Journal of Medicinal Chemistry

Volume 31, Number 9 September 1988

Perspective

An Analysis of Current Methodologies for Conformational Searching of Complex Molecules

Allison E. Howard and Peter A. Kollman*

Department of Pharmaceutical Chemistry, University of California—San Francisco, San Francisco, California 94143.
Received March 28, 1988

1. Introduction

As theoretical methods become more sophisticated and computer resources more accessible, computational chemistry is being seen as an increasingly useful tool to both industrial and academic institutions. The computational description of a molecule generally involves input of the geometry or the configuration of atoms in that molecule and then the application of a theoretical method, such as semiempirical quantum mechanics, ab initio quantum mechanics, or molecular mechanics, to evaluate the physical properties of the molecule. Some small, rigid molecules, such as ethylene, will have only one conformation which contributes to the ground-state properties of the system. Other somewhat larger and more flexible molecules will have several conformations populated at room or physiological temperature. For example, approximately 60% of *n*-butane molecules exist as the anti rotomer, in the gas phase at room temperature, and 40% exist as the gauche rotomers.[1] Hence, a critical calculation of some intramolecular physical properties of *n*-butane may require that both the anti and gauche conformers be used as input geometries. It is generally appreciated that intermolecular interactions may also be very conformationally dependent. The uncomplexed crown ether 1,4,7,10,13,16-hexaoxacyclooctadecane (18-crown-6) has been demonstrated by X-ray crystallography[2] to be in a C_i conformation; however, the 1:1 complex with the potassium cation has been shown[3] to contain the crown ether in a D_{3d} conformation. Here again some calculations, such as the binding energy of 18-crown-6 with a potassium cation, may require geometrical information regarding several possible conformations of the individual molecules.

Given the importance of conformational information with regard to accurate calculations of intramolecular and intermolecular properties in computational chemistry, how does one generate the several possible conformers of *n*-pentane, or the hundreds of feasible conformations of cyclooctane, or the approximately 7×10^{48} plausible backbone conformations[4] of the enzyme papain in order to evaluate which are populated at a specific temperature and which may be important to the chemistry of the molecule? Is is necessary or even possible to generate all of the conformations of a molecule? It is these questions which we address in the present review. Specifically, our goal in writing this review is to acquaint the reader with various methods that have been used to generate conformational surfaces for molecules and to illustrate cases where these methods might be most appropriate. In addition, we hope to show the necessity for examining conformational possibilities in computational chemistry and the need for continuing research in this area. It is not our aim to write a comprehensive review of the work done in this area, rather we hope to provide a discussion of some of the techniques and applications of conformational searching. Also, the reader should appreciate the distinction between the generation of molecular conformations and the subsequent evaluation of their energies to determine their relative importance to the ground-state properties of a molecule, for we examine in some detail only the former issue. Although the latter issue of the calculation of conformational energy is often interwoven with their generation, it is also sufficiently broad and complex to require a separate discussion.

2. Interactive Molecular Modeling Techniques and Database Retrieval

Certainly most, if not all, chemists have had experience using hand-held molecular models such as Dreiding or CPK molecular representations. Many possible conformations of small molecules may be rapidly generated by using this method. Those conformers possessing unfavorable steric or intramolecular interactions may be noted and reasonable conformations may be projected onto paper for calculation of coordinates. Computer graphical systems extend the usefulness of hand-held models to moderately sized compounds and large biomolecules. The geometry of the molecule may be built up from substructures stored

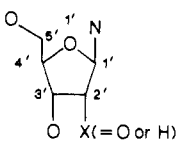
(1) Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1959**, *32*, 748.
(2) (a) Dunitz, J. D.; Seiler, P. *Acta Crystallogr., Sect. B* **1974**, *B30*, 2739. (b) Dale, J. *Israel J. Chem.* **1980**, *20*, 3.
(3) Seiler, P.; Dolzler, M.; Dunitz, J. D. *Acta Crystallogr., Sect. B* **1974**, *B30*, 2744.
(4) Dill, K. A. *Biochemistry* **1985**, *24*, 1501. The number of conformations for the enzyme was estimated from the formula $(z/a)^n = (1.7)^n$, where the number of residues, n, is equal to 212 for the enzyme papain. Further details may be found in the paper.

0022-2623/88/1831-1669$01.50/0

in the computer or may be read in directly from databases similar to the Brookhaven National Laboratory Protein Data Bank[5] or Cambridge Structural Database.[6] Furthermore, the output of molecular geometries is easily accomplished with a computer modeling system.

While the method of interactive molecular modeling is simple, rapid, and widely used, it suffers from two conspicuous deficiencies. The generation of conformers is limited by the knowledge, imagination, patience, and bias of the researcher. Cyclic compounds produce an additional complication in that ring closure cannot always be maintained during the modeling process.

The computer structural databases can be efficiently exploited to obtain conformational possibilities for chemical residues. The basis for this technique is that the large databases of molecules contain information relevant to the low-energy conformations of residues if one can identify the effects of crystal packing forces, experimental errors and the ability of the database to provide a representative sample of compounds. Since a large number of conformations can conceivably be obtained by using this technique, it is desirable to be able to group the conformations into classes after retrieval. Procedures for obtaining and analysing fragments from the Cambridge Structural Database have been described.[7,8] Subsequent statistical analysis[9,10] on the retrieved conformations of one fragment, β-1′-aminoribofuranoside, have revealed (1) the physical factors which govern the conformations of the fragment, (2) the classes of conformations accessible to the fragment in the crystal state, and (3) that the fragment is not conformationally limited[10] to the C(2′)-endo and C(3′)-endo configurations as suggested by other workers.



β-1′-aminofuranoside fragment

The idea that structural databases contain information describing the natural conformations of protein fragments was first described one decade ago[11] and has recently been exploited rather elegantly.[12] Ponder and Richards[12] have attempted to produce tertiary templates for proteins, that is, lists of allowed internal residue sequences compatable with specific protein structural classes. The templates are generated by using three essential criteria: (1) trial residues do not sterically interfere with other protein main-chain atoms, (2) the side chains of the trial residues do not suffer from steric interference with other side chains in the protein, and (3) the packing density of the trial sequence be similar to the native molecule. This algorithm would be unmanageable without restricting the possible side-chain conformations of the amino acid residues. The workers[12] have therefore created a rotomer library of the side-chain conformations from 19 highly resolved proteins (2273 residues) in the Brookhaven Protein Data Bank. In doing so, they have reduced to 67 the side-chain orientation possibilities for 17 of the 20 standard amino acid residues.

In another study pertinent to protein side-chain conformations, it was shown[13] that there is a high transferability of the γ and δ side-chain atomic positions between homologous proteins. A comparison of seven homologous protein pairs from the Brookhaven Protein Data Bank revealed transferability for the γ and δ atoms of identical residues to be 60–97% and 50–90%, respectively (the δ atom percentages were taken from those cases where the γ atoms were found to be matched). The transferability was found to be only slightly lower for dissimilar residues. Poor transferability was typically associated with residues engaging in new intramolecular interactions, a high calculated surface accessibility for the residue side chain, high B-factors, or with residues having broad or multiple rotational minima.

The use of structural databases to obtain conformational possibilities is not limited to small fragments. The structure of retinol binding protein can be reproduced with segments of three proteins from the Brookhaven Protein Data Bank.[14] Using a computer sequence matching algorithm, researchers found that when the root mean square (RMS) fit between C-α carbons is constrained to be less than 1 Å, 169 amino acids of the protein could be built from 15 database fragments and 20 fragments allow the fit to improve to a RMS of 0.5 Å.

3. Systematic Search Methods

The largest structural differences which exist between various conformations of a molecule generally lie not in the bond lengths or bond angles but in the dihedral angles. This is due to the greater energy needed to distort a bond or bond angle as compared to a dihedral angle. Therefore, an obvious method for conformation generation is to vary systematically each of the dihedral angles in the molecule by some increment, while keeping the bond lengths and bond angles fixed, to obtain all of the dihedral angle combinatorial possibilities for the compound. Such a method is known as a grid search. Assuming the increment is appropriately small, the grid search method will generate all conformational possibilities. The disadvantage of the method is that without some way to select only those conformers which are unique and relatively low-energy local minima on the conformational hypersurface, the generation and energy evaluation of conformations may become intractable. For example, a molecule with three dihedral angles will generate 216 configurations with a dihedral angle increment of 60° (i.e., 6^3). A molecule with six dihedral angles will produce 46 656 configurations while nine dihedral angles will lead to more than ten million configurations.

It appears that surprisingly little effort has been given toward determining optimal dihedral angle increments in grid searches. Lipton and Still[15] have explored this issue for a number of hydrocarbon compounds and found that a dihedral step size of 60° was sufficient to obtain all conformational possibilities. They have proposed that a liberal procedure might be to use a dihedral resolution of

(5) Bernstein, F. C. Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
(6) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. *Acta Crystallogr., Sect. B* **1979**, *B35*, 2331.
(7) Murray-Rust, P.; Motherwell, S. *Acta Crystallogr., Sect. B* **1978**, *B34*, 2518.
(8) Murray-Rust, P.; Bland, R. *Acta Crystallogr., Sect. B* **1978**, *B34*, 2527.
(9) Murray-Rust, P.; Motherwell, S. *Acta Crystallogr., Sect. B* **1978**, *B34*, 2534.
(10) Taylor, R. *J. Mol. Graphics* **1986**, *4*, 123.
(11) Janin, J.; Wodak, S.; Levitt, M.; Maigret, B. *J. Mol. Biol.* **1978**, *125*, 357.
(12) Ponder, J. W.; Richards, F. M. *J. Mol. Biol.* **1987**, *193*, 775.
(13) Summers, N. L.; Carlson, W. D.; Karplus, M. *J. Mol. Biol.* **1987**, *196*, 175.
(14) Jones, T. A.; Thirup, S. *EMBO J.* **1986**, *5*, 819.
(15) Lipton, M.; Still, W. C. *J. Comput. Chem.* **1988**, *9*, 343.

120° for X–C_{sp^3}–C_{sp^3}–X dihedral angles and 60° for X–C_{sp^3}–C_{sp^2}–X or X–C_{sp^3}–heteroatom–X angles. However, it is unclear whether the dihedral step size could be even more lenient for classical grid searches since Lipton and Still's[15] conclusions are based on a program which included constraints for van der Waals violations and ring closure for cyclic compounds. Their proposal for the dihedral resolution during grid searches can probably be extented beyond hydrocarbons since several groups[11,12,16] have examined databases of protein structures and have found the populated χ_1 and χ_2 angles for the amino acid residues to be the typically the staggered values.

The grid search method is inherently appealing because it is an exhaustive technique, and much work has been done to improve the method. Advancements in tree searching algorithms,[17] in conjunction with restricted searching in some regions of conformational space, enable the upper limit of dimensionality to be extented from around five to six dihedral angles for acyclic systems to more than 20 dihedral angles. However, most improvements focus on filtering algorithms which eliminate unreasonable portions of conformational space from consideration. In a study of the cyclic polypeptide cyclohexaglycyl, workers[18] first limited their search to allowed regions from the Ramachandran map[19] of *N*-acetyl *N'*-methyl amide. After the generation of the possible dihedral angle sequences within this constraint, unreasonable conformations were rejected prior to energy refinement by testing for ring closure, steric interaction, and *cis*-amide bonds at the site of ring closure. The original high dimensionality problem was reduced to only 152 conformations prior to molecular mechanical energy refinement. Other sequences of filtering criteria have included judicious selection of initial dihedral angle values, ring closure, transannular distances, and rejection of degenerate structures in other studies[20] of cyclic compounds and steric interaction, ring closure, and rejection of degenerate structures coupled with a highly optimized tree searching algorithm in a study of acyclic and cyclic compounds.[15]

The unique chemistry of proteins has been cleverly exploited in a recent study[16] of polypeptide sections within a protein. Conformational space was first restricted to those regions found in previous X-ray crystallographic studies of proteins. A diverse filtering sequence is then used to reduce the conformational possibilities. Of particular interest is the sequential building of the main chain conformations followed by the side chains and the selection of low-energy conformations through the calculation of electrostatic energy and exposed hydrophobic area in the presence of solvent screening. The results from a study of *Streptomyces griseus* trypsin are encouraging. Workers[16] studied two fragments of this protein. After analyzing protein X-ray crystallographic structures, they concluded that 11 pairs of ϕ,ψ angles were sufficient to describe most protein dihedral angles to within 20°. A systematic search of the dihedral angle combinatorial possibilities for these 11 pairs of ϕ,ψ angles was then conducted for subunits of the two fragments. Conformers whose termini could not be joined to the rest of the protein were discarded as well as those having van der Waals violations. The fragments were energy refined with 10 steps of molecular mechanics minimization and degenerate conformations were eliminated. The side chains were then attached to the protein backbone and appropriate conformations were selected with use of van der Waals, electrostatic, and hydrophobic area criteria with the latter two properties calculated with solvent screening. The optimal conformations determined by the filtering process for two polypeptide segments had RMS deviations with respect to the X-ray structure of 0.59 and 1.25 Å.

The concept of conformation filtering has been extended in the program WIZARD[21] to a complete Expert System for conformational analysis. The generation of conformations within this program is CPU efficient as compared to purely numerical methods. In addition, because it is an Expert System, it can easily incorporate knowledge obtained from previous conformational searches to enhance the accuracy and efficiency of searches. Furthermore, although WIZARD utilizes a systematic search method, it is not as rigid as grid searches in that it allows flexibility of bonds and bond angles in addition to dihedral angles. For example, a search[21] of the conformational space accessible to the molecule di-*tert*-butylmethane showed that alterations in the dihedral angles of a standard geometry was insufficient to relieve van der Waals violations. The program relieved these violations by increasing the central carbon bond angle, in agreement with experimental structural data known for di-*tert*-butylmethane.

Although grid search methods normally have limited applicability for molecules with large numbers of dihedral degrees of freedom, "build-up" procedures[22,23] can circumvent that limitation to some degree. The success of this procedure depends on its inherent approximation that the conformation of a molecule is largely dependent on short-range interactions. Thus, a systematic search is made for small segments of a molecule and the lowest energy conformations are determined. The molecule is slowly built up from these segments by combining only the lowest energy conformations during each step, minimizing the combinations, and then again selecting the most favorable conformations for the subsequent step. While this method is useful, especially for those macromolecular systems which are not handled well by other systematic search procedures, it suffers from the disadvantage the optimum conformations for each fragment in the assembled molecule may be rejected in the course of the screening due to the density of low-energy states and the uncertainty of the energies resulting from approximate force fields.

4. Molecular Dynamics Methods

Molecular dynamics is a computational technique in which the time evolution or trajectory of a molecule is described by using the principles of Newtonian mechanics. Thus, the molecule is described as a dynamic structure having atomic coordinates which change with respect to time as influenced by their kinetic energy and by the forces exerted on them by surrounding atoms. In general, the energy of a molecule at a specific time may be described by a potential energy function V. The force field typically contains contributions from bond stretches, bond and dihedral angle bends, and pairwise nonbonded interactions including electrostatic and van der Waals terms. For each atom i with a positional vector x_i and mass m_i, the ac-

(16) Moult, J.; James, M. N. G. *Proteins: Struct., Funct., Genet.* **1986**, *1*, 146.
(17) Ryhänen, T.; Bermejo, F. J.; Santoro, J.; Rico, M. *Comput. Chem.* **1987**, *11*, 13.
(18) White, D. N. J.; Kitson, D. H. *J. Mol. Graphics* **1986**, *4*, 112.
(19) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95.
(20) Wu, Y.; Houk, K., preprint.
(21) Dolata, D. P.; Leach, A. R.; Prout, K. *J. Computer-Aided Molecular Design* **1987**, *1*, 73.
(22) Scheraga, H. A. In *Current Topics in Biochemistry*; Anfinsen, C. B., Schechter, A. N., Eds.; Academic: New York, 1974; p 1.
(23) Gibson, K. D.; Scheraga, H. A. *J. Comput. Chem.* **1987**, *8*, 826.

celeration (a_i) at time t is described by eq 1:

$$\frac{d^2\vec{x}_i}{dt^2} = \vec{a}_i = \frac{\vec{F}_i}{m_i} \quad (1)$$

where

$$\vec{F}_i = -\frac{\partial V}{\partial \vec{x}_i} \quad (2)$$

An excellent introductory discussion of the techniques and applications of molecular dynamics has recently been published.[24]

The time step (Δt) in molecular dynamics is restricted to be smaller than the highest frequency motions, typically bond stretches, found in the system. This necessitates taking small time steps, of the order 1 fs, and usually limits the trajectory length to be of the order 10^1–10^2 ps, for molecules the size of proteins, due to the magnitude of the computation. The necessity of taking small step sizes limits the ability of molecular dynamics to simulate long trajectories given the current computer technology and calculational methodology. The method is not optimally suited either for examining dynamic processes having substantial potential barriers which result in inadequate sampling for statistical evaluations or for sampling large regions of conformational space unless many different conformations are first generated as starting points for different trajectories. Methodology improvements, such as the SHAKE algorithm,[25] allow some increase in trajectory lengths.

Despite the above caveat, the method of molecular dynamics simulation has been frequently used to examine and obtain conformational possibilities for small and moderately sized molecules and macromolecular systems. Although numerous results and insights have been gained, the isolated technique appears to be limited in its ability to generate low-energy conformers due to time restrictions on the trajectory length. Brooks[26] has investigated various strategies for optimal low-energy conformer generation with molecular dynamics. He has concluded that the most efficient use of the method is to generate a moderate number of starting structures (of the order 100) and then minimize the structures. The resulting conformers are then used as starting coordinates for molecular dynamics simulations at 600 K with a trajectory length of approximately one *nanosecond*. Configurations selected from the simulations are then quenched or reminimized to a local minimum. It should be noted that such a procedure was derived from a study of a octapeptide using highly optimized code on a Star Technologies ST-100 computer.[26] Such a molecule could represent the upper size limit for this molecular dynamics procedure and the investigator[26] has stated that the trajectory length of 1 ns is not of an optimal length to fully explore conformational space.

In well-chosen cases, molecular dynamical simulations may provide insights into the dynamics of macromolecular systems despite short trajectory lengths. Recently, workers[27] investigated the active site of a dimeric protein, triose phosphate isomerase (TIM), with molecular dynamics. They performed several simulations: TIM with no substrate, TIM complexed with the natural substrate, dihydroxyacetone phosphate, and TIM complexed with a substrate analogue, dihydroxyacetone sulfate. Although the short trajectory lengths of each simulation, 10.5 ps, were insufficient to allow gross movement of most of the protein, the investigators found extensive motion in the active-site loop. Interesting, the magnitude of the loop movement was greater for the natural substrate than for the substrate analogue. It is proposed that the function of the loop flexibility is to prevent hydrolysis of the substrate during the enzyme-catalyzed isomerization reaction and the workers[27] found evidence that the solvent-accessible surface of the dihydroxyacetone phosphate substrate was significantly reduced by loop conformational changes during the course of the simulation. The dihydroxyacetone sulfate substrate analogue was not as effectively shielded during the simulation.

Experimental spectroscopic data may be used to formulate rational starting structures for molecular dynamics simulations in those cases where structural information is especially sparse. Generally, the spectroscopic data which are used are interproton distances derived from nuclear magnetic resonance spectroscopy NOE (nuclear Overhauser enhancement) experiments. Since one is then presumably starting the simulation in a appropriate part of conformational space, there is a reasonable change of obtaining useful conformations despite the computational limitations of the trajectory lengths. The complex resulting from the binding of ristocetin pseudoaglycon to Ac_2-Lys-D-Ala-D-Ala has been studied[28] with constrained molecular mechanical minimization and molecular dynamics. The glycopeptide–peptide complex was first interactively constructed by computer graphics and then subjected to minimization where the force field used the additional harmonic distance-constraint term shown in eq 3, where

$$K(r - r_0)^2 \quad (3)$$

r and r_0 are the calculationally determined and experimental distances, respectively, and K is a pseudobond force constant. The structure resulting from constrained minimization was in good agreement with the intra- and intermolecular distances obtained from NOE experiments, however, the conformer was not a local minimum on the energy surface and the structure obtained from subsequent unconstrained minimization was not consistent with all of the experimental data. A molecular dynamics simulation[28] showed this and another structure to be interconverting and the average distances calculated from the trajectory were in better agreement with the NOEs. The simulation not only provided additional low-energy conformations of the complex but also illustrates the need to consider more than one conformation when evaluating time-averaged data from mobile systems.

A modification of molecular dynamics which allows enhanced sampling of low-energy conformations involves incorporating NOE distance constraints in the simulation to force the trajectory to sample conformations compatible with spectroscopic data. A particularly dramatic example has recently been reported.[29] The workers looked at the feasibility of generating the structure of a 46-residue protein, crambin, with molecular dynamics and NOE distance constraints. The NOE data was incorporated into the molecular dynamics force field by the addition of the following term to the potential:

(24) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: New York, 1987.
(25) (a) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327. (b) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.
(26) Brooks, B. R. *Proceedings from Supercomputer Research in Chemistry and Chemical Engineering*; Truhler, D., Jensen, K. Eds.; American Chemical Society: Washington, DC, 1987.
(27) Brown, F. K.; Kollman, P. A. *J. Mol. Biol.* **1987**, *198*, 533.
(28) Fesik, S. W.; O'Donnell, T. J.; Gampe, R. T., Jr.; Olejniczak, E. T. *J. Am. Chem. Soc.* **1986**, *108*, 3165.
(29) Brünger, A. T.; Clore, G. M.; Gronenborn, A. M.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3801.

$$\text{for } r_{ij} > r^\circ_{ij},\ E_{NOE} = c_1(r_{ij} - r^\circ_{ij})^2 \qquad (4)$$
$$\text{for } r_{ij} < r^\circ_{ij},\ E_{NOE} = c_2(r_{ij} - r^\circ_{ij})^2$$

where r_{ij} and r°_{ij} are the calculationally determined and experimental distances, respectively, and c_1 and c_2 are constants which incorporate the temperature and the error between the calculated and experimental distances. Regardless of whether the workers started the simulations in a totally extented structure of crambin or one in which some secondary α-helical structure was preformed, they found averaged structures after molecular dynamics which showed a RMS deviation of less than 2 Å, for the main chain atoms, with respect to the X-ray determined structure. It is impressive that the tertiary structure of crambin essentially formed after 9 ps of simulation. The molecular dynamics/distance constraint methodology has also been applied to other proteins and peptides[30-34] and nucleic acids.[35]

Not all molecular dynamics simulations which incorporate distance constraints use the functional form of eq 4. Another method is to use a one-sided potential as shown in eq 5, where r_{ij} and r°_{ij} are the calculational and experimental distances, respectively, and K_{NOE} is the distance constraint force constant. Kessler et al.[34] have discussed the applicability of the two potentials. They point out that molecular dynamic potentials contain a repulsive component in the van der Waals potential, therefore inclusion of repulsion in the distance constraint potential, as shown in equation 6, may be superfluous. Furthermore, if conformational averaging occurs on the nuclear magnetic resonance spectroscopy time scale during the NOE experiment, or if the experimental data is not very accurate, the use of a repulsive harmonic component is not justified.

$$\text{for } r_{ij} > r^\circ_{ij},\ E_{NOE} = \frac{1}{2}K_{NOE}(r_{ij} - r^\circ_{ij})^2 \qquad (5)$$
$$\text{for } r_{ij} \le r^\circ_{ij},\ E_{NOE} = 0$$

5. Monte Carlo Methods

In Monte Carlo calculations, the dynamic behavior of a molecule is simulated by random changes which are made to the system, such as dihedral angle rotation or atom displacement. The energy of the trial configuration of atoms is calculated and the configuration is accepted if the energy has decreased from the previous configuration. If the energy is instead higher after the random change is made, an algorithm is used to determine whether the new configuration should be accepted. Commonly, the Metropolis algorithm[36] is implemented and the new configurations are accepted with a probability given by the Boltzmann distribution (eq 6), where k is the Boltzmann constant and T the absolute temperature. A random number in the interval (0, 1) is generated and compared with the Boltzmann factor, if the random number is smaller, the configuration is accepted, otherwise the previous configuration is again subjected to a random alteration.

$$\exp\left(-\frac{\Delta E_i}{kT}\right) \qquad (6)$$

Although the Monte Carlo method has been used most frequently for the simulation of liquids, some investigations have been carried out to study conformations in small molecules and constrained motion in proteins. The techniques of Monte Carlo and molecular dynamics simulations were combined in a study[37] of an activated structural process in a protein, that of the tyrosine-35 ring rotation in bovine pancreatic trypsin inhibitor (BPTI). Since the tyrosine ring rotation occurs on a time scale of microseconds, molecular dynamics alone cannot be used since it would not provide adequate statistics in a reasonable amount of computational time. The investigators[37] therefore generated conformations representative of the transition-state geometry by manually orienting the tyrosine ring. Monte Carlo simulations were then used to produce relaxed configurations. The transition-state geometry was maintained by accepting only those conformations where the dihedral angle associated with ring rotation was within 2.5° of the initial geometry. The resulting configurations served as starting points for molecular dynamic trajectories through the transition-state geometry. Using this approach, the researchers were able to determine the interactions which promote the tyrosine-35 ring rotation. Other researchers[38] have similarly used the Monte Carlo method to generate configurations of the Ala dipeptide (AcAlaNHMe) near the known C_7,C_5,α_r, and P_{II} local minima in order to calculate intramolecular thermodynamical properties.

The relative dearth of Monte Carlo simulations of macromolecules is due to the inefficiency of the technique for systems which contain many covalent bonds. Workers[39] have compared the effectiveness of two Monte Carlo methods versus molecular dynamics in producing configurations of a protein. Both the Metropolis[36] Monte Carlo method and a Smart Monte Carlo procedure,[40] where the trial configuration is calculated on the basis of the forces which were acting on the previous configuration, were found to be less efficient in generating conformations of BPTI than molecular dynamics. For equivalent RMS displacements in BPTI, the Monte Carlo methods[39] were found to require more than a factor of 10 increase in computational time as compared to that of molecular dynamics. The researchers attribute the ineffectiveness of the Monte Carlo methods to a poor trial configuration acceptance probability due to the energy surface anisotropy in systems having many covalent bonds. This may be understood in the following manner. If a trial displacement of an atom is in the same direction as a covalent bond, the energy surface will be steep and there will be a high probability that the displacement will not be accepted. The surface will be less steep in directions perpendicular to the bond since smaller forces are acting on the atom. Such a displacement will have a higher probability of being accepted. The problem arises because many more dis-

(30) Fesik, S. W.; Bolis, G.; Sham, H. L.; Olejniczak, E. T. *Biochemistry* **1987**, *26*, 1851.
(31) Clore, G. M.; Gronenborn, A. M.; Brünger, A. T.; Karplus, M. *J. Mol. Biol.* **1985**, *186*, 435.
(32) Kaptein, R.; Zuiderweg, E. R. P.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. F. *J. Mol. Biol.* **1985**, *182*, 179.
(33) Nilges, M.; Gronenborn, A. M.; Brünger, A. T; Clore, G. M. *Protein. Eng.* **1988**, *2*, 27.
(34) Kessler, H.; Griesinger, C.; Lautz, J.; Müller, A.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393.
(35) Nilsson, L.; Clore, G. M.; Gronenborn, A. M.; Brünger, A. T.; Karplus, M. *J. Mol. Biol.* **1986**, *188*, 455.
(36) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(37) McCammon, J. A.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 3585.
(38) Ravishanker, G.; Mezei, M.; Beveridge, D. L. *J. Comput. Chem.* **1986**, *7*, 345.
(39) Northrup, S. H.; McCammon, J. A. *Biopolymers* **1980**, *19*, 1001.
(40) Rossky, P. J.; Doll, J. D.; Friedman, H. L. *J. Chem. Phys.* **1978**, *69*, 4628.

placements sample steep surfaces than not in molecules such as proteins; hence the overall acceptance probability will be small. It has been suggested[39] that the development of a Monte Carlo algorithm which compensated for the surface anisotropy would allow the method to be more competitive with molecular dynamical simulations of covalently bonded systems.

Such an improved Monte Carlo method has been recently implemented.[41] In the modified method, the bond lengths and bond angles are held rigid and trial displacements are made for the dihedral angles. The issue of energy surface anisotropy effecting acceptance ratios is treated by taking only small step sizes in regions where the surface is steep and large steps in more shallow regions. The new method was applied to BPTI and was found to be 50–500 times more efficient in generating conformational change as compared to the standard Monte Carlo method utilizing isotropic step sizes. The investigators[41] have also compared the conformational change efficiency of the modified Monte Carlo method to molecular dynamics. They conclude the method is 5–50 times more efficient than molecular dynamics. The method may be quite useful if one is interested in generating conformations of molecules; however, the approximation in keeping the bond lengths and angles rigid may be physically inappropriate.

6. Distance Geometry and Ellipsoid Algorithm Methods

Distance geometry procedures offer another computational aid in conformational searching, particularly in those cases where experimental data is available for intramolecular and/or intermolecular distances. The method[42] utilizes a matrix of all pairwise atomic distances in a molecule to generate a series of Cartesian coordinates for which the matrix is suitable. Standard geometries may be used for some distances while others are obtained through experimental data (i.e., NOEs) and random number generation in the upper and lower bound range provided by known distances. Although distance geometry has been used in small molecule conformational searching,[30,43] it appears to be especially well-suited for macromolecular solution conformations.[44]

The power of distance geometry has been illustrated by a study[45] of the aqueous solution conformations of α-amylase inhibitor (Hoe-467A), a 74-residue polypeptide. Distance geometry was used to generate the backbone conformation of the polypeptide and the conformations of the side chains in the proposed active site. The distance matrix was constructed, in part, from 401 NOE distance constraints. Four solutions to the matrix were found, each with an average nuclear magnetic resonance distance constraint violation of approximately 0.025 Å. The four conformations are similar except for the terminal residues and one five-residue loop. The RMS fit for the main chain heavy atoms when the above-mentioned residues are excluded is 1.6 Å. An independent X-ray crystallographic study[46] was consistent with the proposed backbone conformations.

Distance geometry is similar to many systematic searches in that the generated conformations deviate from relaxed structures due to imperfect internal coordinates and nonbonded interactions. Geometry optimizations do relax the structures, but the process may also decrease the fit to the distance constraints. One partial solution to this problem has been to assume the minimized conformation is in an appropriate general region of conformational space and only local adjustments are needed. Thus, the conformations are often subjected to short molecular dynamics simulations in order to sample the local potential surface.

A technique which is conceptually related to distance geometry is that of the ellipsoid algorithm.[47,48] An ellipsoid large enough to encompass conformational space is generated in n-dimensional space, where n is the number of dihedral angles in the molecule. In each iteration, a violated inequality constraint, such as a NOE distance, is selected. If there are no violated inequality constraints, an objective function, such as van der Waals violations, is chosen. The function is evaluated and the smallest new ellipsoid containing the negative half of the gradient of the previous ellipsoid is generated. The iterations continue until there are no longer violations in the structure and the objective function has converged. The center of the ellipsoid is a vector whose components are the dihedral angles of a conformation. The attractiveness of the algorithm is, in part, due to the fact that the dimensionality of the problem is defined by the number of dihedral angles whereas in distance geometry the dimensionality is determined by the Cartesian coordinates of the atoms.

In a test case[48] of the Ala dipeptide of 36 starting structures which defined a grid covering the two-dimensional space, 33 converged to the global minimum for the molecule. The structures found by the techniques of distance geometry and the ellipsoid algorithm are similar. For example, the ellipsoid algorithm was used[48] to generate conformers of an 11-residue segment [Arg-17–Met-27] of the polypeptide glucagon. Eighty-six distance constraints from NOE experiments were used during the optimizations. A comparison of the RMS deviation for distance geometry and the ellipsoid algorithm of the peptide backbone atoms was 1.4 Å.

The ellipsoid algorithm has also been used[49] to explore the conformational surfaces of a cyclic molecule, 18-crown-6. When followed by molecular dynamics to explore local regions of conformational space, the workers were able to obtain numerous low-energy minima for the molecule, and on several occasions they found the conformer identified by X-ray crystallography. In addition, the researchers were able to exploit the complexation properties of cyclic ethers during the conformational search. They found that, by using a cation as a template during the molecular dynamics simulations, they were also able to enhance the isolation of the apparent conformational minimum for the complexed crown ether.

7. Conclusions

As we mentioned in the introductory section, we have concerned ourselves in this paper with a discussion of the generation of molecular conformations, and not their subsequent refinement and energy evaluation. There is, however, one aspect of energy evaluation which we would like to briefly mention. It is currently possible to evaluate

(41) Noguti, T.; Gō, N. *Biopolymers* **1985**, *24*, 527.
(42) Crippen, G. M. *Distance Geometry and Conformational Calculations*; Research Studies Press: New York, 1981.
(43) Weiner, P. K.; Profeta, S., Jr.; Wipff, G.; Havel, T.; Kuntz, I. D.; Langridge, R.; Kollman, P. A. *Tetrahedron* **1983**, *39*, 1113.
(44) Kuntz, I. D. *Protein Eng.* **1987**, *1*, 147.
(45) Kline, A. D.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1986**, *189*, 377.
(46) Pflugrath, J.; Wiegand, G.; Huber, R.; Vértesy, L. *J. Mol. Biol.* **1986**, *189*, 383.
(47) Shor, N. Z. *Cybernetics* **1977**, *12*, 94.
(48) Billeter, M.; Havel, T. F.; Wüthrich, K. *J. Comput. Chem.* **1987**, *8*, 132.
(49) Billeter, M.; Howard, A. E.; Kuntz, I. D.; Kollman, P. A., *J. Am. Chem. Soc.*, in press.

and order the energies of conformations of molecules such as hydrocarbons and a variety of programs can execute this function. Furthermore, once the conformations have been ordered in terms of energies, one has confidence that the calculated lowest energy conformations are those which contribute to the ground-state properties of the molecule. This is due to the minimal charge and polarizability associated with hydrocarbons. The same cannot be said for molecules with more highly charged or polar atoms such as are found in peptides or proteins. As an example, suppose one generates conformations for a polypeptide and evaluates their energies in vacuo. It is highly probable that the most favored conformations will be those that maximize intramolecular electrostatic interactions and hydrogen bonding. If the energies of those same conformations were now evaluated in the presence of a polar solvent, a significant number of the previously favorable interactions would be replaced by more energetically favorable solvent–polypeptide interactions. The differences in the in vacuo and solvent calculations may be so notable that one would select totally different groups to represent low-energy conformations under the two conditions.

Thus, one would like to calculate the energies of conformations for some molecules in the presence of explicit solvent molecules. Such calculations, however, are often prohibitively expensive in terms of CPU time, even for molecular mechanical force fields. Approximations may be used to simulate the presence of solvent, but often these are also fairly crude and the ordering of the energies of conformations may be inaccurate. Although some recent work has been done in this area, for example in hydration shell calculations,[50] this problem obviously needs to be addressed further since it is foolish to spend a large amount of effort generating conformations if the subsequent energy evaluation is not representative of what is found in nature.

A great deal of work has been done and is currently being pursued in the field of conformational searching, of which we have presented only a sampling. For small and moderately sized systems, systematic searches based upon expert systems perhaps offer the most efficient and convenient approach. Larger molecules may be more amenable to distance geometry methods or molecular dynamics coupled with distance constraints derived from experiment. While these methods may not always exhaustively sample low-energy conformations in molecules, especially in the case of macromolecular systems, they do provide a means of generating conformations which can provide insight into understanding the physical properties of molecules.

Acknowledgment. This work was supported by the National Institutes of Health under Grant No. GM-29072. We gratefully acknowledge M. Billeter, B. R. Brooks, D. A. Case, A. T. Hagler, K. N. Houk, G. Marshall, H. A. Scheraga, W. C. Still, and W. F. van Gunsteren for providing preprints.

(50) (a) Kang, Y. K.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4105. (b) Kang, Y. K.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4109. (c) Kang, Y. K.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4118. (d) Gibson, K. D.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4121.

Articles

Platinum Complexes with Binding Affinity for the Estrogen Receptor

Norbert Knebel and Erwin von Angerer*

Institut für Pharmazie, Universität Regensburg, Universitätsstrasse 31, D-8400 Regensburg, Federal Republic of Germany.
Received December 2, 1987

A number of (1,2-diaminoethane)dichloroplatinum(II) complexes, linked to dihydroxy-2-phenylindole by spacer groups of varying lengths, were synthesized and studied for their binding affinities for the calf uterine estrogen receptor. Best binding conditions were provided by the *n*-hexyl and the *p*-xylene group as spacer with RBA values of 6.5 (**16c**) and 4.4 (**17c**), respectively (17β-estradiol: RBA = 100). These values are only slightly lower than those of the corresponding diaminoethane ligands.

Endocrine therapy of hormone-dependent mammary tumors has proven to be a valuable alternative to chemotherapy with cytostatic agents because of the low toxicity associated with drugs like antiestrogens or aromatase inhibitors.[1-3] However, this treatment is limited by the fact that approximately 40% of the patients with estrogen receptor positive tumors do not respond to endocrine manipulations.[4] The reason for this lack of response is still unclear. The presence of these receptors in tumors that do not respond initiated our search for new compounds that bind to the receptor but exert their antitumor effect by a different mode of action. Substances with receptor affinity and that carry a cytotoxic group were thought to be good candidates for this purpose. The diaminodichloroplatinum(II) group was chosen as the cytostatic function because the parent compound *cis*-platinum is a potent antineoplastic agent against some tumors, especially against testicular cancer, but with low activity against breast cancer. Receptor affinity of platinum complexes might make it possible to overcome the resistance of mammary tumors to *cis*-platinum. Our rationale is based solely on the presence of estrogen receptors in the malignant cells but not on their function as transmitters of hormonal signals. The receptor should only be used to direct the cytotoxic agent toward the target cell.

(1) Pearson, O. H.; Manni, A.; Arafah, B. M. *Cancer Res. (Suppl.)* **1982**, *42*, 3424s.
(2) Jordan, V. C.; Fritz, N. F.; Tormey, D. C. *Cancer Res.* **1987**, *47*, 624 and references therein.
(3) Santen, R. J.; Brodie, A. M. H. In *Clinics in Oncology*; Furr, B. J. A., Ed.; Saunders: London, 1982; Vol. 1, p 77.
(4) Maass, H.; Jonat, W.; Stolzenbach, G.; Trams, G. *Cancer* **1980**, *46*, 2783.

0022-2623/88/1831-1675$01.50/0