AIIM

Association for Information and Image Management

1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910

301/587-8202

Centimeter

Inches

MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

1 of 1

Conf-9407126--1

Energy Research Power Users Symposium II
Rockville, Maryland; July 12, 1994

Molecular Views of Damaged DNA: Adaptation of the Program DUPLEX to Parallel Architectures[1]

Brian E. Hingerty and Oakley H. Crawford
Health Sciences Research Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831

Suse Broyde
Biology Department
New York University
New York, New York 10003

Robert A. Wagner
Department of Computer Science
Duke University
Durham, North Carolina 27708

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

[1]The submitted manuscript has been authored by a contractor of the U.S. Government under contract number DE-AC05-84OR21400. Accordingly, the U.S. Government retains a non-exclusive royalty-free license to publish or reproduce the published form of the contribution, or allow others to do so for U.S. Government purposes.

Molecular Views of Damaged DNA:
Adaptation of the Program DUPLEX to Parallel Architectures

Abstract. The nucleic acids molecular mechanics program DUPLEX has been designed with useful features for surveying the potential energy surface of polynucleotides, especially ones that are modified by polycyclic aromatic carcinogens. The program features helpful strategies for addressing the multiple minimum problem: (1) the reduced variable domain of torsion angle space; (2) search strategies that emphasize large scale searches for smaller subunits, followed by building to larger units by a variety of strategies; (3) the use of penalty functions to aid the minimizer in locating selected structural types in first stage minimizations; penalty functions are released in terminal minimizations to yield final unrestrained minimum energy conformations. Predictive capability is illustrated by DNA modified by activated benzo[a]pyrenes. The first stage of adaptation to parallel computers is described.

Key Words: Molecular Mechanics, DNA, Carcinogens, Structures, Potential Energy Surface, parallel computing.

Carcinogenesis by Polycyclic Aromatic Chemicals

Polycyclic aromatic hydrocarbons and amines are chemicals present in the environment. Included among them are substances that are known to cause cancer. Chemical carcinogenesis by these substances is now understood to involve a series of initiating events which include the following steps [1-4]: (a) enzymatic conversion of the proximate carcinogen to a reactive species, the ultimate carcinogen; (b) the ultimate carcinogen may be quenched by reaction with solvent and/or it may be enzymatically detoxified, in which case no harm results; (c) the ultimate carcinogen may react with DNA to form one or more covalently linked adducts. These adducts can affect the shape of the DNA, and we believe that the conformation of the modified DNA plays a key role in the outcome of the next steps; (d) the adduct may be repaired or by-passed by mechanisms that are error-free or error-prone; (e) if error-free repair fails to occur, the modified DNA may fail to act as a faithful template during replication. A mutation, a change is base sequence can result; in combination with other such events this eventually can produce cell transformation and tumors. Of course, many biological processes occur between the initial steps and the appearance of tumors. Benzo[a]pyrene (Fig. 1), a substance present in automobile exhaust, is a carcinogenic polycyclic aromatic hydrocarbon that is of great interest. Among its metabolic activation products are a pair of mirror image molecules known as (+) and (-) anti-benzo[a]pyrene-diol-epoxide (BPDE) (Fig. 1). The (+) enantiomer is highly tumorigenic in mice while the (-) member of the pair is not; the (+) enantiomer is also considerably more mutagenic in mammalian systems. Both (+) and (-)-anti-BPDE react with DNA to form covalent adducts.

The major adduct for the (+) enantiomer is to the amino group of guanine, by trans epoxide opening; the (-) enantiomer also forms such an adduct (Fig. 1). Since (+)-anti-BPDE is tumorigenic, its major adduct may well be an important contributor to the carcinogenicity of this molecule; however, the corresponding adduct of the non-tumorigenic (-)-anti-BPDE is not harmful. Because such chemically similar DNA adducts can lead to such different biological outcomes, it is considered likely that adduct structure plays a key role in governing its mutagenicity and ultimate tumorigenicity. The extensive early literature on the benzo[a]pyrenes is reviewed in references [3-5].

We will use our conformational studies on the fascinating (+) and (-)-trans-anti-BPDE adduct pair to illustrate our approach, via the molecular mechanics program DUPLEX, to survey the potential energy surface of carcinogen modified DNAs.

The Molecular Mechanics Program DUPLEX

The task in predicting nucleic acid structure *de novo*, with no experimental input, by molecular mechanics is so formidable that it is usually not attempted by most workers in the nucleic acids arena. To make predictions one must find a way to survey the multi-dimensional surface of the potential energy to locate the structure that corresponds to the lowest energy, and other structures accessible at ambient temperature. This multiple-minimum problem limits the

application of both molecular mechanics and molecular dynamics to macromolecules.

DUPLEX deals with this problem for minimization in a number of ways: (1) it employs internal rather than Cartesian coordinates as the minimization variables. The internal coordinates are the flexible torsion angles that describe the large movements that govern DNA shape. Limiting the allowed motions to these variables (while fixing naturally very small movements like bond lengths, bond angles and those dihedral angles that vary little in nature, such as those in an aromatic moiety) reduces the number of variables that must be simultaneously optimized from 3N - 6 (N = # of atoms) in Cartesian space to 7 for a nucleic acid residue; it also avoids the prediction of chemically unrealistic geometries. Furthermore, the simplification exacts little cost for our application: determining conformations of DNA modified covalently by polycyclic aromatic chemicals. Of course, for other types of investigations, such as studies of reaction pathways, all degrees of freedom must be allowed. (2) We have developed strategies for searching the potential energy surface by minimization. These can employ very large numbers of trials, using arbitrary or selected combinations of the torsion angle variables as starting parameters for small nucleic acid subunits. The resulting minima are energy ranked and then built to larger energy minimized structures by various building strategies [6-12]. These include combinatorial build-up, embedding small units in canonical helices, and adding favored carcinogen orientations found in small units to larger standard helices, all with energy minimization. This procedure again reduces the number of variables in the searches. It is especially useful for helical DNA, whose shape is governed by short range forces. Some of our building strategies are philosophically derived from the build-up technique for polypeptides pioneered in Scheraga's laboratory [13, 14]. (3) Penalty functions are employed as searching tools to aid the minimization algorithm in locating structures on the potential energy surface satisfying specified distance bounds between atom pairs. These functions are useful for finding structures with interproton distances that are within bounds of measured NMR data, or for locating structures with designated hydrogen bonding patterns either within a given DNA strand or between strands. In contrast to many other approaches, the penalty functions are released in the final minimization stage, and unrestrained structures that fail to retain the target distances are rejected. Thus, accepted structures are unrestrained low energy minima.

The specific penalty functions we employ are the following: for restraints on interproton distances

$$F_N = W_N \Sigma_1^n (d - d_N)^2 \quad (1)$$

$$F_{NN} = W_{NN} \Sigma_1^n (d - d_{NN}) \quad (2)$$

W_N and W_{NN} are adjustable weights (usually in the range of 10 - 30 kcal/mol-Å^2), d is the current value of the interproton distance, d_N is a target upper bound, and d_{NN} is a target lower bound. Equation (1) is implemented when d is greater than d_N, and equation (2) is implemented when d is less than d_{NN}. The summation is over the n NMR derived distance bounds. These functions can also be used as overall goodness-of-fit indexes for a given structure in relation to the upper or lower bounds, with the d values now being achieved distances. For restraints that guide the minimizer in locating selected hydrogen bonding patterns, the penalty function, unique to DUPLEX, is

$$F_{HB} = W \Sigma_1^n [(d - d_o)^2 + (1 + \cos(\tau))^2 + p] \quad (3)$$

If not implemented at a given site, this function can also be used to locate denatured structures, which can be important in chemically modified DNAs. Here, d is the current value of the hydrogen bond length, d_o is its ideal value, τ is the current value of the angle about the hydrogen participating in the hydrogen bond, and $p = |\mathbf{c}_1 - \mathbf{c}_2|^2$ where $\mathbf{c}_1$ and $\mathbf{c}_2$ are unit vectors perpendicular to the planes of the hydrogen bonded bases. The function is summed over all n hydrogen bonds. It can also be used to evaluate the overall quality of hydrogen bonds achieved in a model, with d and τ now being the model values.

The force field employed by DUPLEX is one that has been developed for nucleic acids in the laboratory of Olson [15 - 17], with potential functions and parameterization now widely accepted in the nucleic acids community. Details are given in reference 6. In DUPLEX, centurion condensation can be mimicked by reduced partial charges on pendant phosphate oxygens [6], or by explicit metal ions [18]; solvent is modeled by distance dependent dielectric functions developed by Srinivason and Olson [17] and by Hingerty and co-workers [18].

Implementation on a Parallel Computer

For each molecular configuration, DUPLEX computes the energy as a sum of terms, one for each pair of atoms. There are many such terms -- almost N**2/2 for N atoms -- so that in a large molecule most of the CPU time goes into their evaluation.

The code has been adapted for an iPSC/860 hypercube computer with 128 nodes after extensive restructuring. Starting from the dihedral angles, each processor computes its own copy of the atomic coordinates. Then it calculates its predetermined share of the energy terms and sends the sum to a master node, which returns the grand total (total energy). Each processor then calculates a new set of dihedral angles and continues as above.

For a case of 2 strands, having 12 base pairs each, and 759 atoms total, tested with up to 16 processors, the speedup is approximately equal to the number of processors. The program is now being prepared for porting to the new Intel Paragon at Oak Ridge, a 2048 node processor.

Studies of (+) and (-)-trans-anti-Benzo[a]pyrene-diol-epoxide Adducts

Our original studies for the two BPDE adducts of Fig. 1 employed no experimental information other than the force field parameters (7,5). Many thousands of energy minimization trials were first performed for a modified deoxydinucleoside monophosphate d(CpG); additional hundreds of trials were carried out for single stranded trimers, and further trials were carried out with modified duplex trimers that explored the orientation space of the carcinogen-base linkage in that context. Finally, important conformers located from these searches were embedded in duplex 12-mers with energy minimization. The hydrogen bond penalty function was employed in all duplex trials, to search for structures with Watson-Crick, Hoogsteen, or no base pairing at the modification site, and Watson-Crick base pairs at all other loci (except at the site next to the lesion where the possibility of no pairing was also explored). The penalty function was released in terminal minimizations and final structures were energy ranked.

Three types of structures resulted from these trials: structures which placed the pyrenyl moiety in (1) and B-DNA minor groove; (2) the B-DNA major groove; of (3) a base-displaced position, with the modified guanine removed from its normal, base-stacked position and the pyrenyl moiety located in its place. A striking observation in all three structural types was that the pyrenyl group was always oriented oppositely in the enantiomer pair adducts: whether in the minor or major groove, the (+) enantiomer adduct pointed in the 5'-direction of the modified strand, while it was oriented in the 3'-direction for the (-) enantiomer adduct. For the base-displaced conformers, the long axis of the pyrene ring in the (+)-adduct pointed toward the major groove, while in the (-) adduct conformer of this type it pointed toward the minor groove. Energetically, the structure in which the pyrene was in the B-DNA minor groove was lowest in energy for the (+) adduct; for the (-) adduct the minor and major groove type structure were about equal in energy. Base displaced structures were higher energy. Figures 2 and 3 show these structures.

These pair of adducts were subsequently synthesized in a duplex 11-mer by Monique Cosman in the laboratory of N. Geacintov, and analyzed by high resolution NMR studies in the laboratory of D. J. Patel [19,20]. Our computed minor groove structures served as starting conformations for minimizations with the NMR derived distance restraints. The NMR distances were achieved in a single round of minimization and were retained once all restraints were released. The root mean square deviations between the predicted starting structures and the final NMR structures, following sequence and length adjustment of the predicted structures, were only about 1Å.

The prediction that adducts of enantiomer pair polycyclic aromatic hydrocarbon diol-epoxides are oppositely oriented with respect to the DNA helix may be a more general phenomenon. Benzo[c]phenanthrene (BPh) is another carcinogenic polycyclic aromatic hydrocarbons of current interest. It is also metabolically activated to anti-diol-epoxides, like BPDE. The (+) and (-)-anti-benzo[c]phenanthrene-diol-epoxide can form (+) and (-) trans-anti adducts to the amino group of adenine (Fig. 1). Reference 11 summarizes the literature on the benzo[c]phenanthrenes. Synthesis and high resolution NMR studies in the laboratories of

N. Geacintov and D. J. Patel, respectively, together with out computational searches which incorporated the NMR data, have shown that the (+) adduct is intercalated between intact base pairs on the 5' side of the modified strand (11), while the (-) adduct is intercalated on the 3' side (21).

Biological Implications

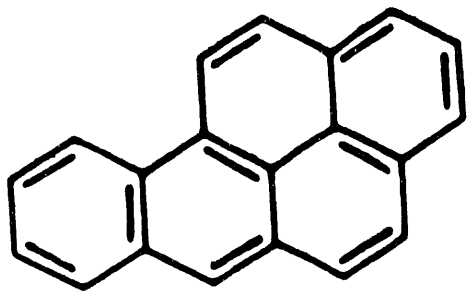
Differing orientations of adducts of mirror-image molecules could well result in different treatments of the lesions by replication or repair enzymes. Indeed, it has been shown in the Geacintov laboratory that exonucleases which digest DNA from the 3'-end treat BPDE modified single strands differently than do those that digest from the 5'-end, depending on the isomer present. Specifically, single strands modified with the (+) adduct slow down a 5'-3' exonuclease more, as would be expected for a 5'-directed orientation of the pyrene, while those modified with the (-) adduct, which is a 3'-directed, slow down a 3'-5' exonuclease more [22]; this shows that enzymes do treat the adduct pair differently, at least *in vitro*. Perhaps the strikingly different biological properties of (+) and (-)-anti-BPDE may result, in some part, from different orientations in DNA.

Figure Captions

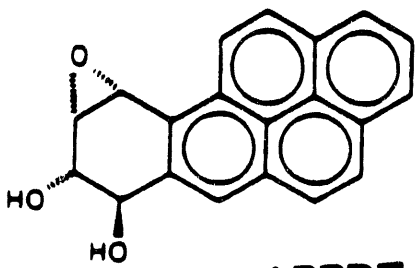
Figure 1. Structures of polycyclic aromatic molecules and DNA adducts.

Figure 2. Structural types computed for (+)-anti-BPDE-trans-N2-dG adducts; A: pyrenyl moiety in B-DNA minor groove; B: pyrenyl moiety in B-DNA major groove; C: pyrenyl moiety in base-displaced position. Reproduced from Figure 2 of *Cancer Research* **51**, 3482-3492, with permission.

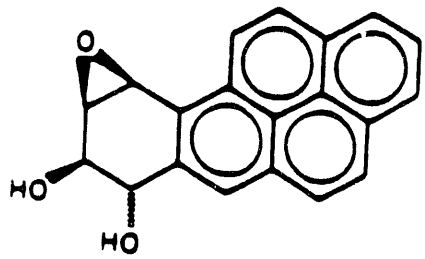
Figure 3. Structural types computed for (-)-anti-BPDE-trans-N2-dG adducts; A: pyrenyl moiety in B-DNA minor groove; B: pyrenyl moiety in B-DNA major groove; C: pyrenyl moiety in base-displaced position. Reproduced from Figure 3 of *Cancer Research* **51**, 3482-3492, with permission.



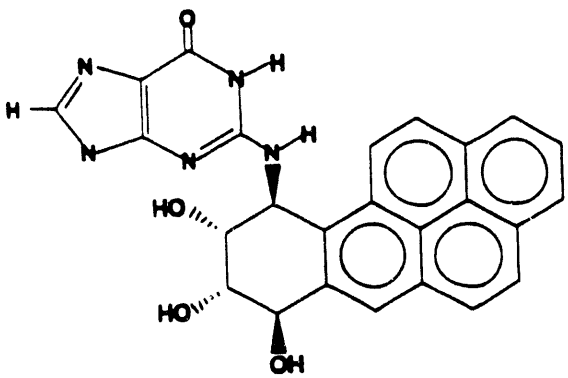
Benzo[a]pyrene



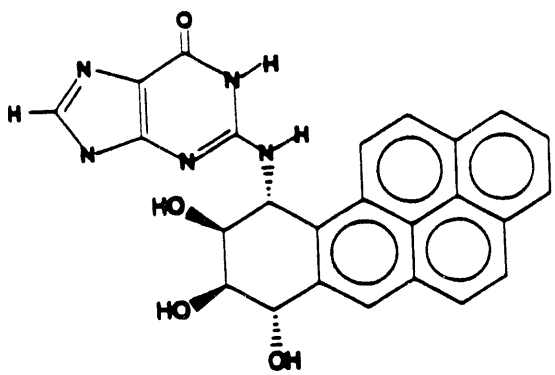
(+)-*anti*-BPDE



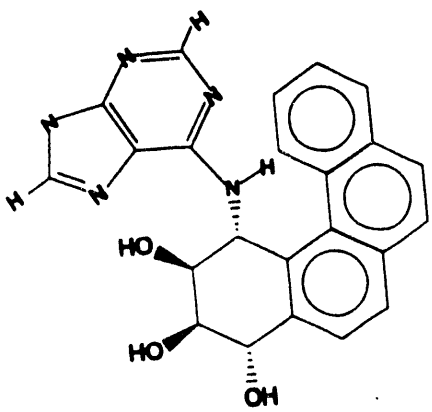
(-)-*anti*-BPDE



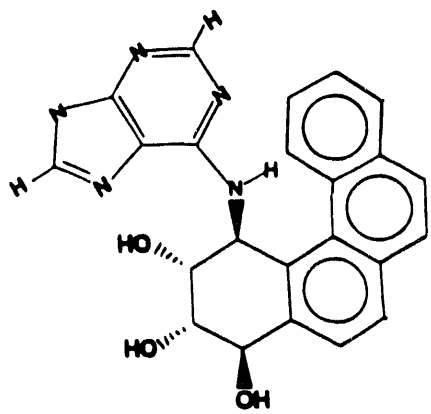
(+)-*anti*-BPDE-*trans*-N2-G



(-)-*anti*-BPDE-*trans*-N2-G

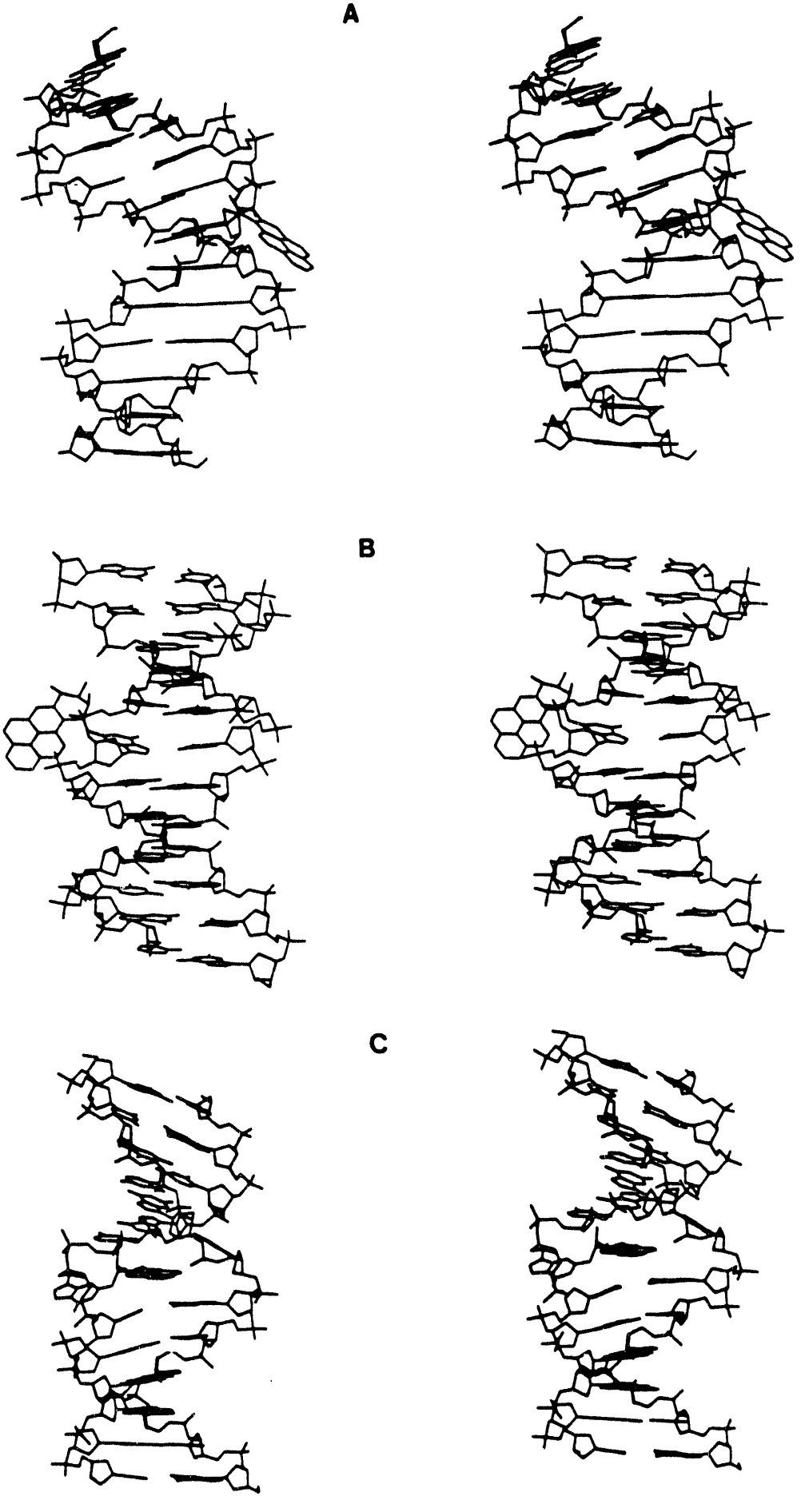


(+)-*anti*-BPh-*trans*-N2-A

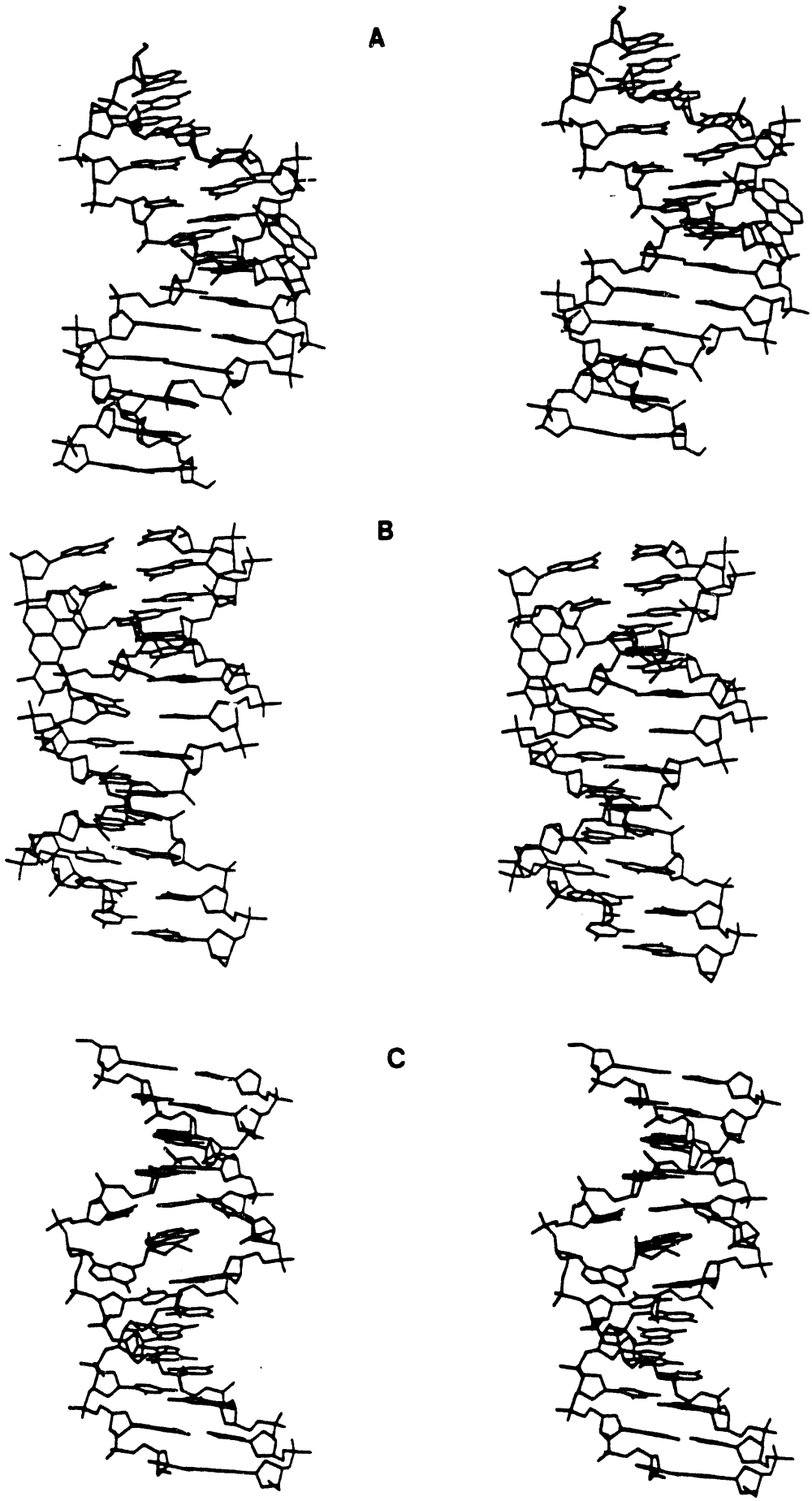


(-)-*anti*-BPh-*trans*-N2-A

Hingerty + Broyde
Fig 1



Mingertzo Broyde
Figure 2



Hingerty + Broyde
Figure 3

References

1. Harris, C. C., *Cancer Res.* (Suppl.) **51**, 50235-50445 (1991).

2. Cohen, S. M. and Ellwein, L. B., *Cancer Res.* **51**, 6493-6505 (1991).

3. Conney, A. H., *Cancer Res.* **42**, 4875-4917 (1982).

4. Harvey, R., *Am. Sci.* **70**, 386-393 (1982).

5. Singh, S. B., Hingerty, B. E., Singh, U. C., Greenberg, J. P., Geacintov, N. E., and Broyde, S., *Cancer Res.* **51**, 3482-3492 (1991).

6. Hingerty, B. E., Figueroa, S., Hayden, T. L., and Broyde, S., *Biopolymers* **28**, 1995-1222 (1989).

7. Hingerty, B. E. and Broyde, S., *Biopolymers* **24**, 2279-2299 (1985).

8. Norman, D., Abuaf, P., Hingerty, B. E., Live, D., Broyde, S., and Patel, D. J., *Biochemistry* **28**, 7462-7476 (1989).

9. O'Handley, S. F., Sanford, D. G., Xu, R., Lester, C. C., Hingerty, B. E., Broyde, S., and Krugh, T. R., *Biochemistry* **32**, 2481-2497 (1993).

10. Carothers, A. M., Yuan, W., Hingerty, B. E., Broyde, S., Grunberger, D., and Snyderwine, E. G., *Chem. Res. Toxicol.* **1**, 209-218 (1994).

11. Cosman, M., Fiala, R., Hingerty, B. E., Laryea, A., Lee, H., Harvey, R., Geacintov, N. E., Broyde, S., and Patel, D. J., *Biochemistry* **32**, 12488-12497 (1993).

12. Shapiro, R., Hingerty, B. E., and Broyde, S., *J. Biomolec. Struct. and Dynam.* **7**, 493-513 (1989).

13. Pincus, M., Klausner, R. and Scheraga, H., *Proc. Natl. Acad. Sci. USA* **79**, 5107-5110 (1982).

14. "The multiple minimum problem in protein folding." Gibson, K. D., and Scheraga, H. A. In *Structure and Expression Volume I: From Proteins to Ribosomes*, Sarma, R. H., and Sarma, M. H., Eds. Adenine Press (1988) pp. 67-94.

15. Olson, W. K. and Srinivasan, A. R. "Classical energy calculations of DNA, RNA and their constituents." In *Landolt-Bornstein, Numerical Data and Functional Relationships in Science and Technology, Group VII Biophysics* **Vol. 1D**, W. Saenger, Editor, Springer Verlag, Berlin (1990) 415-435.

16. Taylor, E. R. and Olson, W. K., *Biopolymers* **22**, 2667-2702 (1983).

17. Srinivasan, A. R. and Olson, W. K., *Fed. Proc.* **39**, 2199 (1980).

18. Hingerty, B. E., Ritchie, R. H., Ferrell, T. L., and Turner, J. E., *Biopolymers* **24**, 427-439 (1985).

19. Cosman, M., de los Santos, C., Fiala, R., Hingerty, B. E., Singh, S. B., Ibanez, V., Margulis, L., Live, D., Geacintov, N. E., Broyde, S., and Patel, D. J., *Proc. Natl. Acad. Sci. USA* **89**, 1914-1918 (1992).

20. de los Santos, C., Cosman, M., Hingerty, B. E., Ibanez, V., Margulis, L., Geacintov, N. E., Broyde, S., and Patel, D. J., *Biochemistry* **31**, 5245-5252 (1992).

21. Cosman, M., Fiala, R., Hingerty, B. E., Laryea, A., Lee, H., Harvey, R. G., Amin, S., Geacintov, N. E., Broyde, S., and Patel, D. J., Manuscript in preparation.

22. Mao, B., Li, B., Amin, S., Cosman, M. and Geacintov, N. E., *Biochemistry* **32** 11785-11793 (1993).

Acknowledgement

Supported in part by NIH Grants CA 28038 (S.B.) and RR 06458 (S.B.), DOE Grant DE FG0290ER60931 (S.B.), DOE contract DE-AC05-84OR21400 with Martin Marietta Energy Systems (B.E.H.), and DOE, Office of Health and Environmental Research, Field Work Proposal ERKP931 (B.E.H.). Calculations were carried out on Cray Supercomputers at the DOE's National Energy Research Supercomputer Center and the NSF's San Diego Supercomputer Center. We thank Professor Nicholas Geacintov and Professor Robert Shapiro, Chemistry Department, New York University, for on-going helpful discussions and advice.

DATE
FILMED
11/15/94

END