kansas	62
	Ramzi Kutteh	Jayhawk	50
5:30 - 5:55	Eric Barth	Kansas	13
	Lars Nyland	Jayhawk	63

Sunday, October 2, 1994 - morning

Time	Name	Room	Abstract Page
8:15 - 8:55	David Chandler	Jayhawk	28
9:00 - 9:40	Paul Bash	Jayhawk	14
9:40 - 10:00	Coffee break		
10:00 - 10:25	Bogdan Lesyng	Jayhawk	54
10:30 - 10:55	Weitao Yang	Jayhawk	98
11:00 - 11:25	Alain St. Amant	Jayhawk	11

PRELIMINARY LIST OF PARTICIPANTS

Alain St. Amant	Dept. of Chemistry, University of Ottawa, 10 Marie Curie, Ottawa, Ontario, K1N 6N5 CANADA, st-amant@theory.chem.uottawa.ca
Eric Barth	New York University/Courant Institute, barth@watson.biomath.nyu.edu
Paul A. Bash	Argonne National Laboratory, Bldg 202/A349, 9700 S. Cass Avenue, Argonne, IL. 60439, bash@mcs.anl.gov
John A. Board, Jr.	Dept. of Electrical Engineering, Duke University, Durham, NC 27706, jab@ee.duke.edu
Greg Brown	Dept. of Physics, Kansas State University, Manhattan, KS 66506, browngrg@phys.ksu.edu
Bernard Brooks	Bldg 12A Room 2041, NIH, Bethesda, MD 20892-5626, brbrooks@helix.nih.gov
John Canny	Computer Science Division, 529 Soda Hall, University of California, Berkeley, CA 94720, jfc@twocs.eecs.berkeley.edu
David A. Case	Dept. of Molecular Biology, MB1, The Scripps Research Institute, 10666 N. Torrey Pines Rd., La Jolla, CA 92307, case@scripps.edu
David Chandler	Dept. of Chemistry, University of California, Berkeley, CA 94720, chandler@hydrogen.CChem.Berkeley.EDU
Shih-I Chu	Dept. of Chemistry, University of Kansas, Lawrence, KS 66045, MPIATI@KUHUB.CC.UKANS.EDU
Hon M. Chun	Moldyn, Inc., 1033 Massachusetts Avenue, Cambridge, MA 02138-5319, hon@moldyn.com
Borries Demeler	University of Texas, Health Science Center, Dept. of Biochemistry, 7703 Floyd Curl Drive, San Antonio, Texas 78284-7760, demeler@bioc02.uthscsa.edu
Ron Elber	Dept. of Chemistry M/C 111, The University of Illinois at Chicago, Science and Engineering South Bldg., 845 West Taylor, Rm 4500, Chicago, IL 60607-7061 The Hebrew University of Jerusalem, Dept. of Physical Chemistry, Jerusalem, 91904 Israel, ron@pap.chem.uic.edu
Mahmoud Fath El-Den	Dept. of Math. & Comp. Sci., FHSU, Hays, KS 67601-4099, MDMFE@FHSUVM.FHSU.EDU
Marcia O. Fenley	Rutgers University, Dept. of Chemistry, P.O. Box 939, Piscataway, New Jersey 08855-0939, FENLEY@chemf.rutgers.edu
Jan Hermans	Dept. of Biochemistry and Biophysics, Univ. of North Carolina, Chapel Hill, NC 27599-7260, hermans@mcnc.org
Michael Holst	Applied Math 217-50, Caltech, Pasadena, CA 91125, holst@ama.caltech.edu
Barry Honig	Dept. of Biochemistry and Molecular Biophysics, Columbia University, 630 West 168 St., New York, NY 10032, honig@cuhhca.hhmi.columbia.edu
D. J. Ierardi	Dept. of Computer Science, SAL 300, University of Southern California, 941 West 37th Place, Los Angeles, CA 90089-0781, ierardi@flash.usc.edu
Sergei Izrailev	Dept. of Physics, University of Illinois, 1110 W Green St., Urbana, IL 61801, sergei@lisboa.ks.uiuc.edu
Eric Jakobsson	University of Illinois at Urbana-Champaign, Beckman Institute, 405 N. Mathews Ave., Urbana, IL 61801, jake@ncsa.uiuc.edu
Laurent Jay	Dept. of Computer Science, University of Minnesota, 4-192 EE/CSci Building, 200 Union Street S.E., Minneapolis, MN 55455-0159, na.ljay@na-net.ornl.gov
Hongmei Jian	New York University, jianh@acf2.nyu.edu

Peter Kollman Dept. of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, pak@cgl.ucsf.edu
Man Kong Dept. of Computer Science, University of Kansas, Lawrence, KS 66045, kong@cs.ukans.edu
Krzysztof Kuczera Dept. of Chemistry, University of Kansas, Lawrence, KS 66045, kuczera@tedybr.chem.ukans.edu
Ramzi Kutteh Molecular Science Research Center, Battelle, PNL, P.O.Box 999 / MSIN K1-96, Richland, WA 99352, ra_kutteh@pnl.gov
Brian Laird Dept. of Chemistry, University of Kansas, Lawrence, KS 66045, laird@pilsner.chem.ukans.edu
Shu-Yun Le Lab of Mathematical Biology, NCI, NIH, Bldg 469, Rm 151, Frederick, MD 21702, SHUYUN@FCRFV2.NCIFCRF.GOV
Tai-Sung Lee Dept. of Chemistry, Duke University, Durham, NC 27710, taisung@chem.duke.edu
Ben Leimkuhler Dept. of mathematics, University of Kansas, Lawrence, KS 66045, leimkuhl@math.ukans.edu
Bogdan Lesyng Univ. of Houston, lesyng@plearn.bitnet
Michael Levitt Beckman Laboratories for Structural Biology, Dept. of Cell Biology, Stanford University School of Medicine, Stanford, CA 94305--5400, levitt@cellbio.stanford.edu
Jie Liang NCSA, Beckman Institute, 405 N. Mathews Ave., Urbana, IL 61801, jliang@silibio.ncsa.uiuc.edu
Carmay Lim Dept. of Molecular & Medical Genetics, Faculty of Medicine, University of Toronto, Toronto, Ontario, CANADA M5S 1A8, carmay@ovid.med.utoronto.ca
Guang H. Liu CB#3290, Dept of Chemistry, Univ. of North Carolina, Chapel Hill, NC 27599-3290, ghliu@gibbs.oit.unc.edu
Ilya Logunov 3149 Beckman Institute, 405 N Mathews Ave., Urbana, IL 61801, ilya@lisboa.ks.uiuc.edu
Eric Jui-Lin Lu Computer Science Dept., Room 325, University of Missouri-Rolla, Rolla, MO 65401, jlu@cs.umr.edu
Zan Luthey-Schulten B56 Noyes Lab, MC 712, 600 S Mathews, Urbana, IL 61801, zan@nsca.uiuc.edu
Jianpeng Ma Dept. of Chemistry, Boston University, 590 Commonwealth Ave., Boston, MA 02215jma@chem.bu.edu
Mark Mallalieu Dept. of Chemistry, University of Kansas, Lawrence, KS 66045
Margaret Mandziuk Chemistry Dept. & Courant Institute of Mathematical Sci., 251 Mercer Street, New York, New York 10012, mandzk@cims.nyu.edu
Dinesh Manocha Dept. of Computer Science, University of North Carolina, Sitterson Hall, CB#3175, Chapel Hill, NC 27599-3175, manocha@cs.unc.edu
Jorge J. Moré Mathematics and Computer Science Division, Argonne National Laboratory, 9700 S. Cass Ave., Argonne, Illinois 60439-4844, more@mcs.anl.gov
Lawrence S. Norris Depts. of Biomedical Engineering and Chemistry, Northwestern University, 2145 Sheridan, Evanston, IL 60208, lnorris@nwu.edu
Lars S. Nyland Dept. of Computer Science, University of North Carolina, Chapel Hill, NC 27599-3175, nyland@cs.unc.edu
Daniel Okunbor University of Missouri-Rolla, okunbor@cs.umr.edu
Britt Park Beckman Laboratories for Structural Biology, Dept. of Cell Biology, Stanford University School of Medicine, Stanford, CA 94305--5400, britt@hyper.stanford.edu
Seongbin Park 667 Levering Ave. #06, Los Angeles, CA 90024, separk@pollux.usc.edu
Montgomery Pettitt Chemistry Dept., University of Houston, Houston, Texas 77004B56, pettitt@kitten.chem.uh.edu

Marios Philippopoulos Dept. of Molecular & Medical Genetics, Faculty of Medicine, University of Toronto, Toronto, Ontario, CANADA M5S 1A8

Carol Post Dept. of Med. Chemistry, Purdue University, Lilly Hall, West Lafayette, IN 47907-1392, b4i@mace.cc.purdue.edu

Jan F. Prins Dept. of Computer Sci., University of North Carolina, Chapel Hill, NC 27599-3175, prins@cs.unc.edu

Mike Prisant Dept. of Computer Science, Duke University, Durham, NC 27706, prisant@cs.duke.edu

Gomathi Ramachandran Courant Institute of Mathematical Sci., 251 Mercer Street, New York, New York 10012, ramachan@franklin.biomath.nyu.edu

Adrian E. Roitberg Dept. of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3113, adrian@silver.chem.nwu.edu

Witold Rudnicki Dept. of Chemistry, University of Houston, Houston, Texas 77204, rudnicki@tracer.chem.uh.edu

Hashim A. Saber Dept. of Mathematical Sciences, Montana State University, Billings, MT 59101, mahas@selway.umt.edu

Tamar Schlick New York University, Courant Institute, 251 Mercer Street, New York, NY 10012-1185, schlick@cims.nyu.edu

Ridgway Scott University of Houston, Dept. of Mathematics, 694 Phillip Guthrie Hoffman, Houston, TX 772043476, scott@math.uh.edu

Yosi Shibberu Mathematics Dept., Rose-Hulman Institute of Technology, Terre Haute, IN 47803, shibberu@nextwork.rose-hulman.edu

K. Srinivas Beckman Institute, University of Illinois, Urbana, IL 61801, cheenu@lisboa.ks.uiuc.edu

Jonathan Simon Dept. of Mathematics, University of Iowa, Iowa City, IA 52242, simon@math.uiowa.edu

Robert D. Skeel Dept. of Computer Sciences, University of Illinois at Urbana Champaign, 1304 W Springfield Ave., Urbana, Illinois 61801, skeel@cs.uiuc.edu

Gerhard Starke Applied Math Program, University of Colorado, Boulder, CO 80309-0526, starke@newton.Colorado.EDU

John Straub Dept. of Chemistry, Boston University, 590 Commonwealth Ave., Boston, MA 02215, straub@chem.bu.edu

Brian Tao Dept. of Molecular & Medical Genetics, Faculty of Medicine, University of Toronto, Toronto, Ontario, CANADA M5S 1A8

Erik Van Vleck Dept of Math and Comp Sci., Colorado School of Mines, Golden, CO 80401, erikvv@lyapunov.mines.colorado.edu

Jingyan Wang Dept. of Chemistry, University of Kansas, Lawrence, KS 66045

Lu Wang Dept. of Biochem. & Biophys., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7260, wanglu@gibbs.oit.unc.edu

Yousong Wang Dept. of Chemistry, University of Kansas, Lawrence, KS 66045, color93@kuhub.cc.ukans.edu

Arieh Warshel Dept. of Chemistry, University of Southern California, Angeles, CA 90089, warshel@usc.usc.edu

Peter Wolynes Dept. of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, wolynes@aries.scs.uiuc.edu

Chung F. Wong Mount Sinai School of Medicine, Dept. of Physiology and Biophysics, Box 1218, 1 Gustave L. Levy Place, New York, NY 10029-6574, WONG@msvax.mssm.edu

William V. Wright Dept. of Computer Science, University of North Carolina, Sitterson Hall, CB# 3175, Chapel Hill, NC 27599-3175, wright@cs.unc.edu

Zhijun Wu 1621 Westminster #106, Naperville, IL 60563, zwu@mcs.anl.gov

Weitao Yang Dept. of Chemistry, Duke University, Durham, NC 27710, yang@chem.duke.edu
Guihua Zhang New York University, zhangui@acf2.nyu.edu
Qingsheng Zhao Dept. of Chemistry, Duke University, Duhram, NC 27706, qzhao@uncvx1.oit.unc.edu

ABSTRACTS

of

TALKS and POSTERS

Gaussian Density Functional Calculations on Large Systems Using Yang's Divide–and–Conquer Approach

Alain St-Amant, Department of Chemistry, University of Ottawa

Density functional theory (DFT) has recently garnered a great deal of attention from computational chemists. Compared to the simplest *ab initio* wavefunction–based approach, the Hartree–Fock approximation, DFT methods incorporate the effects of electron correlation yet require less computational resources. With the advent of gradient–corrected DFT, the popularity of these methods has greatly increased and systematic benchmarks tend to situate the quality of their results at the level of the considerably more computationally demanding MP2 level of theory.

Our particular benchmark of DFT (A. St-Amant, W. D. Cornell, T. A. Halgren, and P. A. Kollman, *J. Comp. Chem.*, submitted) focussed on an ensemble of roughly one hundred organic molecules containing a wide variety of functional groups. Within gradient–corrected DFT, bond lengths are systematically overestimated, the RMS error being 0.18 Å. The RMS error within MP2 is 0.010 Å. For bond angles, the RMS DFT and MP2 errors are respectively 0.8° and 0.7°. For torsional angles, where experimental data is far more scarce and far less reliable, the average DFT and MP2 errors fall within three to four degrees. In a demanding test of DFT's ability to provide reliable energies, relative conformational energies were also studied. For the ensemble of 35 molecules for which experimental data is available, the DFT and MP2 RMS errors are found to be 0.46 and 0.43 kcal mol^{-1}, respectively. The glycine and alanine dipeptide analogs, for which no experimental data is available, were also studied since they represent the simplest models of a polypeptide. The gradient–corrected DFT and MP2 results for these systems' low–lying conformers agreed fairly well upon their structures and relative conformational energies. Results from HF and DFT calculations without gradient–corrections on these dipeptide analogs appeared deficient and the use of results from these two levels of theory for the purpose of force field parameterization would seem to be discouraged.

With the promise of these results in hand, we are currently trying to apply DFT methods to large biological systems. To treat these large systems, we have begun to implement Yang's divide–and–conquer DFT approach (W. Yang, *J. Mol. Struct. (Theochem)*, 255, 461 (1992)) within our gaussian DFT code, DeFT (DeFT may be obtained free of charge by sending a short note to request@theory.chem.uottawa.ca). In Yang's divide–and–conquer scheme, a large system is subdivided into a collection of subsystems. The global Hamiltonian is then projected onto these subsystems and a subsequent independent diagonalization process is carried out on each subsystem to extract a set of subsystem orbitals. These subsystem orbitals are then assigned electrons following the Aufbau principle until we reach the point where by piecing together all the subsystem densities, the total number of electrons is conserved. Communication between the subsystems occurs through

the electrostatic potential and the fact that each subsystem obeys a global Fermi energy. The key point here is that the divide–and–conquer scheme takes us away from conventional DFT methods that scale as N^3 (where the global molecular orbitals must be obtained by diagonalization of a matrix that scales linearly with system size) to a method that scales linearly with system size. The divide–and–conquer method scales linearly since the CPU time is dominated by the independent work on the individual subsystems, which scale linearly with system size. Also, since the CPU–dominating subsystem calculations can be carried out independently, the divide–and–conquer philosophy is ideally suited for massively parallel architectures. The combination of linear scaling and inherent parallelism have led us to test this approach on large biomolecular systems of interest to us.

However, it should be noted at this point that the projection of the global Hamiltonian onto the subsystems is an approximation. To obtain accurate results, the Hamiltonian must be projected not only onto the subsystem's atoms, but also onto a set of associated buffer atoms. A system's buffer atoms are those atoms that are within a certain distance of any of the subsystem's atoms. By adding to the buffer space atoms that are within five or six bonds of a subsystem atom, Yang has been able to reduce the errors in the divide–and–conquer DFT energies to about $\pm$ 0.001 hartrees. However, this can correspond to fluctuations with a range of roughly 1.5 kcal mol^{-1}, a number that is far too high for our studies of biological systems. Buffer space can be extended, but this results in a dramatic increase in the CPU time devoted to each of the subsystem calculations.

In an attempt to improve the precision and economy of our divide–and–conquer gaussian DFT calculations, we have retained the concept of buffer atoms but we have extended it by using various levels of basis set sophistication. In this approach, each atom can be assigned, simultaneously, double–ζ with polarization and diffuse functions, double–ζ with polarization function, double–ζ, and minimal, single–ζ, basis sets. Within a subsystem calculation, the subsystem atoms and nearby buffer atoms are assigned the most sophisticated basis sets while those buffer atoms at the outer limits of buffer space are assigned minimal basis sets. Buffer atoms within these two extremes can be assigned intermediate quality basis sets. Therefore, the quality of an atom's basis set will change from subsystem calculation to subsystem calculation, and its quality will depend on it's position relative to the actual subsystem atoms. This scheme allows us to extend the spatial extent of the buffer space while keeping the number of basis functions as low as possible.

Since this is the first implementation of divide–and–conquer DFT within a gaussian DFT framework, a divide–and–conquer approach to fitting the charge density and exchange–correlation potentials has to be devised as well if we are to retain linear scaling. It can be shown that this can be easily achieved by extending the partitioning scheme already used to obtain the subsystem orbitals.

To date, we have completed test calculations on only dipeptides and tripeptides and have been able to keep the errors down to about 0.5 kcal mol^{-1} in relative conformational energies. This is a promising first step towards performing divide–and–conquer DFT calculations on large biomolecular systems. Ultimately, we would like to use such approaches in modelling enzymatic reactions.

ALGORITHMS FOR CONSTRAINED MOLECULAR DYNAMICS

Eric Barth

Courant Institute of Mathematical Sciences

In molecular dynamics simulations, the fastest components of the potential severely limit the size of timestep for explicit integration schemes. The problem has been treated by replacing these components with algebraic constraints. The imposition of constraints necessitates the solution of a systems of nonlinear equations at each integration step. In a popular discretization scheme for the resulting constrained Hamiltonian system, the algebraic structure of the nonlinear equations is closely related to the bond structure of the molecule being simulated.

This property is exploited to analyse the convergence of the traditional solution algorithm SHAKE, and to motivate alternative methods which enjoy improved performance.

Computer Simulation Methods for Enzyme Reactions

Paul A. Bash, Lawrence Ho, and Surya Panigrahy
Argonne National Laboratory
Argonne, IL. 60439

The Advanced Photon Source at Argonne National Laboratory will provide the capability in 1996 to carry out time-resolved x-ray crystallography experiments to determine the structures of metastable intermediates of enzyme reaction pathways. Such structural data are necessary but insufficient to characterize the mechanisms and energetics of enzyme reactions. Complete atomic mechanisms can only be determined using a first-principles analysis based on quantum and statistical mechanics combined with structural information for the enzyme and its interaction with its substrate. The long-term goal of our work is to develop theoretical and computational methods that combine quantum and classical mechanics methods, implement them on massively parallel computers, and use them to characterize the atomic and electronic events of enzyme reaction mechanisms. These computational techniques will provide a tool for determining the atomic structures and energetics along enzyme reaction pathways. Such methods will complement and enhance time-resolved crystallographic structural data and provide the means to predict the effects of site-specific mutagenesis experiments.

The specific agenda for our work in the near future is:

(1) Develop the capability to determine minimum energy and free energy surfaces and pathways for enzyme reactions. This will be accomplished using a hybrid, semiempirical quantum mechanics and classical molecular mechanics (QM/MM) method, together with free energy perturbation and molecular dynamics techniques.

(2) Test the accuracy of this method on small molecule systems in the gas phase that represent the key functional groups of the enzyme system malate dehydrogenase. Both the absolute energetics of reactions and the structures of the participating molecules, as well as interaction energies, will be characterized by comparisons with high level Hartree-Fock *ab initio* methods.

(3) Demonstrate the utility of these methods by applying them to the reaction catalyzed by the enzyme malate dehydrogenase (MDH). Free energy transition states of native and site-directed mutants of MDH will be calculated and used to determine the relative transition state binding free energies of native and mutant enzymes. These calculated relative free energies will be compared with corresponding experimental values derived from enzyme kinetic data.

(4) Combine a density functional ab initio quantum method with molecular mechanics in order to increase applicability of the QM/MM method to a wider class of enzyme systems. The density functional implementation will be a localized orbital approach that uses (a) a numerical basis set, (b) gradient corrections to the exchange and correlation, and (c) a "divide and conquer" procedure to enhance computational efficiency.

(5) All methods will be implemented to run on massively parallel, multiple instruction, multiple data (MIMD) computers.

To study the properties of condensed phase reactions we have developed an approach that uses (1) a computational quantum mechanics method to treat the part of a system where the electronic structure is changing do to bond making and breaking events and (2) a classical mechanics technique to model the rest of the system.[1,2] This work extended and enhanced the concepts and implementations of earlier efforts to use quantum and classical mechanics methods to study complex molecular systems.[3,4] The specific hybrid quantum and molecular mechanics (QM/MM) method that was developed can model the chemical reactions of an enzyme and include solvation effects due to the protein matrix. For most enzyme reactions, the changes in electronic structure are localized to a small region of the active site (less than 5% of the entire enzyme, or about 25-100 atoms). It is reasonable to hypothesize that only this region of the enzyme needs to be modeled with quantum mechanics. The use of this approach can dramatically reduce computational requirements.

The key to the proper implementation of a QM/MM approach is the representation of different parts of a system with a method whose accuracy is commensurate with relevant experimental data. A particular system can be divided into the following regions:

THE UNIVERSITY OF KANSAS

KANSAS INSTITUTE FOR THEORETICAL AND COMPUTATIONAL SCIENCE

Workshop on Algorithms for Macromolecular Modeling

Lawrence, Kansas

September 30 - October 2, 1994

Protein-protein (solvent-solvent). These interactions are not involved directly with bond breaking or formation and should reproduce structural and dynamical properties near a level of accuracy of protein crystal structures. A molecular mechanics method such as CHARMM[5] has proven to be a reasonable model for the structural properties of proteins.[6]

Reaction region (solute-solute). Components of a system where the electronic structure changes during bond making and breaking must be treated with quantum mechanics. The primary challenge for the successful application of a QM/MM method is the development and use of a QM approach that is both efficient and accurate. These two requirements are not necessarily compatible, and a large portion of our future methods development will deal with solutions to this problem.

Protein-substrate nonbonded (solute-solvent). The accurate modeling of protein-substrate interactions is crucial to the success of a QM/MM method. A primary function of an enzyme is to stabilize reaction intermediates by solvent effects due to its unique amino acid environment. Our experience, and others, with a QM/MM method that uses semiempirical quantum mechanics suggests that solvent effects can be adequately treated for both reactions in enzymes[7] and solution.[1,8] However, this component of the Hamiltonian should be tested for each application.

The utility of our QM/MM approach has been demonstrated by (1) reproducing the experimental free energy of reaction for the S_N2 reaction of chloride plus methyl chloride in solution[1] and (2) suggesting a novel mechanism for the enzyme reaction in triosephosphate isomerase[7] that was subsequently verified by experiment.[9] This method has also been effectively used by Gao[8,10] to gain insights into the behavior of organic reactions in solution, providing further evidence of the utility of a QM/MM approach for the study of chemical reactions in the condensed phase.

Our semiempirical QM/MM method consists of the following three components:

MM-MM interactions. These are modeled with the CHARMM force field[11] and empirical potential energy function.[5]

QM-QM interactions. These are represented by the semiempirical AM1 Hamiltonian and parameterization.[12] We are in the process of adding a density functional quantum mechanical method to increase the reliability and extend the applicability of our method.

QM-MM interactions. These are composed of four different terms. The first is MM (partial charge) and QM (electron) interaction, which is included directly into the one-electron portion of the QM Hamiltonian. The second is the coulomb interaction between charges centered on the atoms, which includes partial charges on MM atoms and the net core positive charge (equal to the number of valence electrons for each QM atom). The third is a molecular mechanics Van der Waals interaction between QM and MM atoms. This emulates the electron repulsion and dispersion interactions between QM and MM atoms that can not be calculated directly because MM atoms have no explicit electrons. The fourth is the interaction of QM and MM atoms that are connected by a chemical bond. In this instance, the QM atom will contain an unpaired electron because the MM atom has no electron to contribute to the bond. To satisfy the valency, an extra hydrogen atom is added between the bonded QM and MM atoms. MM bond, angle, and dihedral interactions are calculated between QM and MM atoms when at least one of the atoms in such terms is MM.

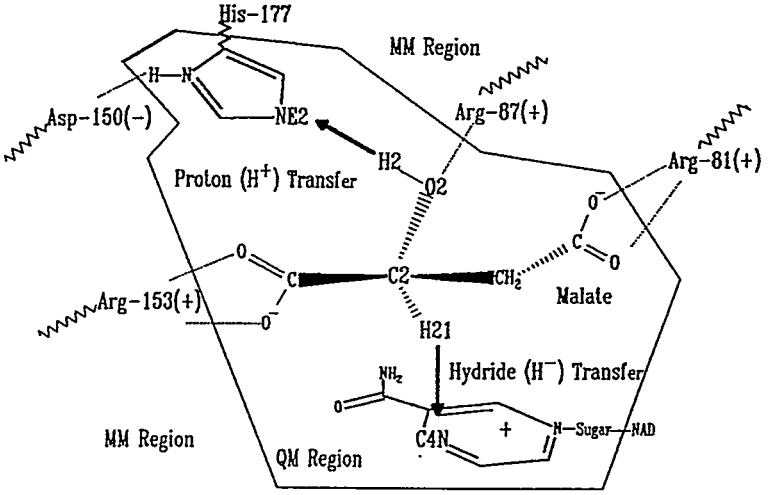

Figure 1. Active site and co-factor NAD for the reaction in malate dehydrogenase.

One of the objectives of our work is to create new methods and procedures for the determination of free energies associated with enzyme reactions. To develop and test such methods, we select an enzyme system as a test bed and verify that our QM/MM Hamiltonian can adequately model the chemical components that characterize this particular enzyme. We use the enzyme malate dehydrogenase (MDH), which catalyzes the conversion of malate to oxaloacetate in the citric acid cycle. There is an excellent x-ray structure available at 1.9Å resolution[13] for a ternary complex consisting of the enzyme, the substrate analog citrate, and nicotinamide adenine diphosphate (NAD), and there is good thermodynamic and kinetic data for site-specific mutants.[14] Figure 1 shows the key elements, their orientation in the active site of MDH, and the QM/MM partition to be used in all our simulations The reaction catalyzed by this enzyme is the transfer of a hydride anion, H21, from C2 on malate to C4N in the nicotinamide ring of NAD, and the transfer of a proton, H2, from O2 on malate to NE2 of His-177 in the enzyme.

To test the applicability of our QM/MM method in the context of this enzyme system, the following calculations were done.

MM-MM interactions. The active site region of MDH was modeled using a stochastic boundary molecular dynamics procedure.[15] An 18Å spherical region was generated about the C2 atom of malate that included the enzyme, malate, NAD, crystallographic water, and TIP3P water[16] added to fill the active site region out to 18 Å. A reaction zone of 16 Å was used with a buffer region of 2 Å (16-18Å from C2). Harmonic restraints are placed on buffer zone atoms with reference to corresponding x-ray coordinates and force constants derived from the crystallographic temperature factors. MM parameters for protein atoms were taken from the new CHARMM all atom force field.[11] MM parameters for the hydroxyl, carboxylate, and methylene functional groups in malate were generated by analogy with corresponding amino acid functional groups. Parameters for NAD were determined by analogy with the CHARMM all atom DNA force field[11] with the exception of charges for the nicotinamide portion of NAD, which were generated using the ESP[17] option from MOPAC 6.0. Using SHAKE[18] to keep bond lengths constant and 2 fs time steps, a constant energy molecular dynamics simulation was carried out. Heating and equilibration was done for 60 ps and the subsequent 40 ps trajectory was used to determine an average structure for comparison with the x-ray crystal structure. The RMS difference between the crystal and simulated structures was 0.55 Å for Cα atoms and 1.05 Å for all atoms. This reasonable agreement with experiment suggests that the CHARMM MM force field should be adequate to model the overall structural features of MDH within our QM/MM scheme.

QM-MM interactions. Tests on small molecule systems[2] indicate that the QM/MM method, using semiempirical QM, can produce nonbonded potential energy surfaces in reasonable agreement to MP2/6-31G* level of *ab initio* theory. Figure 1 depicts the QM/MM partition of the active site of MDH and suggests that the most significant QM/MM interactions are between the guanidinium groups of Arg-81 and Arg-153 and the carboxylate groups of malate. These strong electrostatic interactions dominate the binding of malate to MDH and should have a significant effect on the electronic structure changes that occur during the reaction. We evaluate our QM/MM to model the interaction between these groups by determining the potential energy surface for a guanidinium/formic acid complex in the gas phase using (a) Gaussian Hartree-Fock at an MP2/6-31+G* level of theory with counterpoise-corrections,[19] (b) AM1 semiempirical QM (formic acid) and CHARMM MM (guanidinium), and (c) CHARMM MM (formic acid) and CHARMM MM (guanidinium). The CHARMM parameters for formic acid and guanidinium were determined by analogy to corresponding functional groups in arginine and glutamate from the all atom protein force field. The structures of guanidinium and formic acid were optimized separately in the gas phase using MP2/6-31+G*, AM1, and MM levels of theory. Using these structures, the complex was formed in C_2V symmetry (Figure 2) and the distance between the carbon atoms in the two molecules was varied from 3.1 to 13.0Å.

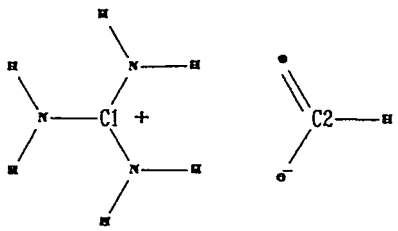

Figure 2. Guanidinium (+) and formic acid (-) complex in C_2V symmetry.

The interaction energy was determined for each conformation with no additional geometry optimization at each configuration. The resultant potential energy surfaces at each level of theory are shown in Figure 3.

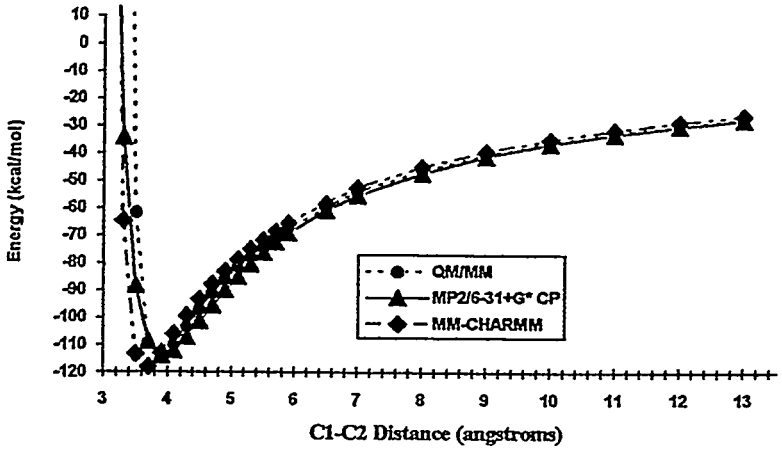

Figure 3. Comparison of the potential energy surface between guanidinium and formic acid using counterpoise-corrected MP2/6-31+G* *ab initio*, combined AM1 semiempirical QM and molecular mechanics, and molecular mechanics methods.

This test indicates that the QM/MM Hamiltonian is able to calculate the potential energy surface for this complex in excellent agreement with a high level of *ab initio* theory. This is a very encouraging result. MP2 counterpoise-corrected interaction energies are not perfect,[20] however, this level of theory provides a reasonable standard with which to calibrate more approximate methods. There is no experiment capable of determining the potential energy surface for this kind of charged complex. The guanidinium and formic acid complex was also optimized at the MP2/6-31+G* level without the imposition of C_2V symmetry to obtain the equilibrium geometry. The resultant *ab intio* structure compares favorably to the QM/MM optimized complex. The agreement between the two calculations is excellent, with bond lengths differing by about 0.05 Å and bond angles by no more than 5°, and the interaction energy is within 2 kcal/mol.

In addition to these gas phase tests, the classical MD simulation of the enzyme, described above, was continued with QM used to describe the active site atoms indicated in Figure 1. Starting at the end of the 100 ps MM simulation, 40 ps of MD (20 ps of equilibration and 20 ps of data collection) was carried out with the QM activated. An average structure was determined during the data collection phase and compared with the crystal structure and the average structure from the MM simulation. The QM/MM and MM structures differed by 0.16Å and 0.45Å RMS for Cα and all atoms, respectively. The QM/MM and x-ray structures differed by 0.6Å and 1.1Å for Cα and all atoms, respectively. There was no fundamental difference between the QM/MM and purely MM simulations with respect to structural characteristics. All these tests suggest that QM/MM interactions appear to be well balanced in the context of the MDH model.

QM-QM interactions. The most problematic aspect of the QM/MM method is the determination of energy barriers for reactions. Although it is a very long-range goal to be able to determine absolute rate constants from first principles, our immediate objective is more realistic: *to develop methods that can determine relative reaction free energies to an accuracy of 1-2 kcal/mol.* However, we characterize the ability of the semiempirical method to calculate absolute reaction energies in order to ascertain its limits of applicability. Tests are carried out on small model systems at AM1 and MP2 levels of theory. For the proton and hydride reactions of the MDH system, we represent the proton transfer by methanol and imidazole and the hydride transfer by methanol and nicotinamide in the gas phase. The reaction profiles are determined by constraining the proton or hydride to distances of 1.0-2.0 Å or 1.1-2.0 Å from NE2(imidazole) or C4N(nicotinamide). Figure 4 shows the energy profiles for the two reactions, and Figure 5 plots the distances of the proton or hydride from donor and acceptor atoms.

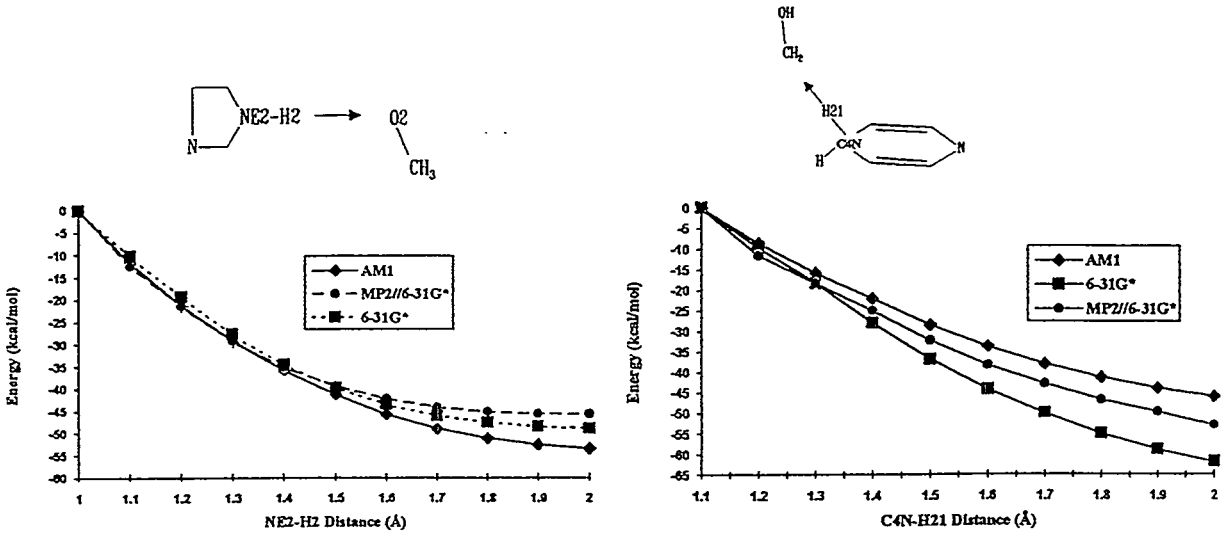

Figure 4. Energy profiles for the hydride transfer (right) from nicotinamide to methanol and proton transfer (left) from imidazole to methanol.

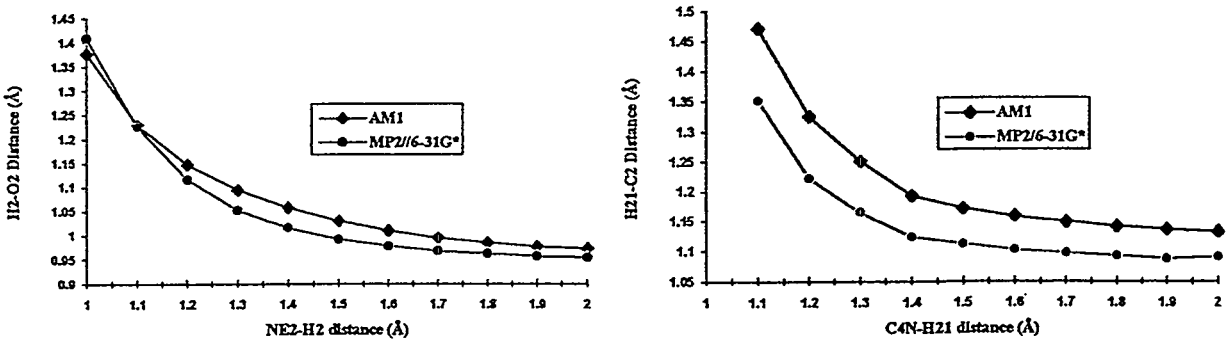

Figure 5. Distances of H2 from donor (NE2) and acceptor (O2) for a proton transfer and H21 from donor (C4N) and acceptor (C2) for a hydride transfer at two levels of theory. C4N-H21 and NE2-H2 are constrained to the ranges of 1.1-2.0 Å and 1.0-2.0 Å, respectively, in 0.1 Å increments.

Figure 5 indicates that AM1 and *ab initio* derived structures agree within 0.05-0.1 Å for donor-hydrogen and acceptor-hydrogen distances, which is similar to results for equilibrium structures of molecules in gas phase. Figure 4 shows that energies agree within 5-10 kcal/mol. Since structures determined with *ab initio* methods, in general, agree well with experiment, our results suggest that absolute geometries for states along a reaction pathway, e.g., the transition state, may be represented reasonably well even at the AM1 level of theory. However, our results indicate that *absolute* reaction energetics cannot be used to obtain quantitative insights into *absolute* rates of reaction. Errors in the AM1 method may be systematic and *relative* reaction energies may be calculated with much higher accuracy. We will be testing this hypothesis through our development of methods to determine relative free energies of reaction and their comparison with data derived from MDH site-directed mutation experiments.

Using the above calibrations and tests as a basis, we calculated the minimum energy surface for the hydride and proton transfer reactions in the MDH enzyme using our QM/MM method. This was done to obtain insights into the details of the minimum energy pathway and transition state(s) for the reaction which are not possible to determine experimentally. We carried out a systematic search over the relevant portions of conformational space associated with the transfer of the proton, H2, from O2 to NE2 and the hydride, H21, from C2 to C4N (atom designations from Figure 1). This was a four-dimensional search over the distances (a) O2-H2, (b) H2-NE2, (c) C2-H21, and (d) H21-C4N. We used a 0.2 Å grid spacing, where the range of distances were restricted such that 2.4 Å < d(O2-H2) + d(H2-NE2) < 3.2 Å and 2.4 Å < d(C2-H21) + d(H21+C4N) < 3.5 Å, d(X-Y) means the distance between X and Y. The same starting point, i.e., the MDH atomic model, was used for each set of distance parameters. This state was determined by molecular dynamics simulated annealing with d(O2-H2), d(H2-NE2),

d(C2-H21), and d(H21-C4N) constrained to be the same 1.3 Å distance, using SHAKE,[18] and all other degrees of freedom allowed to vary consistent with the stochastic boundary procedure[15] used to model the enzyme active site region that was described above. A final minimized structure was obtained where the energy difference over the last 100 steps was less than 0.01 kcal/mol and the final RMS gradient of the forces was less than 0.01 kcal/mol·Å.

Using this starting geometry, a total of 800 separate energy minimizations were done over the set of parameters of the four distances defined above. All calculations were done on the 128 processor IBM SP1 parallel computer at Argonne National Laboratory, where each processor was assigned an MDH model with a different set of distance parameters. Each simulation was run for 500 steps. The energy convergence for all the calculations was $\Delta E = 0.009 \pm 0.01$ kcal/mol over the last 100 steps of minimization where $\Delta E = E(\text{step}=500) - E(\text{step}=400)$. The RMS gradient was less then 0.02 kcal/mol·Å for all calculations. We obtain the minimum energy surface for this reaction by projecting out the two degrees of freedom that are "natural" coordinates for the MDH reaction for the malate to oxaloacetate. These are d(O2-H2) and d(C2-H21). For each pair of distances for these reaction coordinates, e.g. d(O2-H2)=1.3 and d(C2-H21)=1.3, the lowest energy value was selected out of the other possible conformations of d(H2-NE2) and d(H21-C4N). The result of this projection operation is shown in Figure 6.

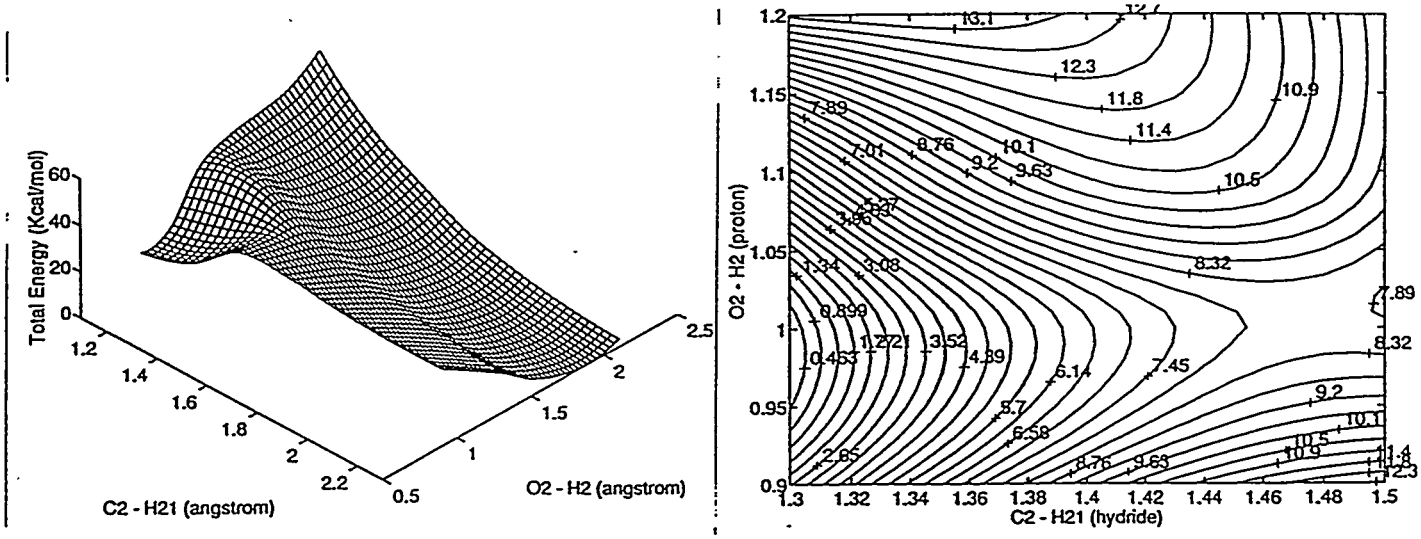

Figure 6. Minimum energy surface (left) for MDH reaction with respect to coordinates d(O2-H2) (Å) and d(C2-H21) (Å). Contour plot (right) of a 0.1 Å grid search in the vacinity of the putative transition state.

From this plot, one can obtain insights into the minimum reaction pathway, energy barriers, and transition states for the MDH enzyme reaction. At this level of theory and grid resolution, the reaction appears to be concerted, with a transition state in the vicinity of d(C2-H21) = 1.5 Å and d(O2-H2) = 1.1 Å

Calculations using a higher resolution, 0.1 Å grid, were made to refine the position of the transition state geometry. The result shown in Figure 6 suggests a transition state at reaction coordinates of d(C2-H21) ≅ 1.48 Å and d(O2-H2) ≅ 1.02 Å. This transition state region can be further characterized by a mathematical procedure that uses the Hessian matrix of the QM/MM energy and an optimization procedure to search for a saddle point on the energy surface for this enzyme reaction. We have implemented the eigenfollower technique of Jon Baker[21] within the context of our QM/MM method. Application of this QM/MM eigenfollower procedure to the region of the energy surface near the transition state suggested by Figure 6 is in progress. Preliminary results are consistent with the grid search.

References

1. Bash P.A., Field M.J., Karplus M., "Free energy perturbation method for chemical reactions in the condensed phase: a dynamical approach based on a combined quantum and molecular mechanics potential", *J. Amer. Chem. Soc.*, **109**, 8092 (1987)
2. Field M.J., Bash P.A., Karplus M., "A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations", *J. Comp. Chem.*, **11**, 700 (1990)

3. Warshel A., Levitt M., "Theoretical studies of enzymatic reactions: dielectric, electrostatic, and steric stabilization of the carbonium ion in the reaction of lysozyme", *J. Mol. Biol.*, **103**, 227 (1976)
4. Singh U.C., Kollman P.A., "A combined *ab initio* quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems", *J. Comp. Chem.*, 7, 718 (1986)
5. Brooks B., Bruccoleri R.E., Olafson B., States D.J., Swaminathan S., Karplus M., "CHARMM: a program for macromolecular energy, minimization, and dynamics calculations", *J. Comp. Chem.*, 4, 187 (1983)
6. Brooks C.L. III, Karplus M., Pettitt B.M. Proteins: a theoretical perspective of dynamics, structure, and thermodynamics. :, 1988 (Adv. Chem. Phys. LXXI; John Wiley & Sons).
7. Bash P.A., Field M.J., Davenport R., Ringe D., Petsko G., Karplus M., "Computer simulation and analysis of the enzyme reaction in triosephosphate isomerase", *Biochemistry*, **30**, 5826 (1991)
8. Gao J., Xia X., "A two-dimensional energy surface for a type II S_N2 reaction in aqueous solution", *J. Am. Chem. Soc.*, **115**, 9667 (1993)
9. Knowles J.R., "Enzyme catalysis: not different, just better", *Nature*, **350**, 121 (1991)
10. Gao J.L., Xia X.F., "*A priori* evaluation of aqueous polarization effects through monte-carlo qm-mm simulations", *Science*, **258**, 631 (1992)
11. Mackerell A.D., Bashford D., Bellot M., Dunbrack R.L., Field M.J., Fischer S., Gao J., Guo H., Ha S., Joseph D., Kuchnir L., Kuczera K., Lau F.T.K., Mattos C., Michnick S., Ngo T., Nguyen D.T., Prodhom B., Roux B., Schlenkrich M., Smith J.C., Stote R., Straub J., Wiorkiewicz-Kuczera J., Karplus M., "Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations", *FASEB Journal*, **6**, A143 (1992)
12. Dewar M.J.S., Zoebisch E.G., "Am1: a new general purpose Quantum Mechanical Model", *J. Am. Chem. Soc.*, **107**, 3902 (1985)
13. Hall M.D., Banaszak L.J., "Crystal structure of a ternary complex of *Escherichia coli* malate dehydrogenase, citrate, and NAD at 1.9 Å resolution", *J. Mol. Biol.*, **232**, 213 (1993)
14. Nicholls D.J., Miller J., Scawen M.D., Clarke A.R., Holbrook J.J., "The importance of arginine 102 for the substrate specificity of Escherichia coli Malate Dehydrogenase", *Biochem. Biophys. Res. Comm.*, **189**, 1057 (1992)
15. Brooks III C.L., Brünger A., "Active site dynamics in protein molecules: A stochastic boundary molecular-dynamics approach", *Biopolymers*, **24**, 843 (1985)
16. Jorgensen W.L., Chandrasekhar J., Madura J., Impey R.W., Klein M.L., "Comparison of simple potential functions for simulating liquid water", *J. Chem. Phys.*, **79**, 926 (1983)
17. Besler B.H., Merz K.M., Kollman P.A., "Atomic charges derived from semiempirical methods", *J. Comp. Chem.*, **11**, 431 (1990)
18. van Gunsteren W.F., Berendsen H.J.C., "Algorithms for macromolecular dynamics and constraint dynamics", *Mol. Phys.*, **34**, 1311 (1977)
19. Schwenke D.W., Truhlar D.G., "Systematic study of basis set superposition errors in the calculated interaction energy of two HF molecules", *J. Chem. Phys.*, **82**, 2418 (1985)
20. Frisch M.J., Del Bene J.E., Binkley J.S., Schaefer III H.F., "Extensive theoretical studies of hydrogen-bonded complexes", *J. Chem. Phys.*, **84**, 2279 (1986)
21. Baker J., "An algorithm for the location of transition states", *J. Comp. Chem.*, 7, 385 (1986)

Multipole-Accelerated Evaluation of Force Fields in Macromolecular Modeling

John A. Board, Jr.
Department of Electrical Engineering
Duke University, Durham NC
jab@ee.duke.edu

September 27, 1994

In molecular dynamics simulation (MD), the N-body problem manifests itself as the need to compute the long-range electrostatic forces acting between all the charged atoms in a molecular system of interest. Straightforward implementation of a Coulomb solver leads to $\mathcal{O}(N^2)$ summation over all pairwise combinations of the N atoms in a system; this quadratic complexity limits the size of systems that can be simulated to a few tens of thousands of atoms. Current MD codes typically truncate the Coulomb interaction at some moderate radius to limit the cost of force evaluation; this can potentially alter the dynamics of the simulated system. The algorithms discussed in this talk permit inclusion of all pair interactions at a runtime cost which grows linearly (or nearly so) in the size of the system.

Though others have contributed important insights to this field, we take as our starting points the "tree code" of Barnes and Hut and the "Fast Multipole Algorithm" (FMA) of Greengard and Rokhlin. Both the tree code and the FMA exploit the observation that the effect of a distant group of charged bodies on a particular charge of interest can be very well approximated by considering a truncated series representation of the potential (and force) due to the distant charges.

We have studied and implemented three different algorithms. Our FMA code follows the original prescription of Greengard and Roklhin, with the additional use of FFT-accelerated multipole manipulations when appropriate [3]. We have also implemented two tree codes, dubbed PMTA (Parallel Multipole Tree Code) and E-PMTA (Enhanced-PMTA). PMTA is similar to the algorithms proposed by Barnes and Hut with a few differences, and E-PMTA hybridizes PMTA with the original FMA [2, 1].

Our package PMTA 4.0 runs on serial workstations and on shared-memory parallel computers such as the Kendall Square KSR-1. A separate distributed

Generation of Models for "Unusual" DNA and RNA: A Computer Language for Structural Exploration

Tom Macke and David A. Case
Department of Molecular Biology
The Scripps Research Institute
La Jolla, California 92037

1. Summary

We have developed a new approach to the modelling of nucleic acids that is implemented as a computer language called *NAB*. The method was primarily designed to construct models of helical and non-helical nucleic acids from a few dozen to a few hundred nucleotides in size, and uses a combination of rigid body transformations and distance geometry to create candidate structures that match input criteria. The language is designed to provide a flexible way to described nucleic acid structures at an atomic level of resolution, and contains built-in connections to the AMBER molecular modelling package, the MEAD programs to compute solvation effects at the Poisson-Boltzmann level, and the AVS visualization system.

2. Background

Many current modeling programs, including those marketed by major vendors like Biosym, Tripos and Polygen, are much better suited for the study of proteins and enzymes than for nucleic acids and their deriviatives. This is partially because computational "drug design" has typically looked more heavily at protein-based targets and peptide-based drugs than at oligonucleotides, partially because it has traditionally been difficult to model poly-ionic nucleic acids in a physically reasonable fashion, and partially because the community has only recently gained a sound understanding of the structural principles behind "unusual" nucleic acids that go beyond base-paired helical structures. Some conventional programs that analyze nucleic acid structures operate at a very detailed level of individual base-pair orientations relative to a helix axis, and are appropriate for the description of deviations from helical geometries [1], but not for more complex structures like hairpins, junctions or pseudoknots. Other nucleic acid modelling algorithms have focussed on supercoiling and other aspects of DNA structure at the level of many hundreds to thousands of base pairs [2]. For several years, we have been working to develop computational tools that will be appropriate for an intermediate range of structures have a few dozen to a hundred or so nucleotides, arranged in a variety of environments. The project began with efforts to understand the structural principles behind four-arm junctions that are models for Holliday-type recombination sites [3], and has since branched out to study hairpins, triplexes and tetraplexes.

The past few years have also seen the development of powerful techniques to estimate energetics of biomolecules in aqueous and non-aqueous environments. Among the most successful of these have been continuum-dielectric models for the response of solvent to charges and dipoles in biomolecular solutes. These models treat the biomolecule as a region of low dielectric, surrounded by a higher-dielectric solvent [4,5]. The Poisson-Boltzmann equation that

describes electrostatic interactions in this system (including "salt" effects arising from mobile counterions in solution) can be solved numerically for aribitrary configurations of solute. This provides a physically realisitic model for the solvation energy as a function of conformation (or binding, etc.) that can be added to more traditional molecular mechanics estimates of energies that arise from molecular mechanics force fields. This ability to model solvent and salt effects on nucleic acid interaction energies promises to provide significant improvements in the realism of energetic evaluations of potential structures.

3. The NAB language

NAB (Nucleic Acid Builder) has been developed by Tom Macke as a part of his graduate research at The Scripps Research Institute. It is a computer language (specified through *lex* and *yacc*) that allows nucleic acid structures to be described in a hierarchical fashion, using a language similar to C or *awk*, but designed especially for the manipulation of nucleic acid structures. *NAB* manipulates nucleic acid components through two principal techniques:

(1) First are base transformations, which are useful in helical or near-helical situations in which the geometric relation of one basepair (or triple) can be specified relative to others in the helix. Under these circumstances, the bases are laid out first to achieve desired helical and base-pairing configurations, and the sugar-phosphate backbone (or derivatives thereof) are added and optimized in a separate step using molecular mechanics energy minimization procedures or distance geometry.

(2) The second pillar of NAB functionality is distance geometry, which allows molecular structures to be built that satisfy sets of distance constraints [6]. Such constaints often form a natural way of describing neighbor relationships, cross-linking or footprinting results, or hydrogen bond and helical constaints in nucleic acids. By systematically exploring databases of known nucleic acid structures [7], we have been able to derive sets of correlated distance constraints that significantly improve the performance of distance geometry techniques as applied to unusual nuclic acid structures. These technques are especially useful in laying out non-helical regions of structures, such as hairpins or loops in pseudo-knot RNA structures.

Fig. 1 gives a simple example of an NAB program that illustrates many of its features. This program sets up potential base triplex structures, and systematically (under program control) searches for the optimal orientation of the bases. The program illustrates that nucleotides and stands, etc. are represented as named objects that are easy to manipulate with *NAB*, and that energetic quantities from molecular mechanics or continuum dielectric models are available as an integral part of the language (in this case, by linking in the AMBER and MEAD programs for molecular mechanics and Poisson-Boltzmann calculations, respectively). The fact that a program loop structure is testing out various options, rather than the more traditional approach of relying upon interactive graphical investigation, means that the results can be readily documented and reproduced by other investigators, and the assumptions of the model can be summarized in compact and testable format. Further, the fact that standard energy analysis programs like AMBER [8] and MEAD [9] can be linked into NAB means that this approach can easily keep pace with scientific advances in compuational chemistry.

4. Examples

In the talk, I will show examples of applications of *NAB* to double- and triple-helical systems, and to hairpins and pseudo-knots. Prospects and progress toward implementation of these methods to larger systems (such as tRNA, the hammerhead ribozyme, and the Group I intron) will be discussed. This language may also be useful for other molecular modelling tasks, and some of its prospects and limitations will be described.

5. References

1. Babcock, M.S., Pednault, E.P.D., and Olson, W.K. (1994) Nucleic acid structure analysis. Mathematics for local cartesian and helical structural parameters that are truly comparable between structures. *J. Mol. Biol.* **237,** 125-156.
2. Malhotra, A., Gabb, H.A., and Harvey, S.C. (1993) Modeling large nucleic acids. *Curr. Opin. Struct. Biol.* **3,** 241-246.
3. Macke, T., Chen, S.-M., and Chazin, W.J. (1992) Analysis of the structure and dynamics of synthetic 32 base-pair models of the Holliday junction intermediate in genetic recombination, in R.H. Sarma and M.H. Sarma, Ed., *Structure and Function, Volume 1: Nucleic Acids*, Adenine Press, Albany. pp. 213-227.
4. Honig, B., Sharp, K., and Yang, A.-S. (1993) Macroscopic models of aqueous solutions: Biological and chemical appications. *J. Phys. Chem.* **97,** 1101-1109.
5. Bashford, D. (1991) Electrostatic effects in biological molecules. *Curr. Op. Struct. Biol.* **1,** 175-184.
6. Crippen, G.M. and Havel, T.F. (1988) "Distance Geometry and Molecular Conformation." Research Studies Press, Taunton, England.
7. Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A.R., and Schneider, B. (1992) The Nucleic Acid Database: A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys J.* **63,** 751-759.
8. Pearlman, D.A., Case, D.A., Caldwell, J.C., Seibel, G.L., Singh, U.C., Weiner, P., and Kollman, P.A. (1991) "AMBER 4.0." University of California, San Francisco,
9. Bashford, D. and Gerwert, K. (1992) Electrostatic Calculations of the pK_a Values of Ionizable Groups in Bacteriorhodopsin. *J Mol. Biol.* **224,** 473-486.

Fig. 1. NAB Program to search for optimal triple helix alignment

```
molecule      m3;
residue             r3;
matrix        r_mat, t_mat;
(...other declarations...)
//
//     create a Watson-Crick base pair
//
m3 = wc_helix(      sbase, "bdna.std.rlb", abase, "arna.std.rlb",
              2.25, 0.0, 36.0, 3.38 );
//
//     create 3rd strand; add 3rd base; save init coords:
//
addstrand( m3, "third" );
r3 = getres( tbase, "bdna.std.rlb" );
addresidue( m3, "third", r3 );
nratoms = getmolxyz( m3, "third::", resxyz );
//
//     Begin energy calc's:
//
cutoff[ 1 ] = 8.0; nbupd[ 1 ] = 1;
ierr = sander_prep_( xyz, box, cutoff );
//
//  start search over the rectangle:
//
min_rz = RZ_UNDEF;
for( x = low_x; x <= hi_x; x = x + x_inc ){
    for( y = low_y; y <= hi_y; y = y + y_inc ){
        b_rz = RZ_UNDEF;
        for( rz = low_rz; rz <= hi_rz; rz = rz + rz_inc ){
        //
        //  position the third strand.....
        //
        setmolxyz( m3, "third::", resxyz );
        r_mat = newtransform(0.0, 0.0, 0.0, 0.0, 0.0, -90.0+rz);
        transformmol( r_mat, m3, "third::" );
        t_mat = newtransform( x, y, 0.0, 0.0, 0.0, 0.0 );
        transformmol( t_mat, m3, "third::" );
        //
        //  .... and calculate its energy:
        //
        natoms = getamberxyz( m3, xyz );
        sander_force_( xyz, force, energy, nbupd );
        }
    }
}
}
//
//  Recreate and write out the best structure:
//
setmolxyz( m3, "third::", resxyz );
r_mat = newtransform(0.0, 0.0, 0.0, 0.0, 0.0, -90.0+min_rz);
transformmol( r_mat, m3, "third::" );
t_mat = newtransform( min_x, min_y, 0.0, 0.0, 0.0, 0.0 );
transformmol( t_mat, m3, "third::" );
putpdb( mfname, m3 );
```

MODELING QUANTUM PROCESSES IN COMPLEX SYSTEMS

David Chandler
Department of Chemistry,
University of California
Berkeley, CA 94720

Abstract

I. Path integrals and the classical isomorphism
II. Computation of rate constants
    A. Classical
    B. Quantal and the centroid approximation
    C. Electron transfer (ET)
        - reaction coordinate, and role of electronic polarization
        - nuclear tunneling
        - harmonic approximation
III. Examples
    A. ET in an aqueous environment
    B. ET in a proteic environment, photosynthetic reaction center

This lecture discusses a few computational techniques based upon the imaginary time path integral formulation of quantum theory [1]. A well known classical-quantal isomorphism results from this formulation [2,3]. The mapping has proved useful in determining from computer simulation the role of quantal dispersion (the uncertainty principle) on the equilibrium properties of water [4]. It has also proved useful in computing the effects nuclear tunneling on electron transfer rates in solution [5,6] and proteins [7,8]. In the case of kinetics, the path integral procedures can be viewed in terms of the so-called "centroid approximation" -- a quantum transition state theory closely related to the instanton approximation [9,10,11]. This view extends the classical correlation function formulation of rate processes and the statistical issues pertaining to the computation of rate constants for rare events [12,13,14]. For long range electron transfer reactions, the centroid approximation coincides very closely with the stationary phase evaluation of the standard golden rule rate formula [15,16].

Much of the lecture will focus specifically on examples of electron transfer, as they illustrate the strengths and limitations of the centroid approach, and are of physical interest as well. The connection with Marcus theory can be made, including a description of the disparate roles of nuclear and electronic polarization [17]. Different approximations to the nuclear dynamics are discussed, such as harmonic approximations including the "disperse polaron" model [18]. The effects of nuclear tunneling are especially pronounced for strongly exothermic electron transfer -- Marcus' inverted regime. Finally, the role of dynamical disorder leading to complex kinetics will be noted [19,20].

The cited references provide detailed descriptions of the techniques covered in this lecture.

References

1. R.P. Feynman and A.R. Hibbs, Quantum Mechanics and Path Integrals (McGraw Hill, New York, 1965).

2. D. Chandler and P.G. Wolynes, J. Chem. Phys. 74, 4078 (1981).

3. D. Chandler, "Theory of Quantum Processes in Liquids," in Les Houches, Part 1, Liquids, Freezing and the Glass Transition, ed. by D. Levesque, J.P. Hansen and J. Zinn-Justin [Elsevier Science Publishers, B.V. (North Holland) 1991], p. 193.

4. R.A. Kuharski and P.J. Rossky, J. Chem. Phys. 82, 5164 (1985).

5. J.S. Bader, R.A. Kuharski and D. Chandler, J. Chem. Phys. 93, 230 (1990).

6. C. Zheng, J.A. McCammon and P.G. Wolynes, Proc. Natl. Acad. Sci. USA 86, 6441 (1989).

7. C. Zheng, J.A. McCammon and P.G. Wolynes, Chem. Phys. 158, 261 (1991).

8. M. Marchi, J.N. Gehlen, D. Chandler and M. Newton, J. Am. Chem. Soc. 115, 4178 (1993).

9. M. J. Gillan, J. Phys. C 20, 3621 (1987).

10. G.A. Voth, D. Chandler and W.H. Miller, J. Chem. Phys. 91, 7749 (1989).

11. A.A. Stuchebrukhov, J. Chem. Phys. 95, 4258 (1991); X. Song and A. A. Stuchebrukhov, J. Chem. Phys. 99, 969 (1993).

12. D. Chandler, J. Chem. Phys. 68, 4213 (1978).

13. D. Chandler, J. Stat. Phys. 42, 49 (1986).

14. D. Chandler, J. Phys C 2, SA9 (1990).

15. R.P. Van Duyne and S.F. Fischer, Chem. Phys. 5, 183 (1974).

16. P.G. Wolynes, J. Chem. Phys. 87, 6559 (1987).

17. J.N. Gehlen, D. Chandler, H.J. Kim and J.T. Hynes, J. Phys. Chem. 96, 1748 (1992).

18. J.-K. Hwang and A. Warshel, J. Am. Chem. Soc. 109, 715 (1987).

19. N. Gehlen, M. Marchi and D. Chandler, Science 263, 499 (1994).

20. J. Wang and P.G. Wolynes, Chem. Phys. 180, 141 (1994).

A SUBSTRUCTURING APPROACH FOR REDUCED VARIABLE MACROMOLECULAR MODELING

H.M. Chun, C.E. Padilla, K.B. Blair, J.H. Li, D.N. Haney, H.E. Alper

This paper presents a new approach to the modeling of macromolecular systems. The method is based on the idea that the essential dynamics of such systems are captured by the low frequency modes of the system (Levy et al, 1984; Ichiye and Karplus, 1991; Horiuchi and Go, 1991; Space et al, 1993; Amadei et al, 1993; Mizuguchi et al, 1994). There are many biological processes that take place in the nano- to milli-second time frames that cannot be modeled with current methods because of the large size of the systems and small time steps needed for numerical integration. By considering only the low frequency behaviors, much larger integration time-steps can be used. Modal approaches also reduce the number of degrees of freedom that need to be modeled. Modal methods have been used in the past to study macromolecular systems. However, because of the linearization that is introduced in order to obtain the mode shapes and frequencies, such models are valid only for a small region near the conformation about which the linearization was performed.

The approach that we have developed is based on the substructuring of a large molecular system into bodies and particles. The substructuring is determined by the amount of motion to be expected between atoms. For regions where motions are expected to be very small, or small enough to be unimportant for the purposes of the simulation, the atoms can be grouped together into rigid bodies. Regions where there are moderate amounts of motion can be modeled as flexible bodies. Regions where large conformational changes are expected can remain atomistic. Since this modeling approach allows large motions between bodies, as well as between individually modeled atoms, it is expected to be valid for a much larger region in conformational space.

Secondary structures of proteins are primary candidates for grouping into bodies. Alpha helices and beta sheets are naturally thought of as having collective motions. In fact, the analysis reported in the paper by Swaminathan et al, 1991 has demonstrated this. Loop and turn regions can be modeled atomistically to allow large conformational changes. Parts of these regions might also be grouped into flexible bodies. Because of the distinct separation of the system into high frequency atomistic regions and low frequency flexible/rigid bodies, this modeling approach is highly amenable to treatment by multiple time scale integration techniques.

The eigensolution process is more tractable for this substructured model because the mode shapes and frequencies are calculated separately for individual bodies, rather than for the entire system. It is computationally less expensive to compute eigensolutions of component bodies within a system than it is to compute the eigensolution of the entire system.

This substructured modeling approach is adapted from spacecraft and mechanical dynamics modeling techniques, where large relative motions are allowed between

articulated bodies which may be either rigid or flexible. Elastic behavior of the individual bodies are modeled by component modes. Constraints at hinges/joints are handled in an exact manner, and there have been recent developments (Chun et al, 1987; Bae and Haug, 1987; Rodriguez et al, 1991; Jain et al, 1993) that have resulted in fast algorithms with Order (n) computational complexity. We have modified these well-understood techniques to handle molecular systems, and have created a code (Turner, Weiner, Chun et al, 1993) called MBO(N)D (Multi-Body Order (N) Dynamics). Exact rigorous bond length and bond angle constraints have been implemented into this code. A hierarchical body-based multipole algorithm has also been implemented for electrostatic interactions so as to take advantage of the fact that invariant multipole coefficients can be computed if they are based on body-fixed coordinate frames.

There are several advantages of this substructured modeling technique. The elimination of high frequency content from aggregated groups of atoms allows larger time-steps, and hence longer time frame simulations or larger molecular systems. The ability to manipulate groups of atoms collectively as bodies allows the modeling and simulation of events that cannot be treated by all-atom models. Examples include flap motion of HIV protease, sub-unit interactions in hemoglobin, and dynamic docking of ligand and substrate. The exact constraint formulation allows the relative motion between bodies to be optionally specified as a function of time, thereby allowing inverse dynamics simulations to be performed based on hypothetical or experimentally observed molecular motions.

Within this substructured modeling approach, there are several alternatives for the calculation of bond and non-bond interactions. Conventional all-atom calculations (Brooks et al, 1983) can be performed. Body forces and torques are obtained by summing up atomistic forces and moments over the atoms that make up each body. Modal forces are computed by multiplying atomistic forces by the mode shapes. This projects the physical forces into the low frequency subspace of the body. A modified approach replaces the body internal interactions by modal stiffness terms. The bond and non-bond pairlists are reduced, resulting in a more efficient calculation. Fast multipole algorithms (Greengard and Rokhlin, 1987, 1989; Ding, Karasawa, and Goddard, 1992) could be applied to speed up non-bond calculations for large systems. As noted above, we have developed a body-based multipole algorithm for better computational efficiency at high levels of aggregation.

The outputs from the substructured modeling approach are of the same type as that from all-atom simulations. This is because the coordinates and velocities of every atom are known once the translations, rotations, and modal amplitudes of the atom's parent body are found. Thus, conventional post-simulation analysis algorithms can be directly applied to these simulations.

The substructured modeling approach provides the framework for dealing with collective motions, models with varying degrees of fidelity (higher fidelity near active site, lower fidelity elsewhere), and a way of treating large systems and long time-frame systems in a computationally tractable manner. The MBO(N)D code is currently in various stages of development and integration with AMBER 4.0, CHARMM, and X-PLOR.

Numerical results will be presented that demonstrate feasibility and potential speedups introduced by this substructuring approach.

LITERATURE CITED

Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C., *Proteins*, **17**, 412 (1993)
Bae, D.-S., and Haug, E.J., *Mechanics of Structures and Machines*, **15**, 359-382 (1987)
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M., *J. Comp. Chem.*, **4**, 187 (1983)
Chun, H. M., Turner, J. D., and Frisch, H. P., Paper AAS-89-457, AAS/AIAA Conf., Stowe, VT (1987)
Ding, H.-Q., Karasawa, N., and Goddard III, W.A., *J. Chem. Phys.*, **97**, 4309, (1992)
Greengard, L., and Rokhlin, V., *J. Comp. Phys.*, **73**, 325 (1987)
Greengard, L., and Rokhlin, V., *Chem. Scr.*, **29A**, 139 (1989)
Horiuchi, T., and Go, N., *Proteins*, **10**, 106 (1991)
Ichiye, T., and Karplus, M., *Proteins*, **11**, 205 (1991)
Jain, A., Vaidehi, N., and Rodriguez, G., *J. Comput. Phys.*, **106**, 258-268 (1993)
Levy, R. M., Karplus, M., Kushick, J., and Perahia, D., *Macromolecules*, **17**, 1370 (1984)
Mizuguchi, K., Kidera, A., and Go, N., *Proteins*, **18**, 34 (1994)
Rodriguez, G., Jain, A., and Kreutz, K. *International Journal of Robotics Research*, **10**, 371-381 (1991)
Space, B., Rabitz, H., and Askar, A., *J. Chem. Phys.*, **99**, 9070, (1993)
Swaminathan, S., Harte, W. E., and Beveridge, D. L., *J. Am. Chem. Soc.*, **113**, 2717 (1991)
Turner, J. D., Weiner, P. K., Chun, H. M., Lupi, V., Gallion, S., and Singh, U. C., Ch. 24., Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, **2**, Gunsteren, W.F., Weiner, P.K., and Wilkinson, A. J., eds., ESCOM, Leiden (1993)

MOLECULAR DYNAMICS WITH THE LOCALLY ENHANCED SAMPLING METHOD: APPLICATIONS TO STRUCTURE AND DYNAMICS OF BIOMOLECULES

Ron Elber,[#,] Carlos Simmerling[*] and Alex Ulitsky,[*]*
The University of Illinois at Chicago[*] and The Hebrew University of Jerusalem[#]

During the last few years we developed and applied a mean field approach called LES (Locally Enhanced Sampling) as a versatile computational tool for treating a variety of problems, including exploration of ligand diffusion pathways[1,2,4], global optimization[6-8], free energy calculations[9] and approximate simulation of ligand-protein dynamics[3,5].

Two separate computational projects that employ the LES protocol will be discussed in the lecture: (i) atomic detail calculations of structures of peptides in explicit solvent, and (ii) simulation of *long time* diffusion of ligands through proteins.

STRUCTURE OF PEPTIDES

It is not at all obvious that short peptides have a unique structure, and in fact many of them do not. However if a unique structure exists, it is of significant interest for a number of reasons:

First, many short peptides transfer a "message" to a receptor. As such, they are the target of pharmaceutical research aiming to identify their stable or biologically active conformation. A known conformation of a biologically active peptide is a useful lead to the design of a new drug.

Second, peptides can be used to test theoretical models. In the last few years there was considerable progress in determining "low resolution" structures of proteins[10-11]. Calculations of optimal structure of peptides (ten to twenty residues) can be pursued using either (a) an atomic level calculation with explicit solvent or (b) reduced models of amino acids. This opens the possibility to examine the accuracy of a reduced model and its applicability to structure determination of *peptides*. We should like to emphasize that data bases with known structures of peptides are very limited and therefore (in contrast to proteins), one is forced to use energy related techniques in structural modeling.

The third motivation for studying the conformations of short peptides is the search for nucleation sites in protein folding.

Levinthal raised the "kinetic" folding problem[12] of how the protein searches for the "correct" configuration through the very large number of possible states. A plausible mechanism for getting around that problem is the introduction of small structural segments that are formed at early phases of the folding. The formation of "nucleation sites" at early times reduces considerably the space that needs to be explored and can solve (in principle) the Levinthal paradox. It is well established that some protein fragments have a significant tendency to assume their correct fold even when the rest of the protein is not folded correctly or even missing[13].

I shall discuss how the LES protocol is used to investigate structures of small peptides as models for folding nucleation sites. The LES approach enables converged and systematic structural investigations of peptides in explicit water. I shall also demonstrate that regular annealing or straightforward molecular dynamics simulations do not converge in the peptides that we investigated. I shall describe calculations done on the peptides CAAAAC, CHDLFC, CSVTC and RVEW. In these investigations we demonstrate the importance of hydrophobic forces in formation of structure in small peptides. The two peptides CHDLFC and REVW form hydrophobic clusters in water solutions. We further show that a water molecule at a specific site can form a bridged hydrogen bond and provides further drive to a unique structure. This is in the peptide CSVTC. As a general conclusion, in all our investigations of short peptides, internal hydrogen bonding makes a minor contribution to the formation of stable structure in solution. Only RVEW has one, partially populated, internal hydrogen bond.

LES APPLICATION TO DYNAMICS

The second part of the talk will focus on simulations of dynamics in non-equilibrium systems. This part is relevant to the diffusion of small ligands in and out of protein matrices[14-19]. The problem of how the ligand penetrates to the buried active site in a globin was discussed first by Perutz[14], and attracted substantial volume of experimental and theoretical work. Recent experimental developments include on one hand advances in physical instrumentation - the development of faster lasers and improvement in detection techniques[18,19] and on the other hand application of genetic engineering methods - the use of mutants to perturb the system[21,22]. In parallel, computational studies benefited from significant progress in computer technology (faster computers) and from the introduction of new simulation methods[1,16,23,24].

The LES makes it possible to run a number of ligand trajectories that share the same protein matrix. This provides more statistics on ligand pathways at a cost comparable to one trajectory. However, LES is an approximation (mean field) when applied to dynamics. This prevents us from studying absolute time scales. We have developed [5,25] a new variant of the mean field that includes a binary collision correction (cLES). This variant provides a significant improvement with respect to the previous mean field results when applied to dynamics. The diffusion constants calculated from exact trajectories and from the new variant are essentially the same[5]. Application will be discussed of the mean field with a binary collision correction to diffusion and recombination of nitric oxide in myoglobin.

We show that the approximate simulation methodology reproduces experimental trends of ligand recombination to protein mutants as well as reproducing diffusion constants obtained from usual molecular dynamics simulation. We further explore plausible diffusion pathways of the ligand and discuss the connection to some recent experiments.

Literature Cited

1. R. Elber and M. Karplus, "Enhanced sampling in molecular dynamics: Use of the time dependent Hartree approximation for a simulation of carbon monoxide through myoglobin", J. Amer. Chem. Soc. 112,9161(1990)
2. R. Czerminski and R. Elber, "Computational studies of ligand diffusion in globins: I. leghemoglobin", Proteins 10,70(1991)
3. Quentin H. Gibson, Rebecca Regan, Ron Elber, John S. Olson and Theodore E. Carver, "Distal Pocket Residues Affect Picosecond Recombination in Myoglobin: An experimental and Molecular Dynamics Study of Position 29 Mutants", J.Biol.Chem. 267,22022(1992)
4. E. Chiancone, R. Elber, W.E. Royer, Jr., R. Regan and Q.H. Gibson,"Ligand binding and conformation change in the dimeric hemoglobin of the clam Scapharca in aequivalvis", J.Biol.Chem. 268,5711(1993)
5. A. Ulitsky and R. Elber, "Application of the Locally Enhanced Sampling (LES) and a mean field with a binary collision correction (cLES) to the simulation of Ar diffusion and NO recombination in myoglobin", J.Phys.Chem., 98,1034(1994)
6. A. Roitberg and R. Elber, "Modeling side chains in peptides and proteins: Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations", J.Chem.Phys. 95,9277(1991)

7. A. Roitberg and R. Elber, "Modeling side chains in peptides and proteins with the locally enhanced sampling/simulated annealing method.", a chapter in "Protein structure determination", Ed. K.M. Mertz and S. Le Grant, Birkhauser Pub., in press
8. C. Simmerling and R. Elber, "Hydrophobic "collapse" in a cyclic hexapeptide: Computer simulations of CHDLFC and CAAAAC in water", J.Amer.Chem.Soc., 116,2534(1994)
9. G. Verkhivker, R. Elber and W. Nowak,"Locally Enhanced Sampling in Free Energy Calculations: Application of Mean Field Approximation to Accurate Calculations of Free Energy Differences", J.Chem.Phys., 97,7838(1992)
10. J. Skolnick and A. Kolinski, "Computer simulations of globular protein folding and tertiary structure", Annu.Rev.Phys.Chem. 40,207(1989)
11. M.S. Friedrich, R.A. Goldstein and P.G. Wolynes, "Generalized protein tertiary structure recognition using associative memory hamiltonian", J.Mol.Biol. 222,1013(1991)
12. C. Levinthal, "Mossbauer Spectroscopy in Biological Systems" (proceedings of a meeting held in Allerton House Monticello, Illinois) University of Illinois Press 1969
13. P.S. Kim and R.L. Baldwin, "Intermediates in the folding reactions of small proteins", Ann. Rev. Biochem. 59,631(1990)
14. M.F. Perutz and F.S. Mathews, "An X ray study of azide met hemoglobin" J.Mol.Biol. 21,199(1965)
15. D.A. Case and M. Karplus, "Dynamics of ligand binding to heme proteins", J.Mol.Biol.132,343(1979)
16. J. Kottalam and D.A. Case, "Dynamics of ligand escape from the heme pocket of myoglobin" J.Amer.Chem.Soc. 110,7690(1988)
17. R.F. Tilton Jr., U.C. Singh, S.J. Weiner, M.L. Connolly, I.D. Kuntz Jr., P.A. Kollman, N. Max, D.A. Case, "Computational studies of the interaction of myoglobin and xenon", J.Mol.Biol. 192,443(1986)
18. H. Frauenfelder, F. Parak and R.D. Young, "Conformational substates in proteins" Annu. Rev. Biophys. Biophys. Chem. 17,451(1988)
19. M.D. Chatfield, K.N. Walda and D. Magde. "Activation parameters for ligand escape from myoglobin proteins at room temperature", J. Amer.Chem. Soc. 112,4680(1990)
20. J.W. Petrich, J.C. Lambry, K. Kuczera, M. Karplus, C. Poyart and J.L. Martin, "Ligand binding and protein relaxation in heme proteins: A room temperature analysis of NO geminate recombination", Biochemistry 30,3975(1991)
21. B.A. Springer and S.G. Sligar, "High level expression of sperm whale myoglobin in escherichia coli", Proc. Natl. Acad. Sci. USA 84,8961(1987)
22. D.G. Lambright, S. Balasubramia, S.G. Boxer, "Ligand and proton dynamics in recombinant human myoglobin mutants", J.Mol.Biol. 207,289(1989)

23. W. Nowak, R. Czerminski and R. Elber, "Reaction path study of ligand diffusion in proteins: Application of the Self Penalty Walk (SPW) method to calculate reaction coordinates for the motion of CO through leghemoglobin", J. Amer. Chem. Soc. 113,5627(1991)
24. G. Verkhivker, Ron Elber and Quentin H. Gibson, "Microscopic modeling of ligand diffusion through the protein leghemoglobin: Computer simulations and experiments", J.Amer.Chem.Soc, 114,7866(1992)
25. A. Ulitsky and R. Elber, "The thermal equilibrium aspects of the time dependent Hartree and the Locally Enhanced Sampling approximations: Formal properties, a correction and computational examples for rare gas clusters", J. Chem. Phys. 98,3380(1993)

A Three-Dimensional Fast Method for the Computation of Electrostatic Energy in Numerical Simulation of Polyelectrolyte DNA

Marcia O. Fenley, Kiat Chua, Alexander H. Boschitsch & Wilma K. Olson
Department of Chemistry, Rutgers, the State University of New Jersey, New Brunswick, NJ 08903; Continuum Dynamics, Inc., Princeton NJ 08543

We present a three-dimensional fast method for the computation of long-range electrostatic interactions in computer simulation of polyelectrolyte DNA. Classically, the computation of electrostatic energy involves a direct summation of all pairwise interactions due to all phosphate groups of DNA. This results in a N-body interaction problem with an asymptotic time complexity of $O(N^2)$. This is computationally very expensive and limits the number of phosphate groups used in numerical simulations of polyelectrolyte DNA to a few hundred. We present a new computational technique that reduces the asymptotic time complexity to O(N). This is achieved by grouping phosphate groups using multipole and Taylor series expansions. The accuracy and speed enhancements of the new method in the computation of the electrostatic energy of circular DNA are studied under the conditions of no added salt and high salt. The fast method is further employed in a Monte Carlo/simulated annealing simulation of closed circular polyelectrolyte DNA. In all cases, order of magnitude improvement in the computational speed is observed with no loss in numerical accuracy. This work was supported by SBIR Phase I Grant No. 1 R43 GM50132-01.

MOLECULAR DYNAMICS OF PROTEINS: WHAT HAS BEEN ACHIEVED, AND WHAT PROBLEMS STAND IN OUR WAY?

Jan Hermans

Department of Biochemistry and Biophysics
Univ. of North Carolina
Chapel Hill, NC 27599-7260

A review of the current status of molecular dynamics simulations, with focus both on methods and on applications in structure and function of proteins. A biophysicist's perspective on the importance of development of new algorithms to deal with the problems of being able to do insufficiently long simulations, and of having to use inaccurate and incomplete force fields.

THE POISSON-BOLTZMANN EQUATION

Michael Holst

Applied Mathematics and CRPC at CalTech

In this talk we consider the nonlinear Poisson-Boltzmann equation, which describes the electrostatic potential of a large complex biomolecule lying in a solvent. This equation has several interesting features impacting both theoretical analysis and numerical solution algorithms, including discontinuous coefficients, rapid nonlinearities, and three spatial dimensions. We develop robust and efficient numerical methods based on multigrid and domain decomposition methods combined with global inexact-Newton methods. Some convergence results are presented using a Schwarz theory framework. Numerical results, illustrating the effectiveness of this method compared to other methods, are presented for several test problems, including superoxidedismutase and the HIV protease. We finish with a look at similar equations arising in gravitation physics.

RECENT DEVELOPMENTS IN THE APPLICATION OF CLASSICAL ELECTROSTATICS TO PROTEINS AND NUCLEIC ACIDS

Barry Honig
Department of Biochemistry and Molecular Biophysics
Columbia University
630 West 168 St.
New York, NY 10032

Classical electrostatics as described by the Poisson-Boltzmann (PB) equation offers an accurate and computationally efficient approach to the study of molecules in aqueous solution. The PB equation is given by

$$\Delta.[\varepsilon(\mathbf{r})\Delta.\phi(\mathbf{r}) - \varepsilon\kappa^2 \sinh[\phi(\mathbf{r})] + 4\pi q\rho^f(\mathbf{r})/kT = 0 \qquad (1)$$

where $\phi(\mathbf{r})$ is the dimensionless electrostatic potential in units of kT/q, k is the Boltzmann constant, T is the absolute temperature, q is the proton charge, ε is the dielectric constant, and ρ^f is the fixed charge density. The term $\kappa^2 = 1/\lambda^2 = 8\pi q^2 I/ekT$, where λ is the Debye length and I is the ionic strength of the bulk solution. The variables $\phi, \varepsilon, \kappa$ and ρ are all functions of the position vector $\mathbf{r}$.

In order to apply the PB equation to molecules of arbitrary shape and charge distribution, it is necessary to map molecular properties into the language of classical electrostatics, that is into the variables that appear in Equation 1. This is accomplished by describing the solute molecule as a set of point charges embedded in a low dielectric object which is surrounded by a material of a different dielectric constant which may contain mobile ions. Once the mapping is accomplished, it is necessary to solve the PB equation for the complex shape and charge distribution defined by the molecule in question. This is carried out in two distinct steps. The first involves formulating the set of equations to be solved and the second is their solution. The first process, for non-regular geometries, involves a discretization of some sort, for example mapping variables onto a grid in finite difference methods. The precision and computational requirements of all methods depends critically on the representation used to define the molecular surface. In the past few a years a number of analytical and numerical techniques have been reported which can describe molecular surfaces, even for complex molecules such as proteins, in a few seconds CPU time on standard processors.

The second process is usually an iterative procedure where an initial guess to a solution is refined successively.

Significant progress has occurred in this area as well. In recent years, a variety of numerical methods have yielded increasing faster and more accurate algorithms. These include successive over-relaxation, conjugate gradient techniques, multigridding, integral boundary element methods, and finite element. Finite difference methods have proved so far to be faster than finite element although the latter offer the possibility of greater accuracy. Recently, we have enhanced the speed of the finite element approach through the use of fast multipole methods.

The most up-to-date version of our own program, DelPhi, employs an efficiently coded multigridding algorithm developed in collaboration with Michael Holst. The mapping of a protein molecule onto 65^3 three dimensional grid and the subsequent numerical solution to the PB equation take a total of about 5 seconds CPU time on an Indigo computer. Thus it is now possible to calculate the electrical potential of a molecule in a solution of an arbitrary ionic strength while working interactively at a graphics workstation.

The presentation will present an overview of recent numerical developments as well as describe new graphical methods used to represent electrostatic potentials. Applications will include salt effects on substrate binding to nucleic acids, combined solvation/quantum mechanical models, conformational free energies of peptides and pH effects on protein stability.

INCORPORATION OF THE SURFACE TENSION INTO SYSTEM BOUNDARY CONDITIONS FOR SIMULATION OF A FLUID PHASE LIPID BILAYER MEMBRANE

Chiu, S.-W.[1], Clark, M.[2], Balaji, V.[1], Subramaniam, S.[1,3,4,5], Scott, H. L.[2], and E. Jakobsson[1,3,4,5]

[1]National Center for Supercomputing Applications, University of Illinois, Urbana, IL 61801, USA, [2]Department of Physics, Oklahoma State University, Stillwater, OK 74078-0444, USA [3]Department of Physiology, [4]Center for Biophysics and Computational Biology, [5]Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana, IL 61801, USA

ABSTRACT

Modeling membranes is not just modeling another kind of macromolecule, but modeling an entire environment for a large class of biomolecular processes. The membrane modeling poses quite a different set of technical problems and scientific issues from modeling proteins. This paper reviews some of these issues and suggests approaches that seem promising for resolving them, based on work in our laboratories and that of others.

INTRODUCTION

Because the membrane in its biological state is a liquid crystal, meaningful and useful models must be dynamic rather than static. It is possible to put phospholipids into crystalline solids (Hauser et al, 1981), but the biological state of membranes is far from this solid crystalline state. Thus molecular dynamics, Monte Carlo, or stochastic dynamics methods are mandatory to describe biological membranes. This paper will focus on several important issues that arise in molecular dynamics simulations of membranes. The choice of the model system, i.e. number of lipid and solvent molecules, choice of boundary conditions, nature of interaction potentials and the method for equilibration bear significantly on the validity of computation and we will address these issues with a view to providing a well-defined set of validation criteria for membrane simulations.

Computations of membrane patches have been reported with stochastic boundary conditions (Heller et al, 1993), with periodic boundary conditions at constant volume (Venable et al, 1993; Alper et al, 1993; Damodaran and Merz, 1994), and with periodic boundary conditions at constant pressure (Berendsen et al, 1992, Chiu et al, 1992). Each choice is fraught with trade-offs. The appropriate choice may depend on the experimental conditions under consideration.

The stochastic boundary conditions suffer from being essentially artificial, especially in that they may inhibit chain tilting of the boundary lipid, which will in turn affect tilting in the rest of the simulation cell. In addition they fail to simulate the essentially infinite bilayer system. However they have the advantage that, with a finite delimited computational space, it is feasible to eliminate cut-offs of electrostatic forces within the computational space by use of a multipole method (Board et al, 1992). But it is not clear that electrostatic forces longer than a reasonable cut-off distance, 15 angstroms or so, are a significant modulator of membrane dynamics and structure. However it probably is important to use neutral group-based rather than atom-based cutoffs. Atom-based cutoffs are likely to introduce spurious polarizations, based on single atoms of a group being on different sides of a cut-off distance. In using group based cut-offs, each water and phospholipid molecule may be treated as a group (Damodaran and Merz, 1994). However it is more economical in non-bonded interactions to divide the phospholipid molecule into several neutral subgroups. This can generally be done with only minor adjustments to the partial charges, well within their inherent range of uncertainty. If it is deemed desirable to eliminate cut-offs altogether, a possible reasonably efficient technique to eliminate cut-offs in

simulations with periodic boundary conditions is the particle mesh method for Ewald sums (Darden et al, 1993).

Periodic boundary conditions are natural for the membrane system. A concern however is that periodic boundary conditions may introduce spurious collective motions, for example an excessive collective tilt of the hydrocarbon chains. This effect might be mitigated by making the box size large enough that the boundaries would have little effect on the interior of the box, but of course this is at the cost of less economical computation. If the periodic boundary conditions are to be used at constant volume, it is necessary to be confident of the density of material in the lipid and at the lipid-electrolyte interface. This is possible in a multi-lammelar preparation of known water content, such as the DOPC membrane that is the subject of a series of papers from the laboratory of White (Wiener and White, 1992). In this case constant volume calculations would seem justified. However in simulating a situation where there is excess water it may be advisable to do the computations at constant pressure. In this method pressure components in each spatial dimension are calculated according to a virial expression, and the dimensions of the box are (slowly) adjusted during the simulation until the mean internal virial matches the external applied pressure, which is set as a boundary condition (Berendsen et al, 1984). This method, with the external pressure at one atmosphere, has been used by the laboratory of Berendsen (Berendsen et al, 1992) and by us (Chiu et al, 1992). Recently in our laboratories we have been using a variant of this method in which the lateral pressure (in the plane of the membrane), rather than being set to the external laboratory pressure of one atmosphere, is set according to the measured surface tension at the membrane-water interface. This surface tension can be derived from the pressure-density curve of a monolayer in a Langmuir trough and the surface tension for a pure air-water interface. (For a clear discussion of surface tension at the membrane-water interface, see pp. 75-78 of Gennis, 1989). For monolayers at the area/lipid characteristic of the fluid phase, this surface tension is of the order of a few tens dyne/cm. Thus the lateral pressure for the simulation is negative and, depending on the thickness of the simulated interface, of the order of a hundred atmospheres. These boundary conditions are shown in schematic form in Figure 1(next page).

The force fields for the lipid and water molecules are also a matter of some concern. In a previous constant pressure simulation at one atmosphere, it was found that the partial charges in the lipid molecules had to be reduced in order to produce a fluid phase in a simulated DPPC membrane (Berendsen et al, 1992). In the later part of this paper it will be seen that a constant pressure simulation as represented in Figure 1 can produce an appropriate fluid phase in a simulated PC membrane with full charges on the lipid molecules.

RESULTS OF A SIMULATION OF A DMPC BILAYER USING ANISOTROPIC CONSTANT PRESSURE BOUNDARY CONDITIONS

We have recently done in our laboratory a simulation of a DMPC bilayer in a fluid state using the constant-pressure boundary conditions shown in Figure 1. A full description of this work will be presented in a journal more suitable for specialized membrane biophysics. A summary of the computation and results is as follows:

The partial charges on the phospholipid molecules were calculated by ab initio electronic structure computations using the GAUSSIAN 92 program at Hartree-Fock SCF level using the 6-31G* basis set. The geometry of the lipid was taken to be that of the x-ray crystal structure. The partial atomic charges were extracted from the SCF total electron density by Mulliken population analysis. Other aspects of the phospholipid force fields were as in a previous simulation of a DPPC bilayer (Egberts, 1988), including the use of the Ryckaert-Bellemans potential for the dihedral angles of the hydrocarbon chains. Unified atom representations were

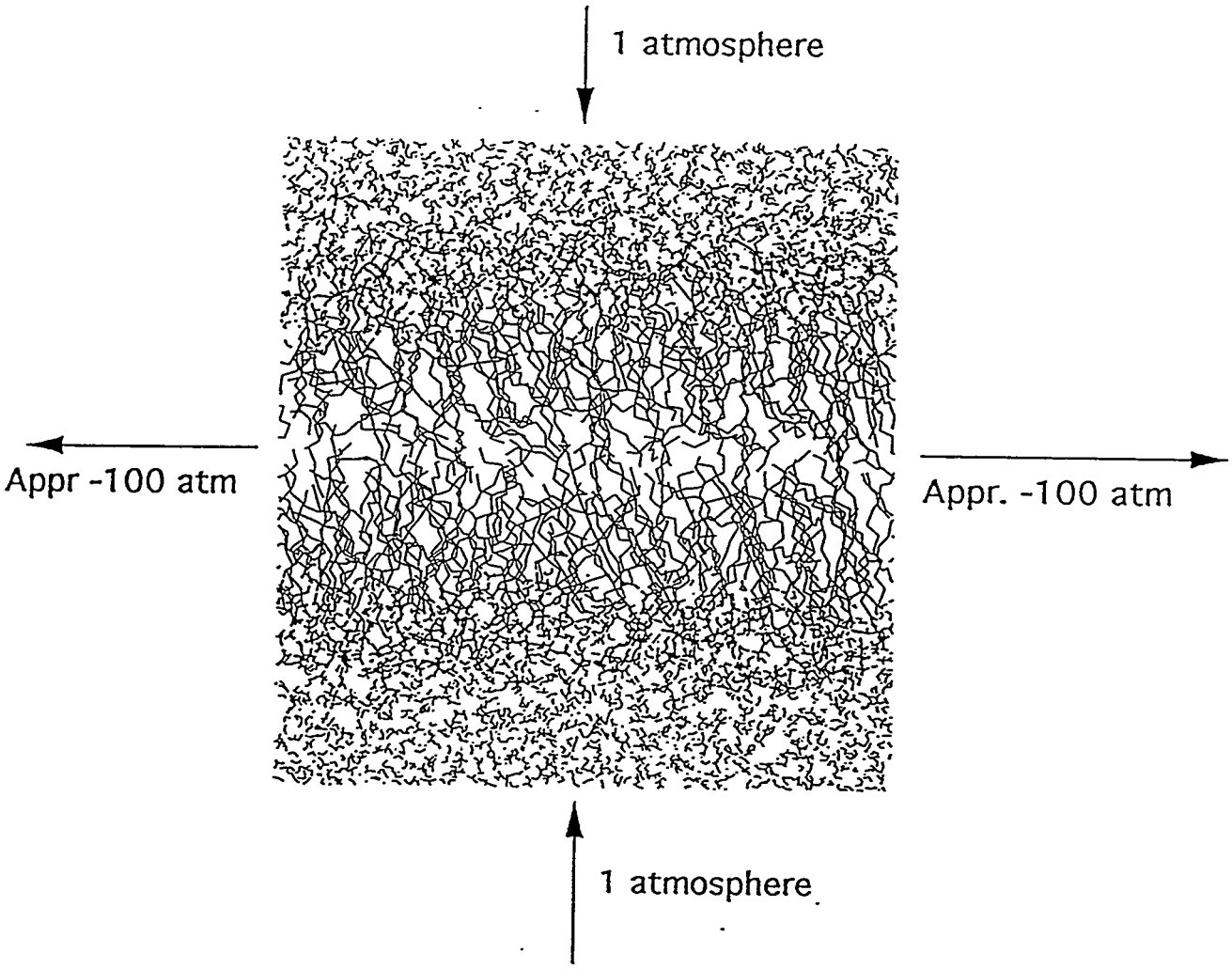

Figure 1. Schematic representation of the constant pressure boundary conditions for computing a fully hydrated fluid phase membrane. The one atmosphere positive pressure normal to the membrane plane is derived from the laboratory conditions. The negative pressure in the membrane plane is derived from the measured surface tension of a phospholipid monolayer.

---

used for the carbons and hydrogens in the hydrocarbon chains. The SPC/E water model was used (Berendsen et al, 1987). Neutral group-based cutoffsof 20 angstroms were used for the non-bonded interactions. The simulation contained 100 phospholipid molecules, 50 in each monolayer, and 2100 water molecules. The starting configuration of the lipid was the crystal structure of Hauser et al (1981), which has a membrane area per lipid of about 38 square angstroms. The pressure tensors were calculated by an internal virial (Berendsen et al, 1984). A coupling time constant of 0.4 psec was used. The system was gradually heated to 325 K. After warming, the system had spread to a cross-sectional area of about 56 square angstroms per lipid molecule, and appeared to stabilize in its dimensions at that value. One change was made in the computational method after heating. When the system reached 325 K, it was seen that about 90% of the hydrocarbon chain dihedrals remained in the trans configuration. At this point the van der Waals interactions were removed for the 1-4 atoms in the hydrocarbon chains, and the simulation was restarted. After this change, the system relaxed to a trans/gauche ratio of about 3/1, appropriate for a fluid phase.

The hydrocarbon order parameters were in good agreement with those expected for a fluid phase PC membrane. Although the actual area of the system was 56 square angstroms/lipid, applying the formula of Nagle (1993) to our order parameters indicated an area of 62 square angstroms. The distribution of particular groups in the direction normal to the membrane is quite similar to that seen experimentally by neutron and x-ray diffraction (Wiener and White, 1992). However it must be added that the neutron and x-ray diffraction experiments were done at a lower hydration, so the measurements may not be strictly comparable to our simulations.

The major discrepancy between our simulated results was the size of the computed dipole potential. This is an electrical potential of the hydrocarbon interior of the membrane positive relative to the external electrolyte (Flewelling and Hubbell, 1986). It is largely due to orientations of the water molecules in the interfacial region, although the lipid charges make a contribution as well. In the case of the PC membrane, the contribution of the lipid is negative, because of the charge distribution in the headgroup in which the choline group carries a net postive charge and the phosphate a net negative charge. Experimentally the net dipole potential should be of the order of a few hundred millivolts. In our simulations we found a computed dipole potential of a few volts, about an order of magnitude too large. It appears our water orientations were too extreme. At this writing we believe this may be due to our periodic boundary conditions combined with not quite enough water in the system to meet the condition of excess hydration, but the precise explanation awaits further computation and analysis.

SUMMARY

Our results show for the first time that a simulated fluid phase phase membrane can be produced with parameters and boundary conditions that are derived from independent experiments and theory, rather than from parameters and boundary conditions that are explicitly designed to give the fluid phase. This has not previously been done. In previous fluid phase constant pressure simulations, the partial charges on the lipid atoms were arbitrarily reduced to produce the fluid phase (Berendsen et al, 1992). In constant volume simulations, the lipid density has been set at a low enough value to ensure the fluid phase, and initial configurations have been randomized (Alper et al, 1993; Venable et al, 1993; Damodaran and Merz, 1994). In our simulations the fluid phase emerged from lipids with partial charges that were computed by ab initio calculations, with the whole system under constant pressure boundary conditions derived from measured surface tension. Further our fluid phase was produced from the crystal structure as the initial conformation, to ensure that the system had no previous bias to be fluid.

It now seems reasonable to consider some future applications of simulated fluid membranes in attacking molecular engineering problems of biological significance.

ACKNOWLEDGEMENTS

We gratefully acknowledge grant support from the National Science Foundation and the use of the facilities of the National Center for Supercomputing Applications. Major computing time was granted from the National Science Foundation Supercomputing Metacenter, for time on the C90 at the Pittsburgh Supercomputing Center and the Convex C3 at the National Center for Supercomputing Applications.

BIBLIOGRAPHY

Alper, H. E., Bassolino-Klimas, D., and T. R. Stouch. 1993. The limiting behavior of water hydrating a phospholipid monolayer: A computer simulation study. *J. Chem. Phys.* 99:5547-5559

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684-3689

Berendsen, H. J. C., Grigera, J. R., and T. P. Straatsma. 1987. The missing term in effective pair potentials. *J. Chem. Phys.* 91:6289-6291

Berendsen, H. J. C., Egberts, B., Marrink, S.-J., and Ahlstrom, P. 1992. Molecular dynamics simulations of phospholipid membranes and their interaction with phospholipase A_2 pp. 457-470 in *Membrane Proteins: Structures, Interactions and Models* A. Pullman, J. Jortner, and B. Pullman, eds. Kluwer Academic Publishers, The Netherlands

Board, Jr., J. A., Causey, J. W., Leathrum, Jr., J. F., Windemuth, A., and K. Schulten. 1992. Accelerated molecular dynamics simulation with the parallel fast multipole algorithm. *Chem. Phys. Lett.* 198:89

Chiu, S. W., Gulukota, K. and E. Jakobsson. 1992. Computational approaches to understanding the ion channel-lipid system. pp. 315-338 in *Membrane Proteins: Structures, Interactions, and Models*. A. Pullman, J. Jortner, and B. Pullman, eds. Kluwer Academic Publishers, The Netherlands

Damodaran, K. V. and K. M. Merz. 1994. A comparison of DMPC- and DLPE-based lipid bilayers. *Biophys. J.* 66:1076-1087

Darden, T., York, D., and L. Pedersen. 1993. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98: 10089-10092

Egberts, E. 1988. Molecular dynamics simulations of multibilayer membranes. Ph.D. Thesis, University of Groningen

Flewelling, R. F., and W. L. Hubbell. 1986. The membrane dipole potential in a total membrane potential model. Applications to hydrophobic ion interactions with membranes. *Biophys. J.* 49:541-552

Gennis, R. B. 1989. *Biomembranes. Molecular Structure and Function.* Springer-Verlag

Hauser, H., Pascher, I., Pearson, R. H., and S. Sundell. 1981. Preferred conformation and molecular packing of phosphatidylethanolamine and phosphatidylcholine. *Biochimica et Biophysics Acta* 650:21-51

Heller, H., Schaefer, M., and K. Schulten. 1993. Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid-drystal phases. *J. Phys. Chem.* 97:8343-8360

Nagle, J. F. 1993. Area/lipid of bilayers from NMR. *Biophys. J.* 64:1476-1481

Venable, R. M., Zhang, Y., Hardy, B. J., and R. W. Pastor. 1993. Molecular dyanmics simulations of a lipid bilayer and of hexadecane: an investigation of membrane fluidity. *Science* 262:223-226

Wiener, M. C., and S. H. White. 1992. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data III. complete structure. *Biophys. J.* 61:434-447

MONTE CARLO STUDY ON POLYETHELYNE GLYCOL (PEG) CHAIN WITH ALAMETHICIN CHANNELS

Hongmei Jian

New York University

We used Monte Carlo simulations to investigate the conformational and thermodynamic properties of a Polyethelyne glycol (PEG) chain with alamethicin channels. Three parameters determine the properties of polymer in this model: Kuhn statistical length, torsional rigidity and channel geometry. We studied the effects of polymer length and channel size on conformational peroperties of polymer inside and outside the channel. It is shown from the simulations that outside the channel the properties of PEG chain are consistent with the statistical results. Inside the channel, the important consideration is the size of the channel. With larger channel size (larger than persistence length of the chain), the chain will fold and twist while it is translocating. Several physical properties have been studied. With small channel size, the simulation shows that it is also possible for the chain to translocate from one side of the membrane to the other side. Some details are still being investigated.

FORCE FIELDS AND FREE ENERGY CALCULATIONS FOR COMPLEX SYSTEMS

Peter Kollman

Dept. of Pharmaceutical Chemistry
UC, San Francisco

We will review our recent work using molecular dynamics techniques to study complex molecules. We will begin with the simplest system, a dioxane, where we have quantum mechanical data on the conformational equilibrium and critically assess force field approaches which can reproduce the high level ab initio and experimental data. Next we will show where non-additive effects can be important in the calculation of molecular interactions both in vacuo and in solution. Particularly striking examples are ion-pi complexes and solvation of complex organic ions. We will then present some free energy calculations on simple systems-amines, amides and phenols-where the results disagree with experiments and try to assess why. We will then turn to applications of molecular dynamics to more complex systems, choosing from molecular dynamics of bpti including water, liquid octanol, octaspherand-alkali cation, thermolysin and its inhibitors, biotin-avidin and distamycin-DNA. We will attempt to critically assess the successes, failures and difficulties in general applicability to non-covalent molecula interactions in complex systems.

A Fast Multipole Algorithm for Molecular Simulations of Very Large Dipolar and Charged Dipolar Systems

Ramzi Kutteh and John B. Nicholas

High Performance Computing Group, Environmental and Molecular Sciences Laboratory,

Pacific Northwest Laboratory, Richland, WA 99352*

Abstract

The fast multipole method (FMM) and cell multipole method (CMM) are efficient algorithms for computing charge-charge interactions in MD simulations. However, dipolar interactions are as important and challenging to compute. We present the dipole cell multipole method (DCMM) for computing the dipolar energy and forces in million-particle systems. DCMM by itself is applicable to systems of permanent dipoles and systems of permanent and induced dipoles. We apply DCMM to arbitrary dipolar systems and find it very fast and accurate. We also present the DCMM/CMM method for efficient computation of all interactions in general charged dipolar systems. These include systems of permanent charges and permanent dipoles, systems of permanent charges and induced dipoles, and polarizable systems of charges and dipoles. The DCMM/CMM approach can be generalized to include higher order point multipoles. Finally, we use DCMM to speed-up the time consuming self-consistent iteration of induced dipoles in polarizable systems. In particular, the DCMM iterative scheme proves very efficient when we apply it to large systems of polarizable simple point charge (PSPC) water.

* *Pacific Northwest Laboratory is a multiprogram national laboratory operated for the U.S. Department of Energy by Battelle Memorial Institute under Contract DE-AC06-76RLO 1830.*

Modeling the Three-Dimensional Folding of RNA Pseudoknots in the Ribosomal Frameshifting

Shu-Yun Le, Hugo M. Martinez,
Mario Chamorro[‡], Jih-H. Chen[†],
N. Pattabiraman[†], and Jacob V. Maizel, Jr.

Laboratory of Mathematical Biology
Division of Cancer Biology, Diagnosis and
Centers, National Cancer Institute, NIH
Frederick, MD 21702

[‡] National Cancer Institute, NIH
Bethesda, MD 20892

[†] Frederick Biomedical Supercomputing Center
Program Resources, Inc./ NCI/FCRDC
P.O. Box B, Frederick, MD 21702

Abstract

Highly efficient ribosomal frameshifting in the gag-pro region of mouse mammary tumor virus (MMTV) requires a higher-order RNA structure, pseudoknot, just downstream of the frameshift site. The importance of the MMTV pseudoknot has been demonstrated by extensive site-directed mutations, in which the distinct conformations of RNA pseudoknot are supposed to be changed in the stacking region of stem 1 and stem 2, and in a crucial bulge loop A that interrupt the two stems, and size of the loop 1. Using a computer modeling protocol, RNA2D3D, we built four MMTV pseudoknots used in the analyses of extensive site-directed mutations on the basis of their established RNA secondary structure and limited amounts of tertiary structural data. That is to say, the stem 1 is supposed to be coaxially stacked with stem 2, and loop 1 and loop 2 are across the major and minor grooves, respectively. These structural models of RNA pseudoknots are further refined by molecular mechanics and molecular dynamics (MD) simulations with sodium ions and waters. The distinct conformations of these possible RNA pseudoknots were analyzed and compared to each other. The important role of nucleotides at the junction of the stem 1 and stem 2 was discussed in the efficient ribosomal frameshifting. Our results indicate that the highly efficient frameshifting pseudoknots are so bent and kinked that stem 1 and stem 2 are not coaxially stacked, and the base pairing in the junction of these two stems is distorted. However, two stems in the poor frameshifting pseudoknots are almost coaxially stacked. These models emphasize the importance of base-base interaction and an unpaired A in the connection region of two stems of the MMTV pseudoknot.

A NEW DEFINITION OF ATOMIC CHARGES BASED ON A VARIATIONAL PRINCIPLE FOR THE ELECTROSTATIC POTENTIAL ENERGY

Tai-Sung Lee

Department of Chemistry
Duke University
Durham, NC 27710

A unique definition of atomic charges in molecules is presented based on a variational principle involving the electrostatic potential energy. The method requires only the electron density as input, and does not rely on an arbitrary set of fitting points as do conventional electrostatic potential fitting procedures. The dipole moments and electrostatic potentials calculated from atomic charges obtained from this method agree well with those from self-consistent-field calculations. The new method also provides a spherical-atom potential model that may be useful in future generation molecular simulation force fields.

Quantum - Classical Molecular Dynamics and Its Applications in Macromolecular Simulations

P.Bala[1], P.Grochowski[2], B.Lesyng[2] and J.A.McCammon[3]

1 Institute of Physics, Nicolas Copernicus University, 5 Grudziadzka, 87-100 Torun, Poland

2 Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw University, 2 Banacha, 02-097 Warsaw, and Department of Biophysics, Warsaw University, 93 Zwirki i Wigury, 02-089 Warsaw, Poland

3 Institute for Molecular Design, Department of Chemistry, University of Houston, Houston, TX 77204, USA

During the past few years we have been developing Quantum-Classical and Quantum-Stochastic Molecular Dynamics models (QCMD/QSMD), to describe time-dependent, quantum dynamical proton transfer processes in hydrogen-bonded systems and in enzymes [1-6]. In general, the models can be used to simulate tunneling processes in time-dependent potentials, e.g., hopping of quantum particles, protons or electrons, in the solid state and in macromolecular systems, see [7]. It should be noted that tunneling effects, previously neglected in enzyme simulations, are important in proton transfer processes [8].

In the QCMD model the proton/electron dynamics are described by the time-dependent Schroedinger equation. The dynamics of the classical atoms are described using classical molecular dynamics. Coupling between the quantum proton/electron and the classical atoms is accomplished *via* extended Hellmann-Feynman forces, as well as the time-dependence of the potential energy function in the Schroedinger equation. The extended theorem developed [2], is as follows:

$$F_\alpha = -\frac{\partial\langle\psi|H|\psi\rangle}{\partial X_\alpha} = -\sum_k \left[\langle\psi\,|\,\psi_k\rangle\left\langle\psi_k\left|\frac{\partial V}{\partial X_\alpha}\right|\psi_k\right\rangle\langle\psi_k\,|\,\psi\rangle + E_k\frac{\partial}{\partial X_\alpha}\left(\langle\psi\,|\,\psi_k\rangle\langle\psi_k\,|\,\psi\rangle\right)\right]$$

X_α are the positions of the nuclei. Ψ_k and E_k are the instantaneous eigenfunctions and eigenvalues of H. The QCMD model is numerically stable and conserves the total energy. The potential energy function is either parametrized prior to the simulations [1-4] or is computed using a parametrized valence bond (VB) orbital method [6].

The parametrizations are done based on *ab initio* calculations. In macromolecular systems, e.g. in enzymes, the potential for the proton motion is modulated with local dynamics (squeezing and stretching forces influencing the shape and the barrier height) and is modified by the electrostatic field generated by the remote classical atoms. In the QSMD model the interaction of the system with its environment is described by Langevin dynamics of the classical atoms/particles.

During the workshop the scheme of the QCMD/QSMD models will be formulated and simulation results will be presented for the quantum proton transfer reactions in proton-bound ammonia-ammonia and water-ammonia dimers as well as in phospholipase A_2. Other quantum-dynamical models, e.g. [8-12], will be compared. The QCMD/QSMD software package contains a variety of numerical tasks (long range two-body interactions, their gradients and second derivatives, matrix diagonalization and polynomial expansions, FFT techniques, etc.). Such problems require mixed parallel-vector type computer architectures. This is quite typical situation when trying to model structure and dynamics of real physical, chemical and biological atomic/molecular systems. This formulates new challenges for next generations of supercomputers and software tools.

References

[1] P.Bala, B.Lesyng and J.A.McCammon, Chem. Phys., **180**, 271-285(1994)

[2] P.Bala, B.Lesyng and J.A.McCammon, Chem. Phys. Letters, **219**, 259-266(1994)

[3] P.Bala, B.Lesyng, T.N.Truong and J.A.McCammon, in *Molecular Aspects of Biotechnology: Computational Models and Theories*, ed. J.Bertran, pp. 299-326, Kluwer Acad. Publishers, Dordrecht, 1992

[4] L.Jaroszewski,B.Lesyng and J.A.McCammon, J.Mol.Struct.(Theochem),**283**,57-65 (1993)

[5] T.N.Truong, J.J.Tanner, P.Bala, J.A.McCammon, D.J.Kouri, B.Lesyng and D.K.Hoffman, J.Chem.Phys., **96**, 2077-2084(1992)

[6] P.Bala, P.Grochowski, B.Lesyng and J.A.McCammon, submitted to Comp.& Chem.

[7] B.Lesyng and J.A.McCammon, *Molecular Modeling Methods. Basic Techniques and Challenging Problems*, Pharmac. Ther., 60, 149-167, 1993

[8] Y.Cha, Ch.J.Murray and J.P.Klinman, Science, **243**, 1325-1330(1989)

[9] R.Car and M.Parinello, Phys. Rev.Letters, **55**, 2471-2476(1985)

[10] U.C.Singh and P.A.Kollman, J.Comput.Chem., **7**, 718-729(1986)

[11] J.-K.Hwang, G.King, S.Creighton and A.Warshel, J.Am.Chem.Soc.,**110**, 5297-5311(1988)

[12] M.J.Field, P.A.Bash and M.Karplus, J.Comput.Chem., **11**, 700-733(1990)

Can the Amino Acid Sequence Distinguish the Correct Fold of a Protein at Low-Resolution?

Michael Levitt, David Hinds and Britt Park
Department of Structural Biology,
Stanford Medical School,
Stanford, CA 94305

September 1994

Simplified models have played an important role in attempts to model protein folding by computer simulation. By eliminating much of the detail, such models are computationally more tractable. They also model physical effects such as the interactions between side chain that are not ordered.

While such models are useful, a central question is whether there is sufficient detail in a particular model to make the native fold stand out from the vast number of alternate non-native folds. More specifically, one hopes that the native fold has a very low free-energy value.

Here we use a simplified representation of proteins that allows all possible conformations to be generated. We then search this conformational space in order to find those folds that have low energies with a simple pair-wise residues contact potential. Results on a wide variety of small proteins show that, even at this low resolution, there is significant selectivity. Using the correct sequence of the protein, selects for those folds that are more similar to the native conformation of the protein. Using a shuffled sequence does not show such selectivity.

Although this proves that the native sequence is able to distinguish the native fold at low resolution, there are always other non-native folds that also have low energies. While such degeneracy may be depend on the energy functions we use, we believe it to an intrinsic property: at low resolution there is simply not enough detail to make the native fold unique.

DISCRETE DYNAMIC POLYMER MODELLING USING CELLULAR AUTOMATA TECHNIQUES

Marc Roussel* and Carmay Lim**

University of Toronto
MSB 4388, 1 King's College Circle
Toronto, Ontario
M5S 1A8 Canada

We have developed a novel simulation strategy based on cellular automata methods which can be used to simulate a variety of physico-chemical processes, including those involved in polymerization. Our approach leads to dynamic, parallel models. This strategy can address several classes of questions in technologically or scientifically important systems for which only limited structural or dynamical information is available by current experimental techniques. We illustrate the use of our methods by creating a model of lignification in vivo. Our lignification model captures the essence of the underlying physical processes, as evidenced by the fact that it reproduces satisfactorily many experimentally determined properties of lignin. Due to the inherent efficiency of parallel cellular automata, our simulation strategy shows great promise, particularly for modeling species of very high molecular weight (over a million daltons).

* Department of Chemistry
** Departments of Molecular and Medical Genetics, Biochemistry and Chemistry

MOLECULAR DYNAMICS STUDY OF THE 13-*cis* FORM OF BACTERIORHODOPSIN AND ITS PHOTOCYCLE

Ilya Logunov

Beckman Institute
405 N Mathews Ave.
Urbana, IL 61801

The 13-*cis* photocycle of bacteriorhodopsin has been studied by means of molecular dynamics simulations. The structure of the 13-*cis* bacteriorhodopsin was obtained through molecular dynamics refinement and tested by altering substituents of retinal and comparing with available observations. The photoisomerization process was simulated. The resulting structures of the J, K, and L intermediates revealed that the protonated Schiff base points to the cytoplasmic side and, hence, cannot form an M intermediate. Our simulations suggest the possibility that leakage from the 13-*cis* cycle to the *trans* cycle occurs during the initial photoisomerization step.

PROTEIN STRUCTURE PREDICTION USING OPTIMIZED HAMILTONIANS

Zan Luthey-Schulten

B56 Noyes Lab, MC 712
600 S Mathews
Urbana, IL 61801

Optimal protein folding codes for residue-residue contact (local) and associative memory type (AM) Hamiltonians lead to algorithms that correctly recognize protein structures in the region of low sequence identity in the overwhelming majority of cases. The optimization is based on simple thermodynamic considerations using spin glass theory. Simulated annealing for the optimally-encoded AM Hamiltonian generally leads to qualitatively correct structures. The process of structure prediction and design using these two approaches is presented for several unknown proteins.

SIMULATION OF MANY BODY SYSTEMS USING THE CLASSICAL DENSITY DISTRIBUTION

Jianpeng Ma

Chemistry Department
Boston, MA02215

Classical Gaussian wave packets approach were applied to the development of varieties kinds of simulated annealing techniques. Constant temperature Gaussian Phase Packet(GPP) algorithm was served as an alternative of classical molecular dynamic simulation and Fokker-Planck(FP) dynamics was also an another constant temperature dynamic algorithm developed for one to perform the annealing simulation. Gaussian Density Annealing(GDA) method were also established for one to pursue an approximate solution for the equilibrium density distribution at zero temperature. All these three annealing methods were applied to the energy global optimization of Lennard-Jones clusters with size range from 8 to 19. Larger size clusters were also tried. A systematic comparison of all methods and the comparison between wave packets approach and conventional annealing algorithms were explored.

WHAT MIGHT EXPLAIN RESONANCE IN THE DYNAMICS OF CHEMICAL SYSTEMS SIMULATED BY THE IMPLICIT MIDPOINT SCHEME?

Margaret Mandziuk

Chemistry Dept. and
Courant Institute of Mathematical Sciences
251 Mercer Street
New York, New York 10012

We study the numerical behavior of the implicit midpoint method (MID) with relatively large timesteps of integration through applications to small nonlinear systems such as an HBr molecule governed by a Morse potential, a Henon-Heiles system, and deoxycytidine, and subsequent analysis.

MID is known to be stable and energy preserving for very large timesteps in the linear regime. However, behavior in the nonlinear regime is not well understood. In particular resonance problems have been noted and limited the application of MID to more complex systems such as biomolecules. For the systems we examine here, the energy fluctuates from step to step, with the range of fluctuations depending on the timestep used. However special behavior is observed at particular timesteps. Namely, resonances occur at certain values of the timestep and instability results. Interestingly, for other values of timestep (even significantly larger than these associated with resonances) the range of the fluctuations remains bound. The resonance at these timesteps is analyzed by careful observation and Fourier analysis.

Conformational analysis of molecular chains using Nano-Kinematics

Dinesh Manocha
Computer Science Department
University of North Carolina
Chapel Hill, NC 27599-3175
{manocha,zhu,wright}@cs.unc.edu
Fax: (919)962-1799

Abstract: We present algorithms for 3-D manipulation and conformational analysis of molecular chains, when bond length, bond angles and related dihedral angles remain fixed. These algorithms are useful for local deformations of linear molecules, exact ring closure in cyclic molecules and molecular embedding for short chains. Other possible applications include structure prediction, protein folding, conformation energy analysis and 3D molecular matching and docking. The algorithms are applicable to all serial molecular chains and make no asssumptions about their geometry. We make use of results on direct and inverse kinematics from robotics and mechanics literature and show the correspondence between kinematics and conformational analysis of molecules. In particular, we pose these problems algebraically and compute all the solutions making use of the structure of these equations and matrix computations. The algorithms have been implemented and perform well in practice. In particular, they take tens of milliseconds on current workstations for local deformations and chain closures on molecular chains consisting of six or fewer rotatable dihedral angles.

ALGORITHMS FOR SHARED MEMORY PARALLEL MOLECULAR DYNAMICS IN SIGMA

Lars Nyland

Dept. of Computer Science
University of North Carolina
Chapel Hill, NC 27599-3175

The goal of our work is to achieve the best possible performance while maintaining full functionality of the SIgMA MD simulator on currently available high-performance computers. Our current efforts target the KSR-1, a shared-memory computer with 48 50MIP processors at the NC Supercomputer Center. Multiple versions of SIgMA have been implemented on the KSR, using different decompositions to assign work to processors. In this talk, we will discuss the decomposition strategies, comparing them with one another and with execution of SIgMA on other supercomputers as well. We will describe the algorithms used, the primitives available for parallel execution on the KSR, analysis capabilities, and desired capabilities. We will also discuss the future plans of high-speed execution of SIgMA, based on what we've learned here.

LOW COMPLEXITY MODELS OF PROTEIN STRUCTURE

Britt Park

Dept. of Cell Biology
Stanford University
Stanford, CA 94305-5400

We have examined the relationship between the complexity and accuracy of various models of protein a-carbon backbone structure, by developing a simple algorithm for generating near optimal fits to X-Ray structure by arbitrary models. We found that even low complexity models can be accurate. By a simple optimization procedure we have generated several 4 state per amino acid residue models which on average fit X-Ray protein structures to within 2.4 Å and preserve 80% of native contacts. We examine the characteristics of these models and discuss their use in the prediction of protein conformation.

MODELING WATER IN AND AROUND PROTEINS

B. M. Pettitt

Department of Chemistry
University of Houston
Houston TX 77204-5641

Modeling the solvent around proteins is an important and notoriously difficult problem for ligand design/evaluation. Both vector and parallel plaforms have been used to obtain simulations of proteins in water; our goal here is to take a step toward reducing the computational workload by reformulating the problem in terms of characteristic solvation correlations derived from simulation. The solvent structure and dynamics around myoglobin has been investigated at the microscopic level of detail from a computer simulation. We analyzed a molecular dynamics trajectory in terms of solvent mobility and probability distributions. A strong correlation between the solvent mobility and density emerges on both global and local scales. We have proposed a simple model where the solvent distribution measured perpendicularly to the protein surface is utilized to reconstruct the simulated network of hydration within 6 \AA\ from the protein surface with a relative error of only 17%. The global precision of this solvation model matches results obtained with more complicated models usually used in refinement procedures in X-ray and neutron experiments but with far fewer parameters. The dramatically improved correspondence between observed and calculated X-ray intensities at low resolution relative to other methods both confirms the validity of the approach used in the MD simulations and allows the results of this study to be implemented in solvent studies on real systems.

IDENTIFYING THE MECHANISM OF PROTEIN LOOP CLOSURE

Marios Philippopoulos, Yue Fang Xiang and Carmay Lim

The mechanism underlying Loop opening in Bacillus Stearothermophillus Lactate DeHydrogenase (BSLDH) is explored through a molecular dynamics simulation at high temperature (1000 K) in the presence of explicit solvent, starting from the X-ray structure of BSLDH complexed with the co-enzyme NAD^+ and oxamate at 2.5 AA. During the simulation, a significant conformational change has occured, as evidenced by sharp dihedral angle transitions, hydrogen bond breaking and formation and large root-mean-square deviations from the starting structure; these changes define the criteria for Loop-opening. The mechanical elements responsible for Loop opening; i.e., Loop hinges and flap, are defined through a combination of order parameters, dihedral angle changes and their correlations, and the dynamical cross-correlation map of atomic displacements for the Loop residues. The results indicate that the Loop consists of two flexible hinge regions on either side of a relatively rigid three-residue segment that undergoes a significant spatial displacement during Loop opening. Loop opening is made possible through an array of correlated dihedral angle changes and intra-Loop re-arrangements of hydrogen-bond interactions.

MOLECULAR DYNAMICS OF RHINO VIRUS: EFFECTS OF AN ANTIVIRAL COMPOUND

C. B. Post and D. Phelps

Dept. of Med. Chemistry
Purdue University
West Lafayette, IN 47906-1333

Structural and biological studies on human rhino virus (HRV) complexed with antiviral agents indicate that one cause of the antiviral activity is the inhibition of disassembly of the viral capsid. We are using molecular dynamics methods to investigate how the alteration in conformational and dynamical properties might lead to changes in disassembly or disruption of quarternary structure. Molecular dynamics simulations are being carried out on free HRV14 and drug-bound HRV14 using a stochastic boundary method (sbmd) that involves a truncated, spherical region centered on the drug-binding pocket. It was found that binding of a drug in the internal cavity of the virus affected the magnitude of the fluctuations in free-volume, a quantity related to thermodynamic compressibility. The computations were repeated with different boundary conditions for the sbmd method, so that the free-volume fluctuations were calculated using a grid method for combined time periods of more than 2.4 ns. Thus, the compressibility was found to increase when drug is bound in the viral pocket, a result which appears counterintuitive but which is consistent with the measured increase in thermal stability and inhibition of disassembly. Plans for future work to understand this effect on compressibility include simulations with different drugs and mutant forms of the virus. The viral dynamics also suggest a path for entry into the fully buried binding pocket. Future work is planned to study drug entry and to simulate larger systems that include important protein-protein interactions. Large-scale simulations are underway and should provide a more accurate description of the dynamics and characterization of concerted, long-range motions.

APPLICATION OF THE RAY-RPRESENTATION AND A MASSIVELY PARALLEL SPECIAL PUPOSE COMPUTER TO PROBLEMS OF PROTEIN STRUCTURE AND FUNCTION

Michael G. Prisant*

Department of Chemistry
Duke University
Durham, NC 27706

Methodology is developed to apply ray-representations to geometric analysis of space-filling models of protein structure. The following specific problems are treated. First, we define the ray-representation for fused-sphere models of proteins. Second, using the ray-representation, we treat computation of molecular contact surfaces in solution via Minkowski dilation and erosion. Third, we describe how all points of the molecular contact surface can be tagged according to their chemical properties. Fourth, we show how equivalence set methods can be applied to ray representations of proteins in order to identify internal empty spaces and classify their connectedness to the outside. Fifth, we develop filters to analyze the morphology of interstitial spaces in proteins which connect to the outside. Sixth, we discuss how Boolean algorithms can be used to determine whether water molecules identified in X-ray crystallography are inside, outside, or intersecting the boundary defined by the solvent accessible surface. Finally we discuss the computation of volume properties from ray-representations. A special purpose massively parallel computer is used to compute the ray-representation.

In an application of ray-casting technology, we consider the dynamical evolution of voids in myoglobin. The analysis begins with the recent molecular dynamics (MD) calculations on myoglobin in water performed by Bernard Brooks and coworkers at NIH. A space-filling model is created based on the atomic coordinates of the protein determined as a function of time in the MD calculations. This model defines the solvent accessible boundary of the protein. Further geometric analysis determines both the captured voids and reneentrant cavities in the protein based on this model. Our preliminary results show that on a 100 picosecond time scale that there is a bubbling motion of voids to the surface from the oxygen binding cavity along three wel defined paths. This motion is due exclusively to thermal fluctiations of the protein in the water bath.

* Secondary Appointment, Department of Computer Science, Duke University, Durham, NC

SOLVENT EFFECTS ON THE DYNAMICS OF SUPERCOILED DNA

Gomathi Ramachandran

Courant Institute of Mathematical Sciences
New York University
251, Mercer Street
New York, NY 10012

The effects of solvent on the dynamics of supercoiled DNA are examined through a simple, macroscopic energy model in the context of Langevin dynamics simulations. Closed circular DNA systems are modeled by B-splines and both elastic and electrostatic (screened Coulomb) potentials are used in the energy function.

In the Langevin formalism, the collision frequency, γ, determines the magnitude of the friction and the variance of the random forces due to molecular collisions and thus characterizes approximately the influence of the solvent on the motion of the solute. Thus, as a first approximation, the Langevin equations of motion can be parametrized to model the dynamics of DNA in solution. Solvent damping is well known to alter the dynamics behavior of DNA, affect various hydrodynamic measurements, and introduce significant entropic effects. By varying γ over 10 orders of magnitude, we identify three distinct physical regimes:

(i) low γ, dominated by globally harmonic motion;

(ii) intermediate γ, characterized by maximal sampling and high mobility of the DNA; and

(iii) high γ, dominated by random forces, where all of the global modes are effectively frozen by extreme overdamping.

The different regimes are explored extensively by Langevin dynamics simulations and very different behavior is observed, offering insights into hydrodynamic effects on supercoiled DNA.

WET PROTEINS AND VIBRATIONAL STATES: BEYOND NORMAL MODES

Adrian Roitberg
(Northwestern University, Evanston, IL)

Ron Elber
(University of Illinois at Chicago, Chicago, IL and
Hebrew University at Jerusalem, Jerusalem, Israel)

Benny Gerber
(University of California at Irvine, Irvine, CA and
Hebrew University at Jerusalem, Jerusalem, Israel)

and
Mark Ratner
(Northwestern University, Evanston, IL)

In the present poster we will study the effect of a water shell on the vibrational density of states of a medium-size protein and the influence of hydration on the protein's RMS fluctuations.

A possible improvement on the normal modes picture, including anharmonicity is presented and a calculation of vibrational ground state wavefunctions and energies for a large system shown.

The effects of the diagonal anharmonicity is shown to be large for a substantial number of modes, while the effect of off-diagonal anharmonicity (mode-mode coupling) is esentially zero.

Structural Analysis of Nucleic Acids. Applicability of Commonly used Parametrizations of Electrostatic Interactions.

W.R.Rudnicki a),b) and B.Lesyng a)
a) Warsaw University, Department of Biophysics,
93 Zwirki i Wigury, 02-089 Warsaw, Poland
b) Agricultural University, Department of Biophysics
26/30 Rakowiecka, 02-528 Warsaw, Poland

The AMBER, CHARMM, BIOSYM (cvff, cff91) and GROMOS force fields were tested by optimization of several DNA double-helical oligonucleotides in their A and B forms, as an extension of our former studies on DNA/RNA structure and dynamics, see e.g. [1-3]. We have applied various sets of electrostatic parameters. The optimized structures were characterized by their helicoidal parameters and compared with those obtained from crystallographic data. Atomic charges are essential for proper description of the DNA structure. In extreme case, the cvff and cff91 force fields with their original charges, fail to preserve the double-helical structure of DNA, but after updating these charges with the AMBER ones gave very good results, better than those obtained with the original AMBER force field. The distance dependent "dielectric constant" is superior to a fixed dielectric constant. Modification of the charges at the phosphate groups exhibit only minimal effects on the structures, but influences noticeably the dynamics of the oligonucleotides and the helicoidal parameters obtained by averaging the parameters over the trajectories. It is worth noting that the barrier to pseudorotation of furanose rings inside the double-helical DNA structures is sensitive to the Van der Waals interaction parameters.

[1] B.Lesyng, "The Barrier to Pseudorotation of Furanose Rings and its Biological Implications", in Topics in Nucleic Acid Structure , Part.3, ed. S.Neidle, Macmillan, London, 1977.
[2] W.R.Rudnicki, B.Lesyng and S.C.Harvey, Biopolimers, 34, 383 (1994)
[3] V.Pechenaya, W.R.Rudnicki, T.Grycuk and B.Lesyng, The Barrier to Pseudorotation of Furanose Rings in Double-Helical DNA Forms - in preparation.

A NUMERICAL TECHNIQUE FOR EVALUATING MOLECULAR PARAMETERS USING ULTRICENTRIFUGE SEDIMENTATION VELOCITY DATA

Hashim Saber

Montana State University - Billings
Department of Mathematics
1500 North 30th St.
Billings, MT 59101

A finite element scheme is adopted to simulate the ultracentrifuge sedimentation velocity data. To estimate the diffusion and sedimentation parameters, we use a non-linear least square algorithm that minimizes the square difference between experimental and simulated data. We also use the conveviance feature of the finite element algorithm to treat the non ideality case where the required parameters depend on the concentraion. In addtion, the multicomponent system where the parameters depend on the concentration is also simulated.

September 22, 1994

WRITEUP FOR KANSAS WORKSHOP, ALGORITHMS FOR MACROMOLECULAR MODELING
(EXCERPTED FROM OTHER WRITTEN MATERIAL)

Numerical Integration Schemes for Molecular Dynamics

Tamar Schlick

The Howard Hughes Medical Institute and New York University
Chemistry Department and Courant Institute of Mathematical Sciences
251 Mercer Street
New York, New York 10012, U.S.A.
(e-mail: schlick@nyu.edu)

Many problems in chemistry can be reduced to the solution of systems of coupled ordinary differential equations (ODE's). Examples include molecular and Langevin dynamics, rate equations of kinetic theory, and the time-dependent Schrödinger equation when expanded in a basis set. The technology of numerical integrators for solving ODE's has a long history with significant interplay between mathematics, physics and chemistry. Many of the earliest integrators, such as Runge-Kutta and predictor-corrector integrators, are still in common use, but there have also been recent advances, driven in part by the need for methods that can treat multiple timescales and have greater stability for the large-scale coupled nonlinear oscillators commonly found in molecular dynamics (MD) of polymers and biological macromolecules [1]. The long-time stability of integrators for such systems is a challenging area of mathematical-analysis research that is still in relative early stages [2], and perhaps the context of biomolecular MD will stimulate important developments.

Symplectic integrators have recently gained favorable attention in the mathematical community and quickly adapted for use in dynamics calculations in chemistry because of their favorable properties. In applications to Hamiltonian systems, symplectic integrators have the property of building in Liouville's theorem, whereby areas in phase space are preserved as the system evolves in time. This strong conservation property translates into stability over long-time integrations, an important property in MD calculations involving millions and more steps. One consequence of this for constant-energy MD simulations is that except for fluctuations, symplectic integrators at small timesteps conserve energy for very long times, whereas non-symplectic integrators typically introduce a systematic drift in the total energy. Time reversibility is another useful practical property of symplectic integrators.

Symplectic integrators may be *implicit* or *explicit*. In explicit methods, the solution at the end of the timestep is obtained by performing operations on the variables at the beginning of each timestep. Symbolically, we write $y^{n+1} = f(y^n, \Delta t, ...)$, where f is some nonlinear function, Δt is the timestep, y^n is the approximation to the solution y at time $n\Delta t$, and the dots indicate other parameters or previous solutions (i.e., y^{n-1}, y^{n-2}). With implicit integrators, the final solutions are functions of both the initial and final variables ($y^{n+1} = f(y^{n+1}, y^n, \Delta t, ...)$, so coupled nonlinear equations must generally be solved at each timestep to propagate the

trajectory. The explicit versions generally involve simple algorithms that (for propagation only) use modest memory, while implicit methods involve more complex algorithms but are often more powerful for treating systems with disparate timescale dynamics.

The development of symplectic integrators has involved significant interplay between mathematicians, physicists, and chemists. Seminal work on symplectic integrators was done by both physicists and mathematicians [3] based on second and third-order explicit approaches and Runge-Kutta methods. Implicit approaches were developed in parallel [4]. Recently these ideas have found their way into the chemistry community with promising results. The Verlet integrator [5], already in common use, was found to be symplectic, thereby explaining the good associated stability observed in practice. However, the Verlet and related methods — while simple to formulate and fast to propagate — impose a severe constraint on the maximum timestep possible

Standard techniques of effectively freezing the fast vibrational modes by a constrained formulation [6]

There are well known numerical techniques for solving differential equations describing physical processes with multiple timescales [9]. Various implicit formulations are available that balance stability, accuracy, and complexity. However, the standard implicit techniques used by numerical analysts [10] have not been directly applicable to MD simulations of macromolecules, for the following reasons.

First, such implicit schemes are often designed to suppress the rapidly-decaying component of the motion. This is a valid approach when the contribution of these components becomes negligible for sufficiently long times, as is the case for the second term in $y(t) = \exp(-t) + \exp(-100t)$. This situation does not hold for biomolecular systems because of the intricate vibrational coupling. It is well recognized that concerted conformational transitions (e.g., in hinge-bending proteins) require a *cooperative* mechanism driven by small-scale fluctuations to make possible a large-scale collective displacement. Thus, while the damping of the high-frequency modes may not by itself be a severe problem, the lower *energies* associated with these modes — see below — may be undesirable for proteins and nucleic acids, as cooperative motions among the correlated vibrational modes may require energy transfer.

Second, implicit schemes with known high-stability (e.g., implicit-Euler) can introduce numerical damping [11]. Therefore, such discretizations of the classical Newton equations of motion will eventually damp out the energy of a system. This has prompted the application of such implicit schemes to the *Langevin* dynamics formulation, which involves frictional and Gaussian random forces in addition to the systematic force to mimic molecular collisions and therefore a thermal reservoir. This stabilizes implicit discretizations and can be used to quench higher-frequency vibrational modes [12], but unphysical increased rigidity can result [11]. Therefore, more rigorous approaches are required to resolve the subdynamics correctly, such as by combining normal-mode techniques with implicit integration [11]; significant linear-algebra work in the spectral decomposition is necessary for feasibility for macromolecular systems. There has also been some work on implicit schemes that do not have inherent damping, but preliminary experience suggests that for nonlinear systems desirable energy conservation properties can only be obtained up to moderate timesteps [13].

Third, implicit schemes increase complexity, since they involve solution of a nonlinear system or minimization of a nonlinear function at each timestep. Therefore, very efficient implementations of these additional computations are necessary and, even then, computational gain (with respect to standard "brute-force" integrations at small timesteps) can be realized only at very large timesteps.

At this point it appears that the optimal algorithms for MD will require a combination of methods and strategies discussed above, including symplectic and implicit numerical integration schemes that have minimal intrinsic damping, and correct resolution of the subdynamics of the system by some other technique (e.g., normal-mode analysis). Undoubtedly, high-performance implementations will make possible a gain of several orders of magnitude in the simulation times, and there are certainly additional gains to be achieved by clever programming strategies.

References

[1] J.A. McCammon and S.C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge (1987).

M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, New York (1990).

[2] A.M. Stuart and A.R. Humphries, "Model Problems in Numerical Stability Theory for Initial Value Problems", *SIAM Review* **36**, 226 (1994).

[3] R.D. Ruth, "A Canonical Integration Technique", *IEEE Trans. Nucl. Sci.* **NS-30**, 2669 (1983).

K. Feng, "Difference Schemes for Hamiltonian Formalism and Symplectic Geometry", *J. Comput. Math.* **4**, 279 (1986).

J. Candy and W. Rozmus, "A Symplectic Integration Algorithm for Separable Hamiltonian Functions", *J. Comp. Phys.* **92**, 230 (1991).

R.I. McLachlan and P. Atela, "The Accuracy of Symplectic Integrators", *Nonlinearity* **5**, 541 (1992).

M.P. Calvo and J.M. Sanz-Serna, "The Development of Variable-step Symplectic Integrators, with Application to the Two-Body Problem", *SIAM J. Sci. Comput.* **14**, 936 (1993).

D.I. Okunbor and R.D. Skeel, "Canonical Numerical Methods for Molecular Dynamics Simulations", *Math. Com.* **59**, 439 (1992).

[4] P.J. Channell and J.C. Scovel, "Symplectic Integration of Hamiltonian Systems", *Nonlinearity* **3**, 231 (1990).

J. De Frutos and J.M. Sanz-Serna, *An Easily Implementable Fourth-Order Method for the Time Integration of Wave Problems*,.*J. Comput. Phys.* **103**, 160 (1992).

[5] L. Verlet, "Computer 'Experiments' on Classical Fluids. i. Thermodynamical Properties of Lennard-Jones Molecules", *Phys. Rev.* **159**, 98 (1967).

[6] W.F. van Gunsteren, "Constrained Dynamics of Flexible Molecules", *Mol. Phys.* **40**, 1015 (1980).

W.F. van Gunsteren and H.J.C. Berendsen, "Algorithms for Macromolecular Dynamics and Constraint Dynamics", *Mol. Phys.* **34**, 1311 (1977).

J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, "Numerical Integration of the cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes", *J. Comp. Phys.* **23**, 327 (1977).

S. Miyamoto and P.A. Kollman, "SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models", *J. Comp. Chem.* **13**, 952 (1992).

[7] W.B. Streett, D.J. Tildesley and G. Saville, "Multiple Time Step Methods in Molecular Dynamics", *Mol. Phys.* **35**, 639 (1978).

M.E. Tuckerman and B.J. Berne, "Molecular Dynamics in Systems with Multiple Time Scales: Systems with Stiff and Soft Degrees of Freedom and with Short and Long Range Forces", *J. Comp. Chem.* **95**, 8362 (1992).

H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten, "Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-Range Interactions", *Mol. Sim.* **6**, 121 (1991).

M. Watanabe and M. Karplus, "Dynamics of Molecules With Internal Degrees of Freedom by Multiple Time-Step Methods", *J. Chem. Phys.* **99**, 8063 (1993).

[8] J.J. Biesiadecki and R.D. Skeel, "Dangers of Multiple-time-step Methods", *J. Comp. Phys.* **109**, 318 (1993).

[9] C.W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Englewood Cliffs, New Jersey (1971).

G. Dahlquist and Å. Björck, *Numerical Methods*, Prentice Hall, New Jersey (1974).

[10] H.-O. Kreiss, "Problems with Different Time Scales", *Acta Numerica* **1**, 101 (1991).

(1991).

[11] G. Zhang and T. Schlick, "LIN: A New Algorithm to Simulate the Dynamics of Biomolecules by Combining Implicit-Integration and Normal Mode Techniques", *J. Comp. Chem.* **14**, 1212 (1993).

[12] C.S. Peskin and T. Schlick, "Molecular Dynamics by the Backward-Euler Method", *Comm. Pure App. Math.* **42**, 1001 (1989).

T. Schlick and C.S. Peskin, "Can Classical Equations Simulate Quantum-mechanical Behavior? A Molecular Dynamics Investigation of a Diatomic Molecule with a Morse Potential", *Comm. Pure App. Math* **42**, 1141 (1989).

[13] J.C. Simo, N. Tarnow, and K.K Wong, "Exact Energy-Momentum Conserving Algorithms and Symplectic Schemes for Nonlinear Dynamics", *Computer Methods in Applied Mechanics and Engineering* **100**, 63 (1992).

G. Zhang and T. Schlick, "Implicit Discretization Schemes for Langevin Dynamics", submitted (1994).

PARALLEL ALGORITHMS FOR BIOMOLECULAR MODELING

Babak Bagheri, Andrew Ilin, J. Andrew McCammon, L. Ridgway Scott, and Dexuan Xie
The Texas Center for Advanced Molecular Computation
University of Houston

Abstract

We describe the development of some parallel iterative techniques for solving boundary value problems for elliptic partial differential equations. Using domain decomposition techniques, we modify standard sequential iterative techniques to obtain effective parallel methods. We contrast implementations on distributed-memory and shared-memory scalable parallel processors. We describe the use of two different programming paradigms, one involving explicit parallelism in a distributed-memory model and the other utilizing simple loop decompositions in a shared-memory model. Our primary conclusion is that parallel computing on existing commercial parallel supercomputers makes it routine to do three-dimensional modeling of electrostatics around biomedically interesting systems.

I. Introduction

We discuss several techniques for solving elliptic boundary value problems via iterative methods which have a high degree of parallelism. These techniques are being developed to solve as broad a class of problems as possible, but our primary motivation has come from computing the electrostatic potential around molecules of biological significance [7]. Moreover, implementation of the methods has been done as part of an existing code UHBD [5]. This makes the code development more complex but also provides an assessment more realistic than would be available by looking only at computational kernels. In addition, we have applied some of the computational techniques to solve prototypical problems related to semiconductor device simulation [3].

We have studied several variants of standard iterative methods which we have designed to have good parallelism. These include variants of the well known ICCG and SOR iterative methods. In addition, we have proposed new types of iterations especially suitable for parallel computation [11]. We anticipate that all of these methods will be useful as coarse grid solvers for parallel multigrid methods [8].

In addition to studying different parallel iterative methods, we have used different parallel programming paradigms. Two of these are (1) IPfortran [1] and (2) shared memory constructs supported by Kendall Square's KSR-1 Fortran [10]. Both approaches have proved adequate for implementing the parallel algorithms presented here, due to the high degree of regularity of the loops involved. Less regular loops in UHBD, related to its Brownian dynamics phase, have been easier to parallelize using shared-memory constructs [4].

II. PSOR

The Jacobi method for approximating the solution of a linear system is naturally parallel, but the typically more efficient Gauss-Seidel method is essentially sequential. In the Jacobi method, each component X_i of the approximate solution vector $X = (X_1, \ldots, X_N)$ can be computed separately of all others, which we can write schematically as

$$X_i^{k+1} = F_i(X_1^k, \ldots, X_N^k), \quad \text{for} \quad i = 1, 2, \ldots, N, \tag{2.1}$$

where the F_i are functions of N variables. For example, $F = (F_1, \ldots, F_N)$ is an affine function in the case of solving a linear system. Typically F is sparse, depending only on entries near the diagonal, which we indicate by $F_i(\ldots, X_{i-1}, X_i, X_{i+1}, \ldots)$. With Gauss-Seidel, it is frequently the case that X_{i+1}^k depends on X_i^k: schematically it is

$$X_i^{k+1} = F_i(\ldots, X_{i-1}^{k+1}, X_i^k, X_{i+1}^k, \ldots) \quad \text{for} \quad i = 1, 2, \ldots, N. \tag{2.2}$$

The same applies for the SOR method, which is just a relaxed (or accelerated) version of Gauss-Seidel.

One approach taken to deal with the sequential nature of SOR is to reorder the unknowns so that one group of components X_i can be computed independently of others. This is often referred to as a coloring of the index set. The most well known case is that of two colors, usually called "red-black" ordering since it is similar to a chess board in simple cases. While this can be quite effective, it requires communication to be done for each color as opposed to just once for each iteration, as is the case for the Conjugate Gradient (CG) method. The number of colors required depends on the extent of the sparcity of F.

A simple technique used in practice is to decompose the index domain (the set of indices i) in a way to minimize the communication (either the number of messages required, or the size) among neighboring domains. Gauss-Seidel (or SOR) is used within each domain, without updating using the appropriate neighboring values. In the two-processor case, it takes the form

$$\begin{aligned} X_i^{k+1} &= F_i(\ldots, X_{i-1}^{k+1}, X_i^k, X_{i+1}^k, \ldots) \quad \forall\, i,\ 1 \le i \le N/2, \\ X_i^{k+1} &= F_i(\ldots, X_{N/2}^k, X_{N/2+1}^{k+1}, \ldots, X_{i-1}^{k+1}, X_i^k, X_{i+1}^k, \ldots) \quad \forall\, i,\ \frac{N}{2} + 1 \le i \le N. \end{aligned} \tag{2.3}$$

Once the local Gauss-Seidel (or SOR) sweep is done, neighboring values are exchanged, similarly to what would be done in the Jacobi iteration. For this reason, we refer to this method as the Jacobi-Gauss-Seidel (JGS) algorithm (or JSOR for its accelerated or relaxed variant). While appealing for its simplicity, this algorithm frequently requires a much larger number of iterations than the sequential case.

Remarkably, a simple alternative [12] to JGS and JSOR has convergence properties similar to the sequential case, but with communication features similar to JGS/JSOR. We will not attempt a complete description of the most general case, but will simply describe an example and present

numerical results. Consider the following algorithm:

$$X_i^{k+1} = F_i(\ldots, X_{i-1}^{k+1}, X_i^k, \ldots, X_{N/2}^k, X_{N/2+1}^{k+1}, \ldots) \quad \forall\, i,\ 1 \le i \le N/2,$$
$$X_i^{k+1} = F_i(\ldots, X_{N/2}^k, X_{N/2+1}^{k+1}, \ldots, X_{i-1}^{k+1}, X_i^k, X_{i+1}^k, \ldots) \quad \forall\, i,\ \frac{N}{2}+1 \le i \le N. \tag{2.4}$$

This algorithm, which we call PGS (and PSOR for its accelerated or relaxed variant) is parallel for sparse F to the extent that the values $X_{N/2}^{k+1}, \ldots, X_N^{k+1}$ which are produced by the processor computing the second line can be computed and made available to the processor computing the first line before they are needed. In the case that the functions F_i are suitably sparse, this constraint poses no practical limitation to parallelism.

Figure 2.1 shows performance analysis for calculations done with the 5-point discretization of Laplace's equation using a strip decomposition (algorithm (2.4) in the case of two processors). We use this type of performance analysis graph to isolate different parts of a code. The computation time decreases even superlinearly [4] whereas the communication time (due to the use of a strip decomposition) remains nearly constant. The category "other time" simply reflects the part of the total time that cannot be accounted for in either of these categories; in this case it is quite small (being less that a second for two and four processors).

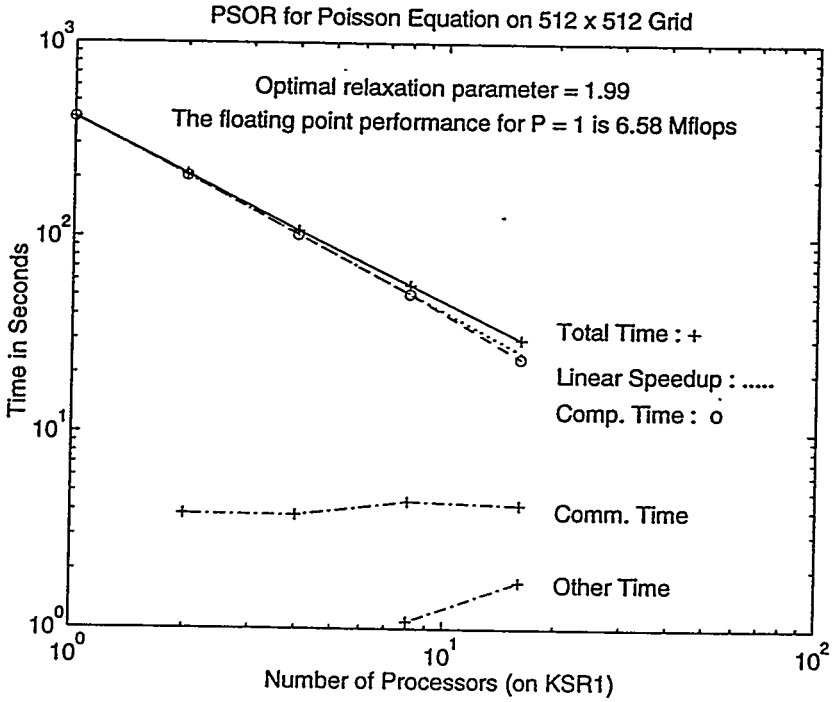

Figure 2.1. Performance analysis for PSOR for the 5-point discretization of Laplace's equation using a strip decomposition on the KSR-1.

We note that the code for this test was implemented in IPfortran and compiled separately for the Delta and KSR-1, without change of source code. The resulting speedup is almost identical for

both systems. In fact, the computation and communications times are largely the same for both systems. Although it is certainly possible to optimize performance for these distinct architectures, this shows that a single programming paradigm can provide efficient execution across a variety of different parallel architectures.

III. PICCG

Our parallel variants of ICCG have been implemented as part of the code UHBD [5,6] which was developed to study the interaction of two molecules of biological significance. One phase involves computing the electrostatic potential around the dominant molecule, and the second phase simulates Brownian motion of the second molecule in this electrostatic force field. The first phase solves the nonlinear Poisson-Boltzmann (NLPB) equation for the electrostatic potential.

We have modified the electrostatic solver to be able to model semiconductor devices [3]. This has provided a stronger test both of the linear and nonlinear parts of the solver, but the principal conclusion is that semiconductor devices can be modeled quite effectively on massively parallel computers. For example, the following table shows that the solver is scalable in the sense that larger problems can be solved without increasing the execution time, by increasing the number of processors used.

Total CPU time in seconds for a MOSFET simulation
on the Intel Delta for P nodes and mesh of size N^3

N^3	$P=1$	2	4	8	16	32	64	128	256
30^3	*22*	14	9	6	4	4	4	5	8
60^3	**192**	99	52	35	*20*	15	12	13	16
90^3			**214**	94	62	36	28	*26*	28
140^3					**184**	127	90	91	76
200^3							**252**	228	197
260^3									**440**

One particular case of interest is the so-called *memory constrained* scaling, the times for which are indicated in bold face. This is the case using the smallest number of processors which can run the problem, i.e., can fit the problem in local memory. We note nearly constant run times for this case. The slanted numbers indicate a different scaling which corresponds to a number of processors yielding an execution time that is an order of magnitude smaller. In this case, local memory is not utilized fully.

Most importantly, this table indicates that very large problems can be solved in just a few minutes (or just a few seconds, depending on resources available), allowing repeated designs to be tested or even optimized. We note also that the best decomposition has not been used for the case of large P and moderate N. If a block decomposition were used in this case, even better performance would be realized for the times away from the diagonal in the table.

One striking conclusion of our work so far [2, 3, 4] is that the total execution time for the elliptic solver portion of UHBD is essentially the same for quite disparate computer architectures and programming paradigms, as shown in Figure 3.1. The computations on each machine have quite distinct internal characteristics. For example, each calculation in done in each machine's single precision, which is 8-bytes on the KSR-1 and 4-bytes on the Delta. Due to the shorter word length, more iterations actually are done to reach the prescribed tolerance (the same for both machines).

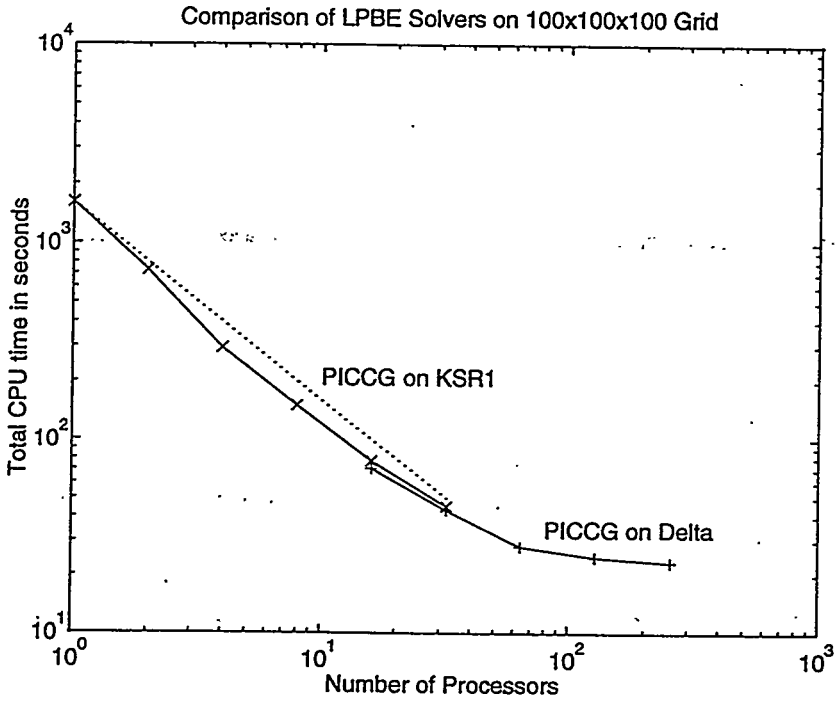

Figure 3.1. Timing for the linear and nonlinear solvers in UHBD on a test problem with a single atom.

In addition, quite different programming paradigms are being used in each case. For the Delta computations, we used IPfortran [1], an explicitly parallel language. For the KSR computations, we used the KSR "tiling" directives [10]. However, the total time is almost identical for 16 and 32 processors for a uniform mesh of size 100^3.

IV. Conclusions and future work

We have identified a number of promising parallel iterative methods, but we have not yet begun to quantify their domains of applicability (and superiority). Moreover, we anticipate these will ultimately find their best application as coarse grid solvers in a parallel multigrid technique. On the other hand, just using these parallel variants of standard iterative methods, we are able to solve two and three dimensional problems of substantial interest remarkably quickly.

References

[1] Bagheri, B., Clark, T. W., and Scott, L. R. IPfortran: a parallel dialect of fortran. *Fortran Forum 11* (Sept. 1992), 20–31.

[2] Bagheri, B., Ilin, A., and Scott, L. R. Parallelizing UHBD. Research Report UH/MD 167, Dept. Math., Univ. Houston, 1993.

[3] Bagheri, B., Ilin, A., and Scott, L. R. Parallel 3-D MOSFET simulation. In *Proceedings of the 27th Annual Hawaii International Conference on System Sciences* vol. 1, T.N.Mudge and B.D. Shriver, ed's, IEEE Computer Soc. Press, 1994, pp. 46–54.

[4] Bagheri, B., Ilin, A., and Scott, L. R. A Comparison of Distributed and Shared Memory Scalable Architectures. 1. KSR Shared Memory, Scalable High Performance Computing Conference, May, 1994.

[5] Davis, M. E., Luty, J. D., Allision, B. A., and McCammon, J. A. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian Dynamics program. *Computer Physics Communications 62* (1990), 187–197.

[6] Davis, M. E., and McCammon, J. A. Solving the finite difference linearized Poisson-Boltzmann equation: A comparison of relaxation and conjugate gradient methods. *Journal of Computational Chemistry 10* (1989), 386–391.

[7] Gilson, M. K., Straatsma, T. P., and McCammon, J. A. Open 'back door' in a molecular dynamics simulation of acetylcholinesterase. *Science March 4* (1994), 386–391.

[8] Holst, M., and Saied, F. Multigrid solution of the Poisson-Boltzmann equation. *Journal of Computational Chemistry 14* (1993), 105–113.

[9] A. Ilin, B. Bagheri, L.R.Scott, J.M.Briggs and J.A.McCammon, Parallelization of Poisson-Boltzmann and Brownian Dynamics Calculation, in Parallel Computing in Computational Chemistry Ed. T. G. Mattson ACS Books, Washington D.C., to appear in 1995.

[10] Kendall Square Research Corporation. *KSR Fortran Programming.* Kendall Square Research, Waltham, MA, 1992.

[11] Scott, L. R. Elliptic preconditioners using fast summation techniques, Domain Decomposition Methods in Science and Engineering (Proceedings of the Seventh International Conference on Domain Decomposition, October 27-30, 1993, The Pennsylvania State University), D. E. Keyes & J. Xu, eds. To appear in Contemporary Mathematics, Americam Math. Soc., Providence, 1995.

[12] Xie, D. Research Report UH/MD, Dept. Math., Univ. Houston, to appear.

The Acceleration of Time in DTH Dynamics

Yosi Shibberu
Mathematics Department
Rose-Hulman Institute of Technology
Terre Haute, IN 47803

email: Yosi.Shibberu@Rose-Hulman.Edu

September 12, 1994

Abstract

A discrete variational principle is used as the basis for a discrete-time theory of Hamiltonian dynamical systems called DTH dynamics. DTH dynamics is governed by implicit difference equations which completely determine piecewise-linear, continuous trajectories. These trajectories exactly conserve the Hamiltonian function at the midpoints of each linear segment and exactly conserve, at the vertices, all conserved quadratic functions. The DTH equations are formally equivariant with respect to a collection of symplectic, piecewise-linear, continuous coordinate transformations. A generating function also exists which determines transformations between the vertices of the piecewise-linear, continuous trajectories. DTH dynamics determines a unique parametrization of time. An asymptotic expansion of this parametrization of time is used as the basis of a new explicit scheme for integrating Hamiltonian dynamics.

COMPLEXITY OF MULTIPLE TIME STEPPING ALGORITHMS

K. Srinivas

Beckman Institute
University of Illinois
Urbana, IL 61801

Though the idea of multiple time stepping dates back to the seventies, there has not been a rigorous analysis of the error due to multiple time stepping and the complexity of N-body integration using multiple time stepping. This poster examines the complexity of a symplectic MTS scheme under various settings.

A practical symplectic distance class algorithm

Robert D. Skeel[1]
Department of Computer Science and Beckman Institute
University of Illinois at Urbana-Champaign
1304 West Springfield Avenue
Urbana, Illinois 61801-2987, U.S.A.
September 1994

Abstract

The high cost of doing macromolecular dynamics simulation is primarily due to the calculation of large numbers of electrostatic interactions. The use of distance classes with multiple time steps (MTS) has been found to reduce dramatically the computing time. The idea is that different terms of the collective force vector are integrated with different time steps. With distance classes the various pairwise interaction terms are partitioned into classes based on the distance between atom pairs. Interactions at greater distances are much more numerous but they are evaluated less often. Further savings are possible by using a method such as the fast multipole method to evaluate forces in the "outermost" distance class, an idea due to A. Windemuth. Although the concept of multiple time steps goes back 20 years, only recently was discovered, independently by groups at Illinois and Columbia, an MTS generalization of the Verlet method, which is both time-reversible and symplectic. Symplecticness is a property of the original analytical differential equations, which is inherited only by special numerical integrators. The significance of being symplectic may be this: a numerical integrator is symplectic if and only if the discretization errors can be interpreted as errors in the energy function (not merely as errors in the collective force vectors). However, the typical use of MTS in conjunction with distance classes destroys both the time reversible and symplectic property of the method. What we propose instead is an artificial splitting of each potential energy term

$$U(r) = U^{\text{hard}}(r) + U^{\text{soft}}(r)$$

such that $U^{\text{hard}}(r)$ vanishes for $r \geq r_{\text{cut}}$ and $U^{\text{soft}}(r)$ is soft for all r. So $U^{\text{soft}}(r)$ never requires a small timestep. The effect of this is to permit a large timestep whenever r exceeds the cutoff. For example with $U(r) = 1/r$ we suggest

$$U^{\text{soft}}(r) = \begin{cases} \frac{3}{2}r_{\text{cut}}^{-1} - \frac{1}{2}r_{\text{cut}}^{-3}r^2, & r \leq r_{\text{cut}}, \\ r^{-1}, & r \geq r_{\text{cut}}; \end{cases}$$

and $U^{\text{hard}} = U - U^{\text{soft}}$. This idea generalizes in a practical way to more than two distance classes. We hope to present results of preliminary experiments with this method.

[1]current address: (Courant Institute of Mathematical Sciences, New York University,) 251 Mercer Street, New York, NY 10012; email: skeel@cs.uiuc.edu

Simulated annealing using a continuous density distribution: Application to a model protein

John E. Straub
Department of Chemistry, Boston University
Boston, Massachusetts 02215

If the native state of the protein is assumed to be the global free energy minimum, the problem of defining the native state configuration is reduced to the problem of finding the global minimum of a complicated many dimensional energy function. This view has been made popular principally through the work of Scheraga and coworkers who have contributed many creative approaches to the difficult problem of optimizing protein structure.[1] The difficulty of finding the global minimum is closely related to the character of the multidimensional energy function – the energy landscape – of the protein. Finding the lowest energy conformation of a large protein, or folding the protein *in simulacro*, is currently an intractable problem – one which cannot be solved in a realizable time using an exhaustive search of configuration space.

Simulated annealing

The paradigm of global optimization of complex systems is simulated annealing.[2] This method makes use of a powerful analogy between the statistical mechanical process of annealing and the challenge of global optimization of a complex cost function. In global energy minimization, where our cost function is the potential energy surface, this requires no stretch of the imagination. The system is simulated at an initially high temperature using molecular dynamics or Monte Carlo. The temperature is slowly lowered according to a cooling schedule, annealing the system to a low or zero temperature. The final configuration of the system is the guess at the global energy minimum. If the cooling schedule is chosen properly, we can expect to find the global minimum with a high probability.

Potential smoothing and coarse graining

Hoare and McInnes[3] noted that softer potentials tend to favor a regular crystal configuration (as the global minimum) while shorter range potentials tend to encourage amorphous structures. In an elegant analysis of the nonlinear optimization problem, Stillinger and Weber[4] recognized that while the number of local energy minima is a strongly (exponentially) increasing function of the system size, how fast the absolute number of minima increases has a great deal to do with the length scale of the interparticle interactions.[3] Systems with short range potentials dominated by nearest neighbor interactions have large numbers of local minima. When the range of interaction is increased, the number of minima can be drastically reduced. The program of smoothing the potential energy hypersurface to remove high lying local minima and deepen and broaden the global energy minimum has been referred to as the "antlion strategy."[4]

My talk will present simulation algorithms which are based on the paradigms of simulated annealing and potential smoothing. In each case, the system is represented by a continuous density distribution $\rho(r,p)$ rather than a simple point in configuration space. Simulated annealing for the continuous classical density distribution in time (by an approximate solution of the Liouville equation $\partial\rho(r,p,t)/\partial t = -\mathcal{L}_0\rho(r,p,t)$) or in temperature (by solution of the Bloch equation, $\partial\rho(r,\beta)/\partial\beta = -\mathcal{H}\rho(r,\beta)$) is described. The algorithms are applied to fold a model protein.

Classical simulated annealing in time

For simulated annealing to be an effective strategy for optimization of proteins, some changes are required. Possibilities are suggested when one thinks of applying the simulated annealing algorithm using an ensemble of systems in parallel. One may then explore a variety of mean field ideas for the approximate integration of an ensemble of systems. Unlike an *ad hoc* potential transform, these methods are rooted in the integration of the classical Liouville equation which describes the time evolution of the classical density distribution (an ensemble of classical systems) just as Newton's equations employed in classical simulated annealing describe the time evolution of a single trajectory.

In the time evolution of an ensemble of classical systems, where each system follows the classical mechanics defined by Newton's equations of motion, the time evolution of the distribution $\rho(r,p,t)$ is described by the Liouville equation

$$\partial\rho(r,p,t)/\partial t = -\mathcal{L}_0\rho(r,p,t) \tag{1}$$

where $\mathcal{L}_0$ is the Liouville operator $\mathcal{L}_0 = (p/m)\cdot(\partial/\partial r) + F(r)\cdot(\partial/\partial p)$, $F(r)$ is the force and m is the mass. $F(r)$, r and p are d-dimensional vectors.

An exact description of the dynamics of $\rho(r,p,t)$ is provided by the equations of motion for the average position and momentum[5]

$$d\langle r\rangle/dt = \langle p\rangle/m \qquad d\langle p\rangle/dt = \langle F\rangle \tag{2}$$

and for the higher-order moments of position and momentum

$$dM_{n,k}/dt = (n/m)M_{n-1,k+1} + kW_{n,k-1}. \tag{3}$$

The moments of the distribution are defined as $M_{n,k} = \langle(r-r_0)^n(p-p_0)^k\rangle$ and $W_{n,k} = \langle(r-r_0)^n(p-p_0)^k(F-F_0)\rangle$ where $r_0 = \langle r\rangle$, $p_0 = \langle p\rangle$, and $F_0 = \langle F(r)\rangle$. The brackets $\langle ...\rangle$ imply an average over the density distribution. Integration of this hierarchy of equations provides an exact description of the dynamics of the ensemble. However, in practice this is intractable since for an anharmonic potential the moments are coupled (up to infinite order moments). Therefore, one must either truncate the moment expansion (which can be numerically unstable) or approximate $\rho(r,p,t)$. As a first approximation, Ma, Hsu and Straub[5] have taken $\rho(r,p,t)$ for a many body system to be a product of single particle distributions represented by d-dimensional spherically symmetric Gaussian phase packets (GPP). Each GPP is completely defined by the first and second moments – the packet center (r_0,p_0) and widths $(M_{2,0},M_{1,1},M_{0,2})$ in phase space.

The appeal of these simple equations of motion comes from their origin as an approximate solution of the Liouville equation for an ensemble of systems, while requiring little more computational investment than is required to integrate Newton's equations for a single representation of the system. In general

$$m\ddot{r}_0 = -\nabla_{r_0}\langle V\rangle \tag{4}$$

which has the form of Newton's equation for the Gaussian center r_0 moving on a coarse-grained effective potential $\langle V\rangle$. When $M_{2,0} = 0$ the equations of motion reduce to the usual equations of classical molecular dynamics. The second moment equations depend on the Laplacian of the effective potential.

In applying the GPP algorithm to an optimization problem we perform a simulated annealing of the continuous density distribution and we must control the temperature during cooling.[5] There

are a variety of ways to do this. One method involves a rigid constraint on the temperature. Other methods involve coupling the system to a heat bath using a Fokker-Planck or BGK collision operator such that the temperature of the system may fluctuate using the generalized Liouvillian[6]

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_c \tag{5}$$

where $\mathcal{L}_0$ is the streaming operator of the system and $\mathcal{L}_c$ is the collision operator which couples the system to the bath.

Classical simulated annealing in temperature

One difficulty in applying simulated annealing to proteins is the definition of a cooling schedule which leads to a reasonable probability of finding the global minimum in a computationally realizable time. An appealing alternative is to replace the cooling schedule necessary when annealing in real or Monte Carlo time with direct integration in temperature.

Ma and Straub[6] have recently explored the possibility of performing simulated annealing directly in temperature by approximately integrating the equation

$$\partial \rho_{eq}/\partial \beta = -(\mathcal{H} - \langle \mathcal{H} \rangle)\rho_{eq}. \tag{6}$$

to obtain the equilibrium classical density distribution for the canonical ensemble $\rho_{eq}(r,p,\beta) = \exp(-\beta\mathcal{H})/Q(\beta)$ at a given temperature. $\mathcal{H}(r,p)$ is the classical Hamiltonian energy function and $Q(\beta) = \int \mathrm{drdp} \exp(-\beta\mathcal{H})$ is the canonical partition function. In fact, Eq. (6) is the classical analog of the imaginary time Schrödinger equation which we have developed into a *quantum mechanical annealing* algorithm.[7] For a many body system it is convenient to make the Hartree approximation to the many body density distribution as a product of single body density distributions (employed in the GPP integration[6] of the Liouville equation)

$$\rho(r,\beta) = (2\pi M_2/d)^{-d/2} \exp[-\frac{d}{2M_2}(r - r_0)^2]. \tag{7}$$

The equations of motion in reciprocal *temperature* for the center r_0 and second moment M_2 of a single Gaussian packet in d-dimensions are

$$\partial r_0/\partial \beta = -(M_2/d)\nabla_{r_0}\langle V \rangle \qquad \partial M_2/\partial \beta = -(M_2/d)^2 \nabla^2_{r_0}\langle V \rangle. \tag{8}$$

$\langle V \rangle$ is the pair potential averaged over the density distribution

$$\langle V \rangle_{GDA}(r_0,\beta) = (2\pi M_2/d)^{-d/2} \int dr' V(r') e^{-d\|r_0 - r'\|^2/2M_2} \tag{9}$$

This effective potential is of the same form that appears in the approximate solution of the Liouville equation using Gaussian phase packets. These equations define the Gaussian Density Annealing (GDA) algorithm.

The form of these equations is quite appealing. The centers move according to a steepest descent energy minimization equation on the effective potential energy surface while the widths of the density distribution adjust themselves to the curvature of the effective potential surface. Therefore, it benefits from the general properties of potential smoothing algorithms. Moreover, the annealing minimization occurs directly in temperature. While the equations define a simulated annealing protocol the algorithm is independent of cooling schedule – if the equations of motion

are integrated accurately we should have the optimal annealing protocol for that representation of the density distribution.

There is an important variation of the GDA algorithm – the "adiabatic GDA" (aGDA) method. For the primitive GDA, problems result when the M_2 values decrease too quickly. Therefore, a useful adiabatic approximation provides that for every set of $\{M_2\}$ the position of the density distribution center relaxes "instantaneously" to its steady state value. The algorithm is applied as follows. (1) The steepest decent equation

$$\dot{r}_o = -\nabla_{r_0}\langle V(r_0, M_2)\rangle \tag{10}$$

is solved with a fixed set of widths $\{M_2\}$ to find the minimum on the effective potential surface for a particular temperature $r_0^{ad}(\beta)$. This can be performed using your favorite local minimizer. (2) The widths of each packet are integrated in β by holding the center position $\{r_0^{ad}\}$ fixed and solving

$$\partial M_2/\partial\beta = -(M_2/d)^2\nabla^2_{r_0}\langle V(r_0^{ad}, M_2)\rangle. \tag{11}$$

(3) Return to step (1) and repeat the cycle.

The resulting algorithm is very much like the Diffusion Equation Method (DEM) of Scheraga and coworkers[8] in that the positions of the centers follow changes in the widths adiabatically. We refer to this method as the adiabatic GDA algorithm (aGDA). An important difference between this algorithm and the DEM is that each width is allowed to evolve in an optimal way which depends on the curvature of the effective potential surface; in the DEM each particle "density" has the same width.

Application to a model protein

Amara and Straub have applied the GDA algorithm to a model protein. The model consists of 22 residues where each residue is a sphere of neutral (N), hydrophobic (B) or hydrophilic (L) nature. The details of this potential can be found elsewhere.[9] The sequence studied is B $(LB)_4$ N_3 $(LB)_5$ which leads to a global energy minimum configuration of a β-sheet. The global energy minimum configuration allows for a maximum number of energetically favorable hydrophobic pair contacts. This is a minimal model for protein folding and represents a serious test of the optimization algorithms. It is a short step from the potential function used for this model protein to the more general all atom empirical potential energy function such as ECEPP or CHARMm.

The more successful algorithms isolate lower energy states with higher probability. Algorithms based on simulated annealing in time (MD, GPP) have the property that a distribution of energies are isolated. Based on these results we find that the aGDA algorithm isolates only the global energy minimum and is most effective. For this system, the GDA algorithm is superior to the dynamical annealing algorithms in providing information on low lying states including the global energy minimum.

Acknowledgments

I thank my students Patricia Amara and Jianpeng Ma who have made many contributions to the work summarized in this chapter. This work was supported in part by a grant from the donors of the Petroleum Research Fund, administered by the American Chemical Society, in part by a grant from the National Science Foundation (CHE-9306375).

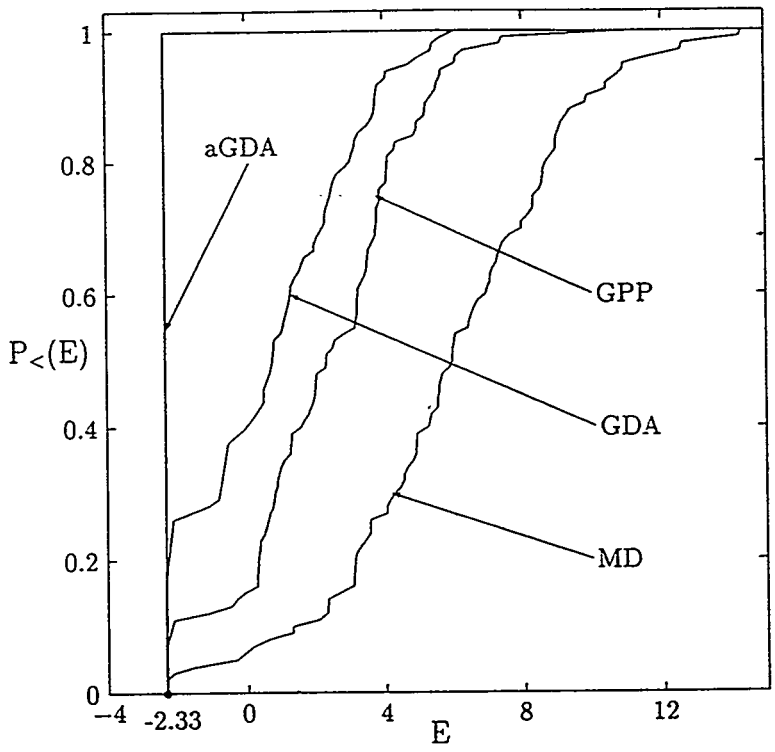

Fig. 1 The probability of locating a configuration with energy E or less from an initial distribution of one hundred independent configurations is plotted for a model protein for four algorithms: simulated annealing in time using molecular dynamics (MD) and Gaussian phase packet dynamics (GPP), simulated annealing in temperature using the Gaussian density annealing (GDA) algorithm, and the "adiabatic" GDA algorithm (aGDA).

References

[1] See H. A. Scheraga, *Rev. Comp. Chem.* III, 73 (1992), Eds. K. B. Lipkowitz and D. B. Boyd; *Theor. Biochem. Mol. Biophys.* **2**, 231 (1991); *Chemica Scripta* **29A**, 3 (1989).

[2] S. Kirkpatrick, Jr. C. D. Gelatt, and M. P. Vecchi. *Science* **220**, 671 (1983).

[3] M. R. Hoare and J. A. McInnes. *Adv. in Phys.* **32**, 791 (1983).

[4] F. H. Stillinger and T. Weber. *J. Stat. Phys.* **52**, 1429 (1988). F. H. Stillinger, *Phys. Rev. B* **32**, 3134 (1985).

[5] J. Ma, D. Hsu, and J. E. Straub. *J. Chem. Phys.* **99**, 4024 (1993).

[6] J. Ma and J. E. Straub. *J. Chem. Phys.* **101**, 533 (1994).

[7] P. Amara, D. Hsu, and J. E. Straub. *J. Phys. Chem.* **97**, 6715 (1993).

[8] J. Kostrowicki, L. Piela, B. J. Cherayil, and H. A. Scheraga. *J. Phys. Chem.* **95**, 4113 (1991).

[9] J. D. Honeycutt and D. Thirumalai. *Biopolymers* **32**, 695 (1992).

MOLECULAR DYNAMICS SIMULATION WITH FRIEDMAN'S IMAGE CHARGE METHOD

Lu Wang and Jan Hermans

Department of Biochemistry & Biophysics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7260

An important technical problem in molecular dynamics simulations is how to simulate a finite region of interest and at the same time incorporate long-range interactions. We have explored the possibility of doing simulations on a droplet system and incorporate the long-range interactions by Friedman's image charge method. This reaction field method is applicable to spherical systems with arbitrary charge distributions and with high dielectric constant such as aqueous solutions. We have tested this scheme with simulations on pure water droplets of radius of 9.5, 11.5 and 14.5 angstroms. We also calculated the hydration free energy of a model cation located at different positions in the droplets. With the reaction field, the radial distribution function is much closer to the result of simulations with periodic boundary condition than that obtained from the simulations without the reaction field. However, the properties of a 3 angstrom surface layer are much perturbed. The calculated hydration free energy of the ion is constant in the interior part of the droplets, but goes up as the ion approaches the surface. Clearly, the reaction field strongly perturbs the surface layer. We found that the problem could be alleviated by not applying the reaction field to the surface layer. With this modification, the calculated hydration free energy of the ion is constant in the major part of the droplets. Also, the reaction field forces between the water molecules could now be ignored for a saving of computer time.

Calculations of Electrostatic Energies in Proteins

Arieh Warshel
Department of Chemistry
University of Southern California
Los Angeles, CA 90089-1062

Calculations of Electrostatic energies provide what is probably the best correlation between structure and function of biological molecules [1-5]. In principle, one can try to use three different strategies. First are microscopic models with explicit solvent molecules and free energy perturbation (FEP) type calculations. Such approaches converge very slowly and could not be used until the mid 80's [6]. Second, simplified microscopic models describe the solvent molecules by dipoles. Such models, in particular the PDLD model [2,5], have allowed one to study proteins quite early [5] and to capture for the first time the physics of solvation effects and electrostatic energies in proteins. Third are macroscopic models that replace the solvent dipoles by polarized volume elements. Such models, which have gained enormous popularity in recent years (e.g. ref. 4), are conceptually very complicated despite the fact that they use extremely familiar electrostatic equations. The main problems are associated with the nonhomogeneous nature of proteins and with the complicated nature of the corresponding dielectric constants. Thus, the early macroscopic studies of proteins have reflected what is basically incorrect physics (see discussion in refs. 2 and 3).

In this lecture we will review the current status of the first two approaches and their relationship to the continuum models. We will concentrate on the following points: (1) new strategies of treating long range interactions in FEP calculations [7]; (2) microscopic FEP calculations of pK_a's in protein [7,9]; (3) microscopic evaluation of the dielectric constants of proteins and the true meaning of these parameters [8]; (4) the advantages of PDLD type models in terms of: (a) simple incorporation of LD dipoles to MD simulations [9], (b) very simple integration with quantum mechanical models including excited electronic states (when the proper continuum treatment is not obvious), and (c) providing a way to move from fully microscopic to a semi-macroscopic description; and finally, (5) we will give some representative examples of recent electrostatic calculations.

References

[1] Warshel, A, Acc. Chem. Res., 14, 284 (1981).

[2] Warshel, A., Russell, S. T., Q. Rev. Biophys., 17, 283 (1984).

[3] Warshel, A., Aqvist, J., Ann. Rev. Biophys., Biophysical Chem., 20, 267 (1991).

[4] Sharp, K. A., Honig, B., Ann. Rev. Biophys., Biophysical Chem., 19, 301 (1990).

[5] Warshel, A., Levitt, M., J. Mol. Biol, 103, 227 (1976).

[6] Warshel, A., Sussman, F., King, G., Biochemistry, 25, 8308 (1980).

[7] Lee, F. S., Warshel, A., J. Chem. Phys. 97, 3100 (1992).

[8] King, G., Lee, F. S., Warshel, A., J. Chem. Phys, 95, 4300 (1991).

[9] Lee, F. S., Chu, Z. T., Warshel., A., J. Comp. Chem., 14, 161 (1993).

Navigating the Energy Landscape of Protein

P. G. Wolynes

School of Chemical Sciences, Univeristy of Illinois, Urbana-Champaign
Urbana, Illinois 68101

I will discuss models of the kinetics of protein folding that go beyond the single reaction coordinate analysis that has been used before. In addition, the effects of correlation in the energy landscape will be discussed, especially vis-a-vis structure prediction.

Continuation-Based Global Optimization for Molecular Conformation and Protein Folding

Zhijun Wu
Mathematics and Computer Science Division
Argonne National Laboratory

This talk is about our recent work on developing a new global optimization algorithm for molecular conformation and protein folding.

A global optimization problem is computationally intractable in general, and is difficult to solve if the problem size is large such as the problems for protein conformation.

The goal of our work is to focus on only the problem for molecular conformation and protein folding, and develop an algorithm that exploits the problem structure so that an efficient solution specific to this class of problems can be found.

In our approach, given a global minimization problem, to avoid directly minimizing a "difficult" objective function, we first use a special technique to transform the function into a class of gradually deformed, but "smoother" or "easier" functions. An optimization procedure is then applied to the new functions successively, to trace their solutions back to the original function.

To deform a given function, we introduce a parametrized integral transformation, transforming a given function into a class of new functions corresponding to a set of parameter values. A transformed function is a coarse approximate to the original function, with small and narrow minimizers being removed while the overall structure of the function is maintained. This property allows an optimization procedure to skip less interesting local minimizers and concentrate on regions with average low function values where a global minimizer is most likely to be located.

The transformed function depends on a parameter λ that controls the degree of smoothing. The original function is obtained if $\lambda = 0$, while smoother functions are obtained as λ increases. We illustrate the transformation process with the problem of finding the global maximizer for the function shown in Figure 1. The transformed functions in Figure 2, clearly

FORCE FIELD AND GEOMETRY OPTIMIZATION FROM A DIVIDE-AND-CONQUER METHOD

Qingsheng Zhao

Department of Chemistry
Duke University
Duhram, NC27706

An efficient way based on divide-and-conquer method in density-functional theory is developed to compute the force field and optimize the geometry for large molecules. Considering the 4th nearest neighbors contribution, the primitive tests show that the average error of the optimized geometry between the method and Kohn-Sham method is very small.

DISCLAIMER