1 of 1

Conf 9306237 4

The Torsion Space Nucleic Acids Molecular Mechanics Program DUPLEX: Surveying Conformation Space by Potential Energy Minimization; A Progress Report

Brian E. Hingerty
Health Sciences Research Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831

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

RECEIVED
NOV 08 1993
OSTI

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 this contribution, or allow others to do so for U.S. Government purposes.

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

ds
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

It has been 21 years since the inception of the program DUPLEX. Now that "molecular modeling" (encompassing quantum mechanics, molecular mechanics and dynamics, and computer graphics) has become something of a trade, the time seems ripe for us to offer a brief perspective of our efforts since the beginnings of the field.

DUPLEX originated in the pioneering Computer Graphics Laboratory of Robert Langridge in the Department of Biochemical Sciences, Frick Chemistry Laboratory, at Princeton University. Computer graphics and molecular mechanics as tools for the solution of macromolecular structures were in their infancy, but were major research focuses in that laboratory. The original version of DUPLEX was written by Steven D. Stellman, Brian E. Hingerty and Suse Broyde. The purpose in writing the program was to solve the crystal structure of Ca^{+2} -GpC, an RNA subunit for which x-ray diffraction data had been collected by T. Sato. No trial structure was available to aid in solving the phase problem in the Fall of 1972 and we hoped to generate one by potential energy minimization. We hoped to compute not only the conformation of the GpC molecule, but also its packing scheme in the unit cell, a feat that had never before been performed for a nucleic acid subunit. Indeed, only one dinucleoside monophosphate crystal structure, that of UpA, had been previously solved (Sussman et al., 1972; Rubin et al., 1972) and its structure was very different from that of GpC. To our amazement, the effort succeeded in a matter of months (Stellman et al., 1972). As far as we are aware, it has actually not been duplicated since then.

Many of the features of the current version of DUPLEX were included in the original version. We followed the philosophy of Scheraga's pioneering work on hydrocarbons and amino acids in employing the reduced variable domain of torsion angle

space and in using the linked atom algorithm to generate coordinates, in employing helix conformational building blocks, in our case, the dinucleoside monophosphates (DNA or RNA), as the smallest modules treated, and in using the Powell algorithm (Powell, 1964) for potential energy minimization (Scott and Scheraga 1965; 1966a; 1966b). Thus, the only variable parameters were the flexible torsion angles about single bonds. Sugar puckers were generated in any desired conformation using a pseudorotation treatment (Altona and Sundaralingam, 1972) devised by Sato (Sasisekharan, 1973; Sato, 1983), but the conformation was not varied during minimizations. The force field, also following Scheraga, employed a 6-12 Lennard-Jones potential, a coulombic term with fixed dielectric constant of 4, and a term to account for the intrinsic barrier to rotation about single bonds, using published parameters (Renugopalakrishnan, et al., 1967; Lakshminarayanan and Sasisekharan, 1969a; 1969b). To aid in the search for correct intermolecular interactions, we added a hydrogen bond penalty function, a pseudo-potential that could locate Watson-Crick base paired partners. A subroutine called ORIENT was employed to generate coordinates of a properly positioned anti-parallel Watson-Crick partner strand to begin minimizations. Only one single or double stranded dinucleoside monophosphate could be treated. During minimizations the conformations of the two strands were the same. The philosophy of surveying conformation space by performing potential energy minimization trials (as many as computing resources would allow—at that time we worked on a DEC PDP-10 and made 18 trials to compute the GpC conformation) from starting conformations that simply combined selected rotamers for each flexible torsion (to give high energy starting structures) also emanated from this work. Details about this version of DUPLEX, including results of the successful conformational and packing searches, and

features of the program unique to the need for solving a crystal structure are described in Stellman et al., 1973. Computer graphics visualization was performed with software developed in the Langridge laboratory on an Evans and Sutherland LDS-1.

Over the years DUPLEX has been improved. Single and double stranded structures with up to 24 residues of any base sequence on each strand can now be investigated, and each strand is conformationally independent. Covalent modification of the bases by certain mutagenic and tumorigenic aromatic hydrocarbons and amines, the focus of our interest for over a decade, is implemented and can readily be extended to new lesions. In addition to the original variables, namely, the torsion angles, the sugar pucker is also a flexible parameter, with the pseudorotation parameter as a variable. Mismatched base pairs, bulged out structures with different numbers of residues on each strand and hairpin loops in single strands can now be treated using the hydrogen bond penalty function as a tool for locating structures with designated hydrogen bonding patterns. This function can also be employed to search for any type of hydrogen bonding scheme between Watson-Crick partners, or, if the function is not implemented at a chosen residue, a denatured site may be located. The various structures are energy ranked, permitting an assessment of the relative importance of different hydrogen bonding schemes. This is especially valuable in light of our interest in mutagenesis, since structures with unusual hydrogen bonding may be responsible for miscoding (Kunkel, 1990; 1992). Use of the hydrogen bond penalty function to locate non-Watson-Crick pairing schemes, together with a modified subroutine ORIENT, now permits generation of parallel as well as anti-parallel double helices. Terminal minimizations without the penalty function ensure that final structures are unconstrained energy minima and that the penalty function is merely a search technique for addressing

the multiple minima problem.

The current force field is essentially one developed and tested in the laboratory of Wilma Olson (Srinivasan and Olson, 1980; Taylor and Olson, 1983; reviewed in Olson and Srinivasan, 1990), and is detailed in Hingerty et al. (1989). It includes anomeric correction terms for the sugars (Olson, 1982) and phosphates, as well as a treatment for hydrogen bonding (Olson, 1978) that employs the Lennard-Jones potential in conjunction with altered parameters that permit the necessary closer approach of designated hydrogen bonding partners. Distance dependent dielectric functions (Taylor et al., 1983; Olson and Srinivasan, 1990; Hingerty et al., 1985) are implemented in the coulombic term to simulate the increased dielectric constant that would result when water interpenetrates an atom pair whose separation increases. Partial charges in the coulombic term are from Ornstein and Rein (1979). A number of different metal ions can be treated explicitly (Hingerty et al., 1985). However, counterion condensation is usually mimicked by a reduced partial charge on the pendant phosphate oxygens.

Penalty functions to permit upper and lower bound distance constraints, such as the interproton distances derived from NMR data, have been implemented (see for example O'Handley et al., 1993). Like the hydrogen bond penalty function, these are employed only as an aid to the minimizer in locating structures within the bounds of the data. Final minimizations are always performed without penalty functions, so that completed structures are unconstrained energy minima.

DUPLEX is now running on the Cray Supercomputers at the National Energy Research Supercomputer Center of the Department of Energy, and the National Science Foundation's San Diego Supercomputer Center.

To survey conformation space without experimental information from high resolution NMR data, we have developed search strategies that employ many trials for small subunits followed by various strategies to generate larger structures. These include build-up and building block repetition (Hingerty et al., 1989), add-on (Shapiro et al., 1989) and embedding (Singh et al., 1991). With these techniques it has been possible to achieve some encouraging results. Structures for 2-aminofluorene (Eckel and Krugh, 1993) and 4-aminobiphenyl (Cho et al., 1992) modified DNA that place the carcinogen in the B-DNA major groove, recently observed in solution by high resolution NMR, were predicted in our computations some years ago (Broyde and Hingerty, 1983; Broyde et al., 1985). We computed B and Z DNA helices as the two lowest energy structures in alternating G-C duplexes (Hingerty et al., 1989). Also, we predicted that the pyrenyl moieties, in (+) and (-) anti-benzo[a]pyrene-diol-epoxide bound covalently to the guanine amino group via trans epoxide opening, point in opposite directions in an alternating G-C sequence (Singh et al., 1991). High resolution NMR data for a different sequence independently showed that opposite orientations do occur in solution (de los Santos et al., 1992; Cosman et al., 1992). Since (+) anti-benzo[a]pyrene-diol-epoxide is tumorigenic in mice while its (-) enantiomer is not (reviewed in Conney, 1982), the opposite alignment of the pyrenyl moiety in relation to the DNA may play a role in producing the different biological effects.

When high resolution NMR data is available to us from collaborators, in the form of interproton distances, we also employ these in our conformational searches via the penalty functions. In this case the number of trials needed is much less, but multiple trials are still carried out. Typically we survey the orientation space of a carcinogen in relation to the DNA via a series of energy minimizations, starting with a small subunit such as a

duplex trimer. In these trials we utilize certain other data from the NMR results, such as the overall form of the DNA (B-form, for example), the hydrogen bonding patterns, and the conformation of glycosidic linkages, to generate starting conformations and to select targets for the hydrogen bond penalty function. The searches yield energy ranked structures that are evaluated in relation to the NMR derived interproton distances by goodness-of-fit-functions. We find that the structures with the best goodness-of-fit are also lowest in energy. The best small subunits are built to larger polymers with energy minimization and all constraints are released. The structures retain their NMR characteristics when unconstrained. Thus, the NMR guided structures generated by our approach are not only models that match the data, but, importantly, low energy conformations that are not restrained. A number of carcinogen modified DNA solution structures have been elucidated by this approach (Norman et al., 1989; Cosman et al., 1992; de los Santos et al., 1992; Cosman et al., 1993; O'Handley et al., 1993; Cosman et al., submitted).

In conclusion, the essential features of DUPLEX, put in place more than two decades ago, seem to have well stood the test of time. Our current good force field from Wilma Olson's laboratory, together with the original choice of using torsion angles as the conformational variables, have yielded a useful mechanism for surveying conformation space. The reduced variable set of torsion space, together with search strategies that explore small subunits first, permit location of low energy structures from high energy starting positions. In addition, geometric structural anomalies (such as non-planar aromatic rings) are avoided when only torsion angles about single bonds (and pseudorotation parameters) are allowed to flex. Prediction of structure in the absence of data, though

difficult, is possible with large numbers of trials. With NMR data to guide the searches in first stage trials, generation of unconstrained low energy forms that are within bounds of the data is feasible with fewer trials.

References

- Altona, C. and Sundaralingam, M. (1972) *J. Am. Chem. Soc.* 94, 8205-8212.
- Broyde, S. and Hingerty, B. (1983) *Biopolymers* 22, 2423-2441.
- Broyde, S., Hingerty, B. E., and Srinivasan, A. R. (1985) *Carcinogenesis* 6, 719-725.
- Cho, B. P., Beland, F. A., and Marques, M. M. (1992) *Biochemistry* 31, 9587-9602.
- Conney, A. H. (1982) *Cancer Res.* 42, 4875-4917.
- Cosman, M., de los Santos, C., Fiala, R., Hingerty, B. E., Singh, S. B., Ibanez, V., Margulis, L. A., Live, D., Geacintov, N. E., Broyde, S., and Patel, D. J. (1992) *Proc. Natl. Acad. Sci. USA* 89, 1914-1918.
- Cosman, M., de los Santos, C., Fiala, R., Hingerty, B. E., Ibanez, V., Luna, E., Harvey, R., Geacintov, N. E., Broyde, S., and Patel, D. J. (1993) *Biochemistry* 32, 4145-4155.
- Cosman, M., Fiala, R., Hingerty, B. E., Laryea, A., Lee, H., Harvey, R. G., Amin, S., Geacintov, N. E., Broyde, S., and Patel, D. (submitted for publication)
- de los Santos, C., Cosman, M., Hingerty, B. E., Ibanez, V., Margulis, L. A., Geacintov, N. E., Broyde, S., and Patel, D. J. (1992) *Biochemistry* 31, 5245-5252.
- Eckel, L. M. and Krugh, T. R. (1993) *Eighth Conversation in Biomolecular Stereodynamics, Journal of Biomolecular Structure and Dynamics*, 10, a40.
- Hingerty, B. E., Figueroa, S., Hayden, T., and Broyde, S. (1989) *Biopolymers* 28, 1195-1222.
- Kunkel, T. A. (1990) *Biochem.* 29, 8003-8011.
- Kunkel, T. A. (1992) *J. Biol. Chem.* 267, 18251-18254.
- Lakshminarayanan, A. V. and Sasiseskharan, V. (1969a) *Biopolymers* 8, 475-488.
- Lakshminarayanan, A. V. and Sasiseskharan, V. (1969b) *Biopolymers* 8, 489-503.

- Norman, D., Abuaf, P., Hingerty, B. E., Live, D. Grunberger, D., Broyde, S., and Patel D. (1989) *Biochemistry* 28, 7462-7476.
- O'Handley, S. F., Sandford, D. G., Xu, R., Lester, C. C., Hingerty, B. E., Broyde, S., and Krugh, T. R. (1993) *Biochemistry* 32, 2481-2497.
- Olson, W. K. (1978) *Biopolymers* 17, 1015-1040.
- Olson, W. K. (1982) *J. Am. Chem. Soc.* 104, 270-286.
- Olson, W. K. and Srinivasan, A. R. (1990) "Classical Energy Calculations of DNA, RNA and their Constituents" in Landolt-Börnstein, Numerical Data and Functional Relationships in Science and Technology, Group VII. Biophysics, Volume 1D, W. Saenger, Editor, Springer Verlag, Berlin, 1990, pp. 415-435.
- Ornstein, R. and Rein, R. (1979) *Biopolymers* 18, 1277-1291.
- Powell, M. (1964) *Comput. J.* 7, 155-159.
- Renugopalakrishnan, V., Lakshminarayanan, A., and Sasisekharan, V. (1991) *Biopolymers*, 10, 1159-1167.
- Rubin, J., Brennan, T., and Sundaralingam, M. (1972) *Biochemistry* 11, 3112-3218.
- Sasisekharan, V. (1973) *Jerusalem Symp. Quant. Chem. Biochem.* 5, 247-260.
- Sato, T. (1983) *Nucleic Acids Res.* 11, 4933-4938.
- Scott, R. and Scheraga, H. A. (1965) *J. Chem. Phys.* 42, 2209-2215.
- Scott, R. and Scheraga, H. A. (1966a) *J. Chem. Phys.* 44, 3054-3069.
- Scott, R. and Schegara, H. A. (1966b) *J. Chem. Phys.* 45, 2091-2101.
- Shapiro, R., Hingerty, B. E., and Broyde, S. (1989) *J. Biomolec. Struct. and Dynam.* 7, 493-513.
- Singh, S. B., Hingerty, B. E., Singh, U. C., Greenberg, J. P., Geacintov, N. E., and Broyde,

S. (1991) Cancer Res. 51, 3432-3492.

Srinivasan, A. R. and Olson, W. K. (1980) Fed. Proc. Fed. Am. Soc. Exp. Biol. 39, 2199.

Stellman, S. D., Hingerty, B., Broyde, S., Subramanian, E., Sato, T., and Langridge, R.

(1973) Biopolymers 12, 2731-2750.

Sussman, J., Seeman, N., Kim., S.-H., and Berman, H. (1972) J. Mol. Biol. 66, 403-421.

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

Acknowledgement

This work is supported by grants NIH CA 28038, NIH RR06458 and DOE DE-FG0290ER60931 to SB, and DOE Contract DE-AC05-84OR21400 with Martin Marietta Energy Systems and DOE, Office of Health and Environmental Research, Field Work Proposal ERKP931 (BEH). The idea of creating a trial structure to solve the phase problem by potential energy minimization searches was Steven D. Stellman's and he wrote the bulk of the original code E. Subramanian introduced BEH to the methods of crystallography. The scientifically exciting atmosphere created by Robert Langridge in his laboratory at Princeton provided the impetus for this work.

**DATE
FILMED**

12 / 20 / 93

END