Theory of Complex Molecular Interactions: Computer Graphics, Distance Geometry, Molecular Mechanics, and Quantum Mechanics

PETER KOLLMAN

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Received July 5, 1984 (Revised Manuscript Received December 8, 1984)

Complex molecules such as proteins and nucleic acids have fascinated chemists because of their essential role in the functioning of biological systems. The chemist would like to know the structure and energies of the conformations of these molecules. To a first approximation, the situation looks hopeless. Although modern quantum mechanics can give powerful insights into the structure and potential energy surface of small molecules,[1,2] macromolecular systems may be composed of 500 atoms or more with uncountable local energy minima.

The methods of single-crystal X-ray crystallography and high-resolution NMR become essential in studies of such complex molecules. X-ray crystallography gives an average picture of low-energy structures of the macromolecule in the crystal[3] and NMR is becoming increasingly powerful for studying complex molecules in solution.[4] But the theoretician's lot still looks pretty hopeless. What can he do but make worse the structure given to him by the crystallographer? In fact, the macromolecular crystallographer has made use of the principles of chemical structure and molecular mechanics in refining his structure, since his data is usually not good enough to lead to an atomic resolution structure ab initio. Thus, the experimentalist must work very hard to get an experimental structure and any clues he gets from the theoretician are useful.

One reason that a theoretician can contribute to understanding complex molecular systems is the recent development of a number of powerful approaches for studying large molecules. A goal of this Account is to show that theoretical methods can give useful insights into the structure and energy of macromolecular systems.

Methods

The methods used in theoretical studies of complex molecules can be divided into two classes: the first is "model building" in which molecular structure is represented by experimental data as input and this structure is manipulated with use of stereochemical rules. Model building includes physical molecular models, computer graphics, and distance geometry methods. The second class relies on energy functions. This class includes ab initio and semiempirical quantum and molecular mechanical calculations, including energy minimization and Monte Carlo and molecular dynamics simulations.

Model building has long been important in helping chemists understand stereochemical properties of molecules.[5,6] Recently, computer graphics methods employing real-time "clipping", color, stereo, and molecular surface representations[7] has allowed one to understand molecular structure and interactions in complex molecules at a comparable level of detail as a small molecule chemist manipulating his substituted cyclohexane ring with manual models. A second "model building" method is distance geometry,[8] which relies less heavily on X-ray crystallographic results for its input and has a large numerical calculation component. In principle, one can generate a complete (albeit low resolution) three-dimensional structure of a molecule with distance geometry with use of information from stereochemical rules and NMR NOE measurements. Distance geometry has the nice feature of being applicable to all kinds of molecular structure, from ribosomes[9] to 18-crown-6,[10] and the additional feature of "telling" one exactly how much one has learned from a given set of experimental data and what further experiments need to be done.

Ab initio quantum mechanical calculations have been applied to a wide range of chemical phenomena and the recent developments of gradient optimization and correlation methods incorporated into the programs made available to the scientific community by Pople and co-workers (GAUSSIAN 80[11] and 82[12]) has been essential to some of the calculations described below. Semiempirical quantum mechanical calculations have also been extensively improved in the last decade,[13] but

Peter Kollman received his B.A. from Grinnell College in 1966 and his Ph.D. from Princeton University in 1970. After a postdoctoral year as a NATO fellow in Cambridge, he joined the faculty at the University of California, San Francisco. Most important in his life, after his family and academic activities, are basketball, touch football, softball, and tennis, although at age 40—he writes—some of these activities are in danger of becoming "academic" too.

(1) T. Clark and P. Schleyer, *J. Am. Chem. Soc.*, **101**, 7747 (1979).
(2) M. Yoshimine, A. McLean, B. Liu, D. DeFrees, and J. Binkley, *J. Am. Chem. Soc.*, **105**, 6185 (1983).
(3) J. Kraut, *Ann. Rev. Biochem.*, **46**, 331 (1977).
(4) K. Wüthrich, G. Wider, G. Wagner, and W. Braun, *J. Mol. Biol.*, **155**, 311 (1982).
(5) L. Pauling, "The Nature of the Chemical Bond", 3rd ed., Cornell University Press, Ithaca, NY, 1960.
(6) J. Watson and F. H. Crick, *Nature (London)*, **171**, 964 (1953).
(7) R. Langridge, T. Ferrin, I. D. Kuntz, and M. L. Connolly, *Science*, **221**, 661 (1981).
(8) G. Crippen, "Distance Geometry and Conformational Calculations", Research Studies Press, Wiley, New York, 1981.
(9) I. D. Kuntz and G. Crippen, *Biophys. J.*, **32**, 677 (1980).
(10) P. Weiner, S. Profeta, G. Wipff, T. Havel, I. D. Kuntz, R. Langridge, and P. A. Kollman, *Tetrahedron*, **39**, 1113 (1983).
(11) Gaussian 80, Quantum Chemistry Program Exchange Program No. 406, J. S. Binkley, R. A. Whiteside, R. Krishnan, R. Seeger, D. DeFrees, H. Schlegel, S. Topiol, L. Kahn, and J. A. Pople, Bloomington, IN, 1980.
(12) Gaussian 82, J. A. Pople and Carnegie-Mellon University, unpublished, 1982.
(13) M. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99**, 4989 (1977); M. Zerner, G. Loew, R. Kirschner, and U. Müller-Westerhoff, *J. Am. Chem. Soc.*, **102**, 589 (1980).

0001-4842/85/0118-0105$01.50/0

such methods are not yet really useful for studying intermolecular interactions.

Molecular mechanical methods have been very powerful in the study of organic molecules in nonpolar solvents.[14] Such methods use simple analytical functions to represent bond stretching, bond bending, and torsional and nonbonded (dispersion attraction, exchange repulsion, and electrostatic interaction) energies of atoms and molecules. These analytical functions and their first and second derivatives can be rapidly and efficiently evaluated, and thus this method can be applied to much larger molecules than can quantum mechanical methods. Programs such as REFINE,[15] ECEPP,[16] AMBER,[17] and CHARMM[18] use topological structure methods to construct polymers from monomer units and this has facilitated recent molecular mechanics studies.

One of the major difficulties in molecular mechanics is how to represent the electrostatic properties of atoms in molecules. U. Chandra Singh has generalized an approach in the literature for representing molecular charge distributions using quantum mechanically calculated electrostatic potentials and incorporated this approach, plus other modifications, into GAUSSIAN 80.[19] S. Weiner and I have used such charge distributions in determining new molecular mechanical parameters for proteins and nucleic acids.[20] T. Lybrand and I have developed methods to incorporate nonadditivity effects in such molecular mechanical simulations,[21] following closely previous approaches.[22-24]

Molecular mechanics approaches include energy minimization, which leads to a OK structure, Monte Carlo,[25,26] which can only be used in an efficient way to study intermolecular degrees of freedom, and molecular dynamics,[27,28] which is the most general as well as time-consuming method to evaluate structures and energies of complex molecules. Even with such a powerful approach as molecular dynamics, there are the difficulties of searching the local energy minima in an efficient way, in correctly representing the effect of water on the energetics of the system, and in efficiently evaluating the free energy as opposed to the internal energy of complex molecular systems. Thus, ab initio theoretical approaches to studies of such systems are currently rather hopeless. In such situations one can only extract *some* information on the systems of interest, but, if the problems are well-chosen, this information can be very valuable.

Applications

Below, we present applications of the four methods noted in the title in studying molecular interactions, taking examples mainly from work in our own laboratory.

Computer Graphics Model Building Per Se

Quantum mechanical energy component analysis of small molecule interactions and thermodynamic studies of nonpolar groups in aqueous solutions give convincing support to the proposition that the major factors governing molecular association are three: van der Waals (which includes both steric-exchange repulsion and dispersion attraction), electrostatic (which includes H bonding), and hydrophobic. Using the Connolly surface algorithm,[29] we developed a method to represent the electrostatic potential at molecular surfaces to allow a qualitative representation of all three of these factors in molecular association. Such a representation allows one to manually dock two molecules in such a fashion to *avoid* steric repulsion, *maximize* dispersion attraction by maximizing the molecular surface areas just touching, optimize electrostatic complimentarity by matching opposite color surfaces, and maximize hydrophobic contacts by optimizing nonpolar (electrostatically neutral) contacts. This representation[30] has enabled better understanding of the electrostatic complementarity of the trypsin–trypsin inhibitor interaction, the preference for 3,5-diiodo over 3,5-dialkyl groups in maximizing thyroxine analogue binding to prealbumin, the lack of affinity of prealbumin for DNA, the general preference for GC-rich regions in restriction endonuclease sites, and the preference of netropsin for AT-rich DNA. Recently, we have presented a representation of the electrostatic potential gradient (using arrows in addition to surface dots) in order to make very clear how superoxide dismutase (SOD) achieves its enormous catalytic efficiency despite the fact that Cu^{2+} is only on such a small fraction of the SOD surface.[31] Koppenol had earlier noted that SOD reacts much more efficiently than predicted by collision theory and had suggested an electrostatic basis for this.[32] In related work, Hansch and co-workers[33] have used a different surface representation (polar or hydrophobic) to help understand quantitative structure–activity relationships earlier derived from experimental protein–ligand binding data.

The fact that much of the above is retrospective rather than predictive illustrates the difficulty in fully "understanding" complex molecular systems, since a complete "theory" should be predictive as well. In this context, the study by Blaney et al.[34] stands out. In that

(14) N. L. Allinger, *J. Am. Chem. Soc.*, **99**, 8127 (1977).
(15) J. Hermans and D. Ferro, *Biopolymers*, **10**, 1121 (1971).
(16) ECEPP, Quantum Chemistry Program Exchange No. 454 (Bloomington, IN), and F. Momany, R. McGuire, A. Burgess, and H. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).
(17) P. Weiner and P. Kollman, *J. Comput. Chem.*, **2**, 287 (1981).
(18) B. Brooks, R. Brucioleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187 (1983).
(19) U. C. Singh and P. Kollman, Quantum Chemistry Program Exchange, Program No. 446, Bloomington, IN, 1982.
(20) S. Weiner, P. Kollman, D. Case, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, *J. Am. Chem. Soc.*, **106**, 785 (1984).
(21) T. Lybrand and P. Kollman, submitted to *Nature (London)*; T. Lybrand and P. Kollman, submitted to *J. Chem. Phys.*
(22) A. Warshel and M. Levitt, *J. Mol. Biol.*, **103**, 227 (1976).
(23) F. H. Stillinger and C. W. David, *J. Chem. Phys.*, **69**, 1473 (1978).
(24) P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn, *Nature (London)*, **282**, 459 (1979).
(25) W. Jorgensen, J. Chandresekhar, J. Madera, R. Impey, and M. Klein, *J. Chem. Phys.*, **79**, 926 (1983), and references therein.
(26) G. Alagona, C. Ghio, and P. Kollman, *J. Am. Chem. Soc.*, in press.
(27) J. A. McCammon and M. Karplus, *Acc. Chem. Res.*, **16**, 187 (1983).
(28) U. C. Singh, S. Weiner, and P. Kollman, *Proc. Nat. Acad. Sci. U.S.A.*, in press.
(29) M. Connolly, Program No. 427, Quantum Chemistry Program Exchange, Bloomington, IN, 1982.
(30) P. Weiner, R. Langridge, J. Blaney, R. Schaefer, and P. Kollmann, *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 3954 (1982).
(31) E. Getzoff, J. Tainer, P. Weiner, P. Kollman, J. Richardson, and D. Richardson, *Nature (London)*, **306**, 287 (1983).
(32) W. Koppenol in "Oxygen and Oxy-Radicals in Chemistry and Biology", E. Powers and M. Rogers, Ed., Academic Press, New York, 1981, p 671.
(33) R. Nelson Smith, C. Hansch, K. Kim, B. Omiya, G. Fukumura, C. Selassie, P. Jow, J. Blaney, and R. Langridge, *Arch. Biochem. Biophys.*, **215**, 319 (1982).

study, the molecular surface representation was crucial in enabling one to predict the relative binding affinities of thyroxine analogues to prealbumin, predictions subsequently confirmed by equilibrium dialysis experiments.

Computer Graphics Model Building and Molecular Mechanics

Computer graphics model building and molecular mechanics are complementary techniques in studying ligand–macromolecule interactions. The model building is used to dock the ligand into the macromolecule site; such structures are then completely energy refined by molecular mechanics.

What is one trying to achieve in these studies? Are molecular mechanics methods capable of representing the strength and directionality of intermolecular interactions involving small, polar organic molecules? We have presented evidence[20,35] that they are. This is true because the electrostatic energy component is the dominant term in such interactions and, provided one accurately represents the charge distribution of the molecule in the molecular mechanical atomic charges, then molecular mechanics should be able to mimic more accurate calculations quite well. This "electrostatic" dominance also underlies the success of simple water models in well reproducing the properties of liquid water.[25] This is not to say that the more subtle and difficult polarization and charge-transfer effects are not important. However, it is likely that such inclusion is not *essential* in a qualitatively correct description of many interesting molecular interactions.

The more difficult problem is one of adequately representing the solvent effect on the intermolecular interactions. Without inclusion of water and extensive simulation, it is impossible to calculate ΔG for molecular associations. We must resign ourselves to a more limited goal: the calculation of $\Delta\Delta G$, the relative free energy of association of different ligands to a given macromolecule, or the relative free energy of association of a given ligand to different (but structurally very similar) macromolecules.

Our approach has been to choose examples from the experimental literature where we had both experimental binding affinities (reaction rates) and an X-ray structure of the macromolecule and see if we could even reproduce the experimental $\Delta\Delta G$ values. Furthermore, we first focused on the relative binding energies of *stereoisomers*, where solvation effects should be least critical. Such studies enabled one to understand and reproduce experimental side-chain stereospecificity and ionic preferences in thyroxine analogues binding to prealbumin. In that study, the molecular mechanics refinement led to some qualitatively different energy minima in some cases than in the model-built one.

In these molecular mechanical simulations we were able to reproduce the preference of prealbumin for L- vs. D-thyroxine, with side chain R = $CH_2C(H)(NH_3^+)$-COO^-. Furthermore, by using a strictly empirical solvation model along with the molecular mechanics energies, we could correctly order the relative binding affinities of des-NH_3^+-thyroxine (R = $CH_2CH_2COO^-$) > L-thyroxine > D-thyroxine > des-COO^--thyroxine (R = $CH_2CH_2NH_3^+$).[36] In fact, the important role of solvation in correctly reproducing the relative binding free energies of the amino acids L- and D-thyroxine and the carboxylic acid des-NH_3^+-thyroxine led to the *idea* of trying a nonionized side chain (R = CH_2CH_2OH) in prealbumin binding, and this molecule was found to have a rather high affinity, greater than that of thyroxine itself.[37]

Also insightful were the results of such studies on models for the Michaelis complex and tetrahedral "intermediates" in α-chymotrypsin interactions with *N*-acetyl-D- and -L-tryptophan amide.[38] Experimentally, both stereoisomers bind noncovalently to α-chymotrypsin with similar affinities, but the L isomer is hydrolyzed at least 10^7 faster than the D isomer, implying that almost all the stereospecificity occurs at or near the transition state for catalysis. The fact that our calculations and that by De Tar[39] on tetrahedral models (with Ser-195 O_γ attached to the hydrolyzable amide C of the substrate, with O_γ proton transferred to His-57) find L more stable than D (in our calculations by about 10 kcal/mol) and we find the Michaelis complexes of L and D almost equoenergetic strongly supports the prevailing dogma that amide hydrolysis goes through a transition state such as our model-built one. Even more exciting than the correspondence between experimental and theoretical energies were the mechanistic insights afforded by our study. The single largest energy component difference between L and D in the tetrahedral model was that the protonated His-57 N–H···N hydrogen bond to the former amide of the substrate, now an amine nitrogen in the L isomer, was significantly stronger than the corresponding protonated His-57 N–H···O_γ (Ser-195) H bond in the D isomer. Since the next logical step in the mechanism for amide catalysis is *transfer* of this His-57 proton to a group on the tetrahedral (formerly carbonyl) carbon, the L isomer structure will lead to transfer to the amine, leading to the acyl enzyme and subsequent hydrolysis. In the D isomer, transfer of the proton to Ser O_γ bond would cause the C–O_γ to break, regenerating the starting materials. Such mechanistic insights were not put into the computer graphics model building but emerged after molecular mechanics refinement. Figures 1 and 2 show stereo views of these L and D tetrahedral intermediates. We have also made a prediction on the NMR properties of the *N*-(trifluoroacetyl)-D-tryptophan amide–α-chymotrypsin complex.[38]

In the area of drug–DNA intercalations, we have made extensive use of model building and subsequent molecular mechanical energy refinement in elucidating the sequence specificity of the interactions of 4-nitroquinoline *N*-oxide[40] with dinucleoside phosphates and ethidium cation[41] and actinomycin D[42] with di- and

(34) J. Blaney, E. C. Jorgensen, M. L. Connolly, T. Ferrin, R. Langridge, S. Oatley, J. Burridge, and C. Blake, *J. Med. Chem.*, **25**, 785 (1982).
(35) P. Kollman, *J. Am. Chem. Soc.*, **100**, 2974 (1978); U. C. Singh and P. Kollman, *J. Chem. Phys.*, **80**, 353 (1984).
(36) J. Blaney, P. Weiner, A. Dearing, P. Kollman, E. C. Jorgensen, S. Oatley, J. Burridge, and C. Blake, *J. Am. Chem. Soc.*, **104**, 6424 (1982).
(37) J. Blaney, Ph.D. Thesis, University of California, San Francisco, 1983.
(38) G. Wipff, A. Dearing, P. Weiner, J. Blaney, and P. Kollman, *J. Am. Chem. Soc.*, **105**, 997 (1983).
(39) D. Detar, *J. Am. Chem. Soc.*, **103**, 107 (1981).
(40) T. L. Lybrand, A. Dearing, and P. Kollman, *Nucleic Acids Res.*, **9**, 6995 (1981).
(41) M. E. Nuss, F. Marsh, and P. Kollman, *J. Am. Chem. Soc.*, **101**, 825 (1979); T. Lybrand and P. Kollman, *Biopolymers*, in press.

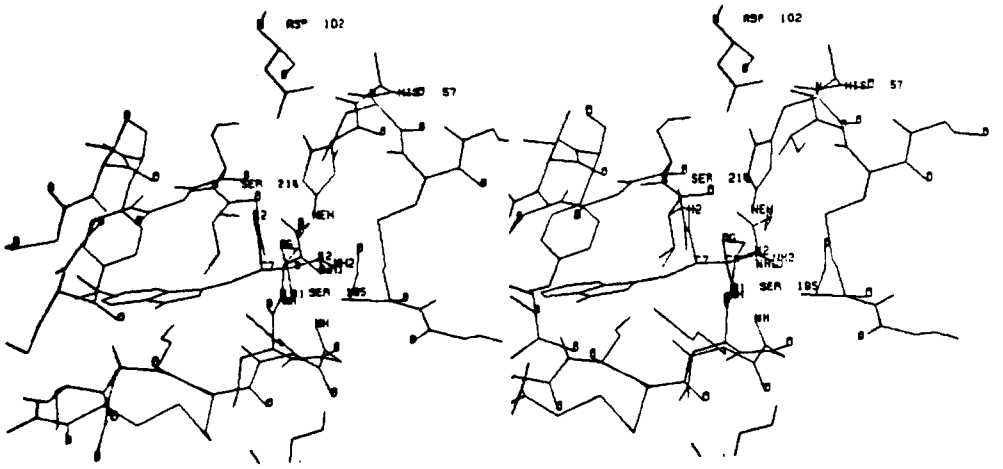


Figure 1. Stereo view of model of tetrahedral intermediate for *N*-acetyl-L-tryptophanamide catalysis by α-chymotrypsin.

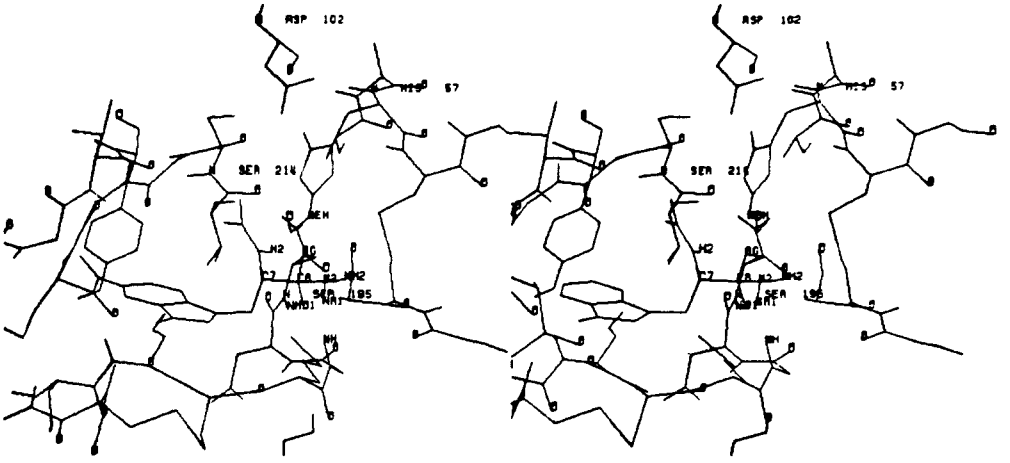


Figure 2. Same as Figure 1 for the D isomer.

hexanucleoside phosphates. Here we are focusing on simulations where we compare binding affinities and properties of double-helical DNA with the same ligand but different DNA sequences. Again, the assumption of *similar* DNA geometry for different sequences seems reasonable for both the nonintercalated DNA structure and the intercalated drug–DNA complexes. These simulations have been generally successful in reproducing base-sequence preferences. For example, we are able to reproduce the actinomycin D (ACTD) intercalation preference for dG(3′–5′)dC sites over the other three possibilities and the reasons for this preference come from specific H-bonding interactions of ACTD peptide groups with the guanine proton acceptor N3 and 2-NH_2 proton donor, in line with earlier suggestions by Sobell and our calculated model for the $d(ATGCAT)_2$–ACTD complex[42] is consistent with that deduced from 2D NMR studies.

Studies in progress on model-building and energy-refining *nonintercalative* noncovalent complexes of DNA with netropsin,[43] *covalent* adducts to double-helical DNA with stereoisomers of benzpyrenediol epoxides[44] and mitomycin[45] and DNA-damaged structures such as thymine dimer and 6–4 photodimer structures show promise of being equally illuminating.[46,47]

Thus, to summarize our simulations combining computer graphics model building with molecular mechanics, we note that they are mainly *retrospective*, in that even reproducing experimental relative energies is a minor feat, but important *ideas* in ligand design (the nonpolar side chain in thyroxine) interpretations of enzyme mechanisms (the D/L stereospecificity in α-chymotrypsin) or structural predictions (the location of the *N*-trifluoroacetyl group in the *N*-(trifluoroacetyl)-D-tryptophanamide–α chymotrypsin complex) can emerge from such studies.

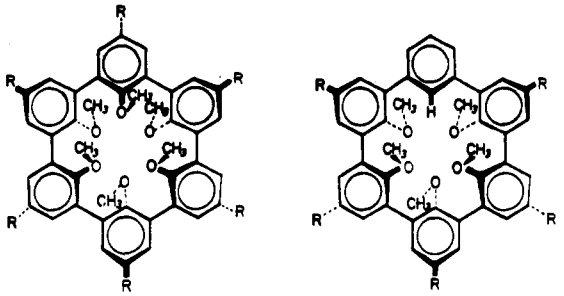
Distance Geometry and Molecular Mechanics

Above, we have alluded to the fact that complex molecules have many local energy minima, and it is difficult to describe all these minima, let alone evaluate their energies. Distance geometry techniques allow one to construct many qualitatively different conformations of noncyclic or cyclic molecules and thus to create a set of well-defined conformations for further evaluation. In conjunction with NMR data, a combined use of distance geometry, NMR NOE measurements, and molecular mechanics calculations is being used to help elucidate the solution conformation of a molecule as complicated as bleomycin.[48] For smaller systems such as 18-crown-6

(42) T. Lybrand, S. Brown, R. Shafer, and P. Kollman, manuscript to be submitted.
(43) J. E. Caldwell and P. Kollman, *Biopolymers*, in press.
(44) S. N. Rao, T. Lybrand, and P. Kollman, manuscript to be submitted.
(45) U. C. Singh, S. N. Rao, and P. A. Kollman, submitted to *J. Am. Chem. Soc.*
(46) S. N. Rao, J. W. Keepers, and P. Kollman, *Nucleic Acid Res.*, in press.
(47) S. N. Rao and P. Kollman, unpublished calculations on 6–4 photodimers.

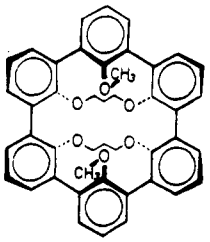
and anisole spherands, the number of energetically reasonable conformations is small enough that one can begin to attack in a systematic way the conformational profiles of these molecules. The goals of these studies has been to find low-energy conformations that can explain the nonzero dipole moment of the molecule (18-crown-6)[49] and find new conformations of the molecule that have a higher ion binding affinity than those previously observed (anisole spherand). In the case of 18-crown-6, we succeeded in finding a low-energy, noncentrosymmetric conformation, which likely contributes to its solution properties (significant dipole moment) and which has not been found in X-ray crystallographic structural studies of the molecule.

The recent studies by Cram et al.[50] and Trueblood et al.[50] on spherand 1 have also proven to be very fertile ground for theoretical study. In simulating spherands **1–3**, we[51] did not use any X-ray coordinates as input, constructing the spherands and their alkali cation (Li^+, Na^+, K^+) complexes by using a template-driven distance-geometry program and then energy refining them with molecular mechanics. Using the same parameters that reproduced the preference of 18-crown-6 (**4**) for K^+ over Na^+, we found the opposite in spherand 1, consistent with the observation that 1 has a very high affinity for Li^+ and Na^+, but none for K^+. By using the

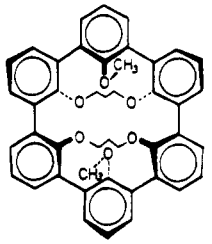


1

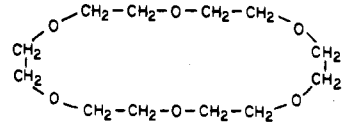
2



3a



3b



4

second derivative of the energy function, we were able to calculate the normal modes of 1, its complexes, and (gas phase) free energies for complexation. Using such an approach, we calculated a $\Delta\Delta G$ of Li^+ association

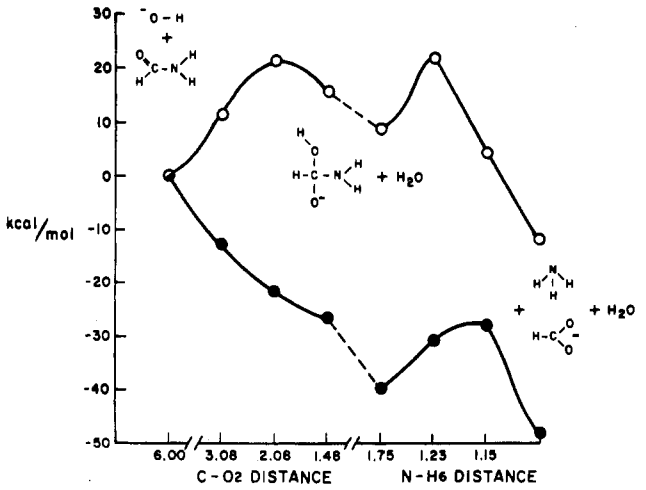


Figure 3. Reaction profile for formamide hydrolysis by OH^- in the gas phase (●) and in solution (○).

upon replacing one OCH_3 with H 1 vs. 2 of 13.6 kcal/mol, in very good agreement with the experimental $\Delta\Delta G$ of >12.6 kcal/mol. Even more exciting was our finding that a second isomer of **3**,[49] which had not been found in the reaction mixture, was calculated to be significantly lower in energy and would be expected to have the highest known Li^+ affinity. Our calculations suggest a very high barrier between **3a** and **3b** and, thus, help rationalize why **3a** does not easily isomerize to the more stable **3b**.

We emphasize again the power of constructing the coordinates of such an isomer a priori using distance geometry and molecular mechanics refinement (no X-ray coordinates were used as input in this study) but also stress that these spherands are particularly favorable cases for theoretical study, since the molecules are so intrinsically constrained that the local minimum problem is not so severe.

Quantum Mechanics and Molecular Mechanics

Molecular mechanics is an inherently ad hoc model to represent the structure and energy of molecular systems. It is quite capable of giving an excellent description of the energies and structures of normal covalently bonded molecules and complexes where the predominant attractive force is electrostatic. However, it is much more difficult to use it successfully for reliably calculating the energetics of processes during which bonds are made or broken. On the other hand, the reactants may be surrounded by many "normal" molecules interacting with normal noncovalent forces, making the overall system much too large to completely study by quantum mechanical methods. In this context, we have recently[52] begun a combined application of quantum and molecular mechanics to study a chemical reaction in the gas phase, in solution, and, eventually, in an enzyme active site. The reaction we have chosen is amide hydrolysis in basic solution, $OH^- + H_2NCHO \rightarrow HCOO^- + NH_3$. We have used ab initio gradient methods to study this reaction, which, we suggest, is likely to proceed in two steps: OH^- attack and water-catalyzed proton transfer/C–N bond breakage. In the gas phase, the first step of the reaction is completely "downhill" energetically and the second in-

(48) N. J. Oppenheimer, I. D. Kuntz, and P. Kollman, unpublished calculations on bleomycin with use of distance geometry and molecular mechanics.

(49) G. Wipff, P. Weiner, and P. Kollman, *J. Am. Chem. Soc.*, 104, 3249 (1982).

(50) D. J. Cram, T. Kaneda, R. C. Helgeson, and G. M. Lein, *J. Am. Chem. Soc.*, 101, 6752 (1979); K. N. Trueblood, C. B. Knobler, E. Maverick, R. C. Helgeson, S. B. Brown, and D. J. Cram, *ibid.*, 103, 5594 (1981).

(51) P. Kollman, G. Wipff, and U. C. Singh, *J. Am. Chem. Soc.*, in press.

(52) S. J. Weiner, U. C. Singh, and P. A. Kollman, *J. Am. Chem. Soc.*, in press.

volves a modest barrier. When we place the reactants in a solvent bath of 216 TIPS3P H_2O molecules[33] and carry out molecular mechanical refinement to estimate the solvation energy, the first step "acquires" a barrier[52] of about 20 kcal/mol, qualitatively consistent with experimental barriers[53] for amide hydrolysis in basic solution, whereas the second step involves a similar barrier. Figure 3 illustrates the difference between the reaction energetics in the gas phase and in solution. The next step in this project has been to carry out the same reaction in the active site of the serine protease enzymes,[54] where use of computer graphics has been helpful. In such a study, we include protein *and* water *and* substrate and evaluate the energies in a combined quantum/molecular mechanical fashion. Such an approach has shown that one can simulate reactions that may occur in vacuo, in solution, and in an enzyme in a quantitatively reasonable way and thus have begun to more fully understand enzyme catalysis. The elegant approach of Warshel[55] is an alternative way to reach the same goal, in which one calibrates the model to reproduce free energy of the reaction in solution and then carries out the simulation in the enzyme using the same parameters. Our approach has the advantage in that fewer empirical adjustments need be made at the quantum mechanical level but the disadvantage that (currently) our model of handling solvation (use of internal rather than free energies) is more primitive. But future developments in carrying out molecular dynamics[56] (rather than energy minimization) on complex systems (enzyme + substrate + water) and in evaluating free energies[57] and potentials of mean force[58] for such systems show promise of being able to remove this disadvantage.

Computer Graphics + Quantum Mechanics + Molecular Mechanics

In the above study of the formamide reaction pathway, computer graphics/model building did not play a *primary* role in the construction of the geometry of the system. But in studies of enzymatic reactions, as in the molecular mechanical calculations on protein–ligand interactions discussed above, model building does play an essential and primary role.

We have already studied one enzyme-catalyzed reaction in the gas phase and in the enzyme using a combination of computer graphics, quantum mechanics, and molecular mechanics to study the dihydroxyacetone phosphate (DHAP) → glyceraldehyde 3-phosphate (GAP) isomerization catalyzed by a carboxylate base (Glu-165) of triose phosphate isomerase (TIM).[59] In the absence of the enzyme, COO^- abstraction of a proton from DHAP is calculated to be unfavorable by ~20 kcal/mol. Although we did not simulate this reaction in aqueous solution, it is obvious that the proton-transfer reaction might be even more unfavorable there, since one is exchanging a COO^- anion for a more delocalized enediolate and the latter would be expected to be less effectively solvated in aqueous solution. In the enzyme, the active site His-95 and the Lys-13–Glu-97 ion pair stabilize the enediolate sufficiently to make it of approximately equal energy to the reactants and thereby to lower the barrier to proton transfer to the order of 10 kcal/mol. The calculations thus enable one to understand how the TIM reaction becomes diffusion controlled, in that the enzyme lowers the barriers for the chemical steps of the reaction. In these calculations, we used computer graphics along with X-ray structures based on difference density maps of Alber and Petsko[60] to dock DHAP in the TIM active site and then carried out molecular mechanical refinement of this structure. These refined geometries were used to incorporate the partial charges of the enzyme groups into the quantum mechanical calculation. With both the molecular mechanical and quantum mechanical models, the enzymatic stabilization of the intermediate endiolate relative to DHAP was ~20–25 kcal/mol, just the appropriate amount to facilitate rapid proton transfer.

One of the most exciting developments in recent years in studies of enzymes and their catalytic mechanisms has been the availability of recombinant DNA techniques. These enable one to make single amino acid changes in protein and, thus, to critically assess which residues are essential for efficient enzyme catalysis. In view of the fact that Davenport and Petsko[61] had begun to make a mutant TIM, in which His-95 was replaced by Gln-95, we carried out comparative molecular mechanics simulations on the modified enzyme.

These lead to an interesting new structure, which has Gln-95–Glu-165 hydrogen bonding. Although the lack of inclusion of water and the fact that only a small fragment of the protein was included in the calculation makes this result tentative, it suggested an alternative rationalization for why the Gln mutant may be less active (other than the difference in pK_a of Gln and His). Thus, it should be looked at as a new idea/hypothesis that has emerged from the molecular mechanical simulation that needs to be further tested with more refined theory or, preferably, by experiments.

Together with the powerful genetic engineering methods, we see the combination of computer graphics, quantum mechanics and molecular mechanical approaches as very useful tools in helping to suggest new interesting proteins for study as well as allowing further elucidation of the nature of enzyme catalysis.

Molecular Mechanics and Computer Graphics in Analyses of DNA

An example of such a study has been the examination of base-pair opening and backbone flexibility in DNA.[62] By carrying out *constrained* molecular mechanics calculations, we have been able to demonstrate coupled low-energy torsional motions that retain base stacking and hydrogen bonding. These calculations were carried out by forcing one torsional angle to change (e.g., ω' g → t) and allowing complete energy refinement of all other geometrical parameters. Constrained molecular mechanics calculation forcing base-pair opening have suggested a mechanism for the relatively rapid A-T

(53) J. Guthrie, *J. Am. Chem. Soc.*, **96**, 3608 (1974).
(54) S. Weiner, Ph. D. Thesis, University of California, San Francisco, 1984.
(55) A. Warshel, *J. Phys. Chem.*, **83**, 1640 (1979).
(56) M. Karplus and J. McCammon, *Ann. Rev. Biochem.*, **52**, 263 (1983).
(57) K. Wilson, *J. Chem. Phys.*, in press.
(58) S. Northrup, M. Pear, J. Morgan, J. McCammon, and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 4035 (1982).
(59) G. Alagona, P. Desmeules, C. Ghio, and P. Kollman, *J. Am. Chem. Soc.*, **106**, 3623 (1984).
(60) We thank T. Alber and G. Petsko for sending us this information prior to publication.
(61) R. Davenport and G. Petsko, unpublished results.
(62) J. Keepers, P. Kollman, P. Weiner, and T. James, Proc. *Natl. Acad. Sci. U.S.A.*, **79**, 5537 (1982).

base-pair proton exchange in double-helical B-DNA. Such coupled motions and base-pair opening structures can be illustrated by using the movie option in the program CHEM.[63]

A second application is the use of graphics to illustrate the low-frequency normal modes and the molecular dynamics[64] of complex molecules. Such applications have nicely illustrated base-pair motions and how drugs affect these motions in fragments of double-helical DNA and its complexes.

Finally, we note some calculations recently completed[65] on the DNA sequences $d(ATATAT)_2$, $d(TATATA)_2$, $d(CGCGCG)_2$, $d(GCGCGC)_2$, and $dA_6 \cdot dT_6$ where molecular mechanics has played a primary role and graphics a secondary analysis role. In this set of studies, the use of specific *dihedral* restraints to study conformational variants of the above structures has found, for right-handed B-DNA like structures of $d(ATATAT)_2$ and $d(TATATA)_2$, both uniform C2′ endo and mixed-sugar pucker models (adenine-C3′ endo; thymine-C2′ endo) are of comparable calculated energy and both are more stable than structures with thymine-C3′ endo and adenine-C2′ endo. These results are consistent with the crystal structure of $(pApTpApT)_2$,[66] which has a mixed A-C3′ endo T-C2′ endo geometry and with NMR data on poly[d(A-T)]; poly[d(A-T)],[67] which shows two ^{31}P resonances. On the other hand, $d(CGCGCG)_2$ and $d(GCGCGC)_2$ prefer uniform C2′ endo geometries, consistent with the low salt solution structure[68] of poly[d(G-C)] (only one ^{31}P resonance). Finally, $dA_6 \cdot dT_6$ prefers a structure with the thymine sugars C3′ endo and adenines C2′ endo to the model with all C2′ endo sugars and to the model with thymine sugars C2′ endo and adenine sugars C3′ endo. Peticolas[69] has shown that in poly(dA)·poly(dT), half the sugars are C2′ endo and half C3′ endo but not which sugars are which.

Energy component analysis on these structures suggests that there are three critical factors that lead to the above-calculated preferences: first, intrinsic deoxyribo sugars are more stable in C2′ endo than C3′ endo geometries by ~0.6 kcal/mol. Secondly, C3′ endo sugars in a B-DNA chain shorten the phosphate–phosphate distances and, thus, increase phosphate–phosphate repulsions. Finally, there are favorable base–phosphate interactions that occur only when thymine is the base and the sugar on the 5′-end of the sugar to which the thymine is attached is C3′ endo. The above physical effects can apparently explain all the available experimental data, including the salt dependence of the sugar geometries[69] and the *seeming* contradiction that in poly[dA-T)]·poly[d(A-T)], the low-energy structure has adenine-C3′ endo; thymine C2′ endo, whereas in poly(dA)·poly(dT), the low-energy structure has adenine-C2′ endo; thymine-C3′ endo. Both of these results are simply explicable when one realizes that in these structures, the sugar on the 5′-end of the thymine sugar is C3′ endo and thus contains the favorable base–backbone interaction noted above. We hope that 2D NMR techniques will enable our prediction for the sugar puckering (*which* sugars are C3′ endo) of poly(dA)·poly(dT) to be tested.

Summary and Conclusions

In our view, one can make a convincing argument that theoretical simulations are useful in aiding understanding of complex systems and that model building and energy calculations used together are clearly synergistic in achieving this understanding. We see the theoretician working with the crystallographer and NMR spectroscopist playing an increasingly important role in aiding protein modification and drug design.

I am very pleased to acknowledge the essential input that my colleagues Dave Case (UC, Davis), Tack Kuntz, and Bob Langridge and co-workers Paul Weiner, Andrew Dearing, Georges Wipff, Sal Profeta, Joe Keepers, Jeff Blaney, Chandra Singh, Kitty Ghio, Guiliano Alagona, Terry Lybrand, Jim Caldwell, Shashi Rao, and Scott Weiner have had on the research reported here. These studies have been supported by CA-25644 and GM-29072 from the National Institutes of Health and Grant CHE-80-26560 from the National Science Foundation.

(63) CHEM, program written at University of California, San Francisco, computer graphics laboratory by A. Dearing.

(64) NMODES, display programs written by P. Weiner and U. C. Singh, and DYNAMICS, program written by M. Levitt and modified by U. Singh.

(65) S. N. Rao and P. Kollman, *J. Am. Chem. Soc.*, in press.

(66) M. Viswamitra, O. Kennard, P. G. Jones, G. M. Sheldrick, S. A. Saliburg, S. Falvello, and Z. Shakked, *Nature (London)*, **273**, 687 (1978).

(67) D. Patel, S. A. Kozlowski, J. Suggs, and S. D. Cox, *Proc. Natl. Acad. Sci. U.S.A.*, **78**, 4063 (1981).

(68) J. S. Cohen, J. S. Wooten, and C. L. Chatterjee, *Biochemistry*, **20**, 3049 (1981).

(69) G. A. Thomas and W. Peticolas, *J. Am. Chem. Soc.*, **105**, 993 (1983).