mechanisms of redox-induced metal-metal bond breaking.

Because the connection between the electronic structure and reactivity of molecular clusters of increasing size and small metal particles bears on problems relevant to catalysis and materials science, a continued interest in cluster research seems likely. Both experimental and theoretical chemists will be challenged to develop reliable methods for elucidating the chemistry and electronic structures of large clusters.

This work is based on research sponsored by the National Science Foundation. The efforts of the many student and faculty collaborators in the references cited are gratefully acknowledged.

Computer Modeling of the Interactions of Complex Molecules

PETER A. KOLLMAN* and KENNETH M. MERZ, JR.†

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

Received October 9, 1989 (Revised Manuscript Received April 3, 1990)

The usefulness of computer-based approaches to simulate the properties of molecules is now accepted in most areas of chemistry. The accuracy with which one can approximate the solution to the Schrödinger equation ab initio for small molecules is constantly increasing,[1] and semiempirical molecular orbital methods have made useful contributions to medium-sized molecules.[2] Computer simulation methods[3] have led to exciting new insights into the properties of condensed phases of molecules. The major focus of this Account is on some of the advances in applications of computer approaches to biological macromolecules.

These computer-based applications can vary from the use of data bases and computer graphics to numerical calculations in the areas of molecular quantum mechanics, molecular mechanics, and statistical mechanics. The focus of our research is to understand the nature of molecular interactions in complex molecules; both numerical calculations and computer graphics methods have been used in our pursuit of this understanding. We wrote an Account some years ago,[4] and it will give a useful historical perspective to this Account.

There are three fundamental problems in the computer simulations of complex biological molecules. The first and most difficult to overcome is the "global minimum" problem. Globular proteins and nucleic acids have hundreds or even thousands of atoms and correspondingly many degrees of conformational freedom. Even if one could exhaustively search conformational space, one still needs to correctly evaluate and rank the relative free energies of all the conformations. This is currently impossible even for systems of ~100 atoms.

Secondly, the study of complex molecules with energy calculations requires molecular mechanics methods. This stands in contrast to studies of small molecules in the gas phase where Schrödinger's equation can be solved to an adequate degree of accuracy to represent the structure, reactivity, and energies of the molecules. Can we say anything useful with such simple approaches to represent the structure and complexation energies of systems?

Thirdly, the quantity of relevance in understanding the energetics of complex systems is the free energy; what can we say about free energies of such systems?

In the last five years there have been useful advances in conformational searching methods to overcome the global minimum problem. There is a growing acceptance that the question about the usefulness of molecular mechanics methods should be answered definitively yes, and there have been tremendous advances in our ability to calculate free energy differences to make direct and meaningful contact with experiment. This Account will focus on our contributions in these area of research, but will conclude that some other applications to demonstrate that various computational methods can be synergistic in giving useful qualitative insight into complex systems. With the latter we reinforce a major theme of our previous Account.

Conformational Searching and the Global Minimum Problem

Development of methods for conformational searching on "small" constrained systems such as crowns[5] and cyclic peptides or antibody combining loops[6,7] provides

Peter Kollman was born in Iowa City, IA, on July 24, 1944, and attended Grinnell College and Princeton University, where he received his Ph.D. in 1970, working under the supervision of Leland Allen. After a postdoctoral year in Cambridge, England, with David Buckingham, he joined the faculty of the School of Pharmacy, Department of Pharmaceutical Chemistry, University of California, San Francisco, where he is now Professor of Pharmaceutical Chemistry. He is the father of two children, Sarah and Eli, and the husband of Mercedes Acosta. This year he achieved a personal best by making 26 free throws in a row.

Kenneth Merz was born in Niagara Falls, NY, in 1959. He received his B.S. in 1981 from Washington College in Chestertown, MD, and his Ph.D. in 1985 from the University of Texas at Austin under the supervision of M. J. S. Dewar. Postdoctoral work followed, with R. Hoffmann at Cornell University and with P. Kollman at the University of California, San Francisco. In 1989 he joined the faculty at the Pennsylvania State University (University Park Campus), where he is now an Assistant Professor of both Chemistry and Molecular and Cell Biology. In 1990 he received an Office of Naval Research Young Investigator Award. His research interests lie in method development and the study of the structure and function of zinc metalloenzymes and antibiotic ionophores.

†Current address: Department of Chemistry, Pennsylvania State University, State College, PA 16802.

(1) Schaefer, H. F. *Science* **186**, *231*, 1100.
(2) Dewar, M.; Zoebisch, E.; Healy, E.; Stewart, J. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(3) McCammon, J. A.; Harvey, S. C. *Molecular Dynamics of Proteins and Nucleic Acids*; Cambrdige Univ. Press: New York, 1987.
(4) Kollman, P. *Acc. Chem. Res.* **1985**, *18*, 105.
(5) Billeter, M.; Howard, A.; Kuntz, I.; Kollman, P. *J. Am. Chem. Soc.* **1988**, *110*, 8385.
(6) Bruccoleri, R.; Karplus, M. *Biopolymers* **1987**, *26*, 137.
(7) Fine, R.; Wang, H.; Shenkin, P.; Yarmuth, N.; Levinthal, C. *Proteins* **1986**, *1*, 342.

0001-4842/90/0123-0246$02.50/0

training for the ultimate target, prediction of the tertiary structure of proteins from amino acid sequence.[8] We thus began a study of the conformations of 18-crown-6 using a combination of an ellipse/distance geometry method to generate structures and molecular mechanics/dynamics to refine them.[5] Although we did not succeed in exhaustively searching conformational space with this approach, it is likely that we found the global minimum structure in the absence and presence of K^+.

Other approaches to conformational searching include systematic searching with closure and van der Waals screening,[9] systematic searching using a data base of allowed conformations to efficiently prune the conformational tree,[6] Monte Carlo searching in Cartesian space,[10] and Monte Carlo searching in dihedral space.[11] These approaches appear capable of finding all the low-energy structures for cycloheptadecane, provided a sufficient investment of computer time is made. An imaginative way to generate ring conformations using a Fourier series has been presented,[12] as have approaches using simulated annealing.[13,14] At this point it is not clear which is the most efficient method for most fully searching conformational space. However, it is clear that systematic approaches will run into a combinatoric "block", and molecular mechanics and dynamics sample usefully locally, but are not especially efficient for global searching by themselves. Nonetheless, dynamics methods can be used along with template ligands such as K^+, to direct conformational searches in particular "directions".[5]

Searching conformational space is only the first half of the problem when one deals with anything more polar than hydrocarbons and considers the molecule in a condensed phase. One must then reliably rank the relative free energies of the conformations in solution, which is a substantial feat if one is dealing with many conformations. No one has addressed this issue adequately yet.

It is significant that the use of restrained molecular dynamics to refine structures forced to be consistent with NMR NOE data[15,16] with *explicit inclusion of water* reproduces the properties of the structure significantly better than the corresponding simulation in vacuo.[17] Furthermore, Chiche et al. reinforce previous conclusions that empirical solvation free energies using surface area calculations are a very valuable guide to distinguish correct from incorrect protein structures. Daggett has developed a new sigmoidal dielectric function,[18] which is a modification of one that Hingerty[19] and Lavery[20] have employed, and demonstrated that molecular dynamics with such a function on an α-helical peptide is significantly better at reproducing the properties of simulations with explicit water molecules than the standard distance-dependent dielectric model. Thus, progress is being made at evaluating and improving various models for molecular simulations.

Molecular Mechanics Energy Function: Can It Be Useful and Reliable?

There is growing evidence that the answer to the above question is yes. We have made a considerable effort to validate molecular mechanical models on a wide variety of test cases, including conformational energies, barriers, and intermolecular interactions.[21,22] The force field equation used by us is as follows:

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] + \sum_{\text{H bonds}}\left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right] \quad (1)$$

We have discussed the development of this force field in considerable detail[21,22] and emphasize again the importance of the electrostatic term which is critical in the reproduction of polar and ionic intermolecular interactions. Our main contribution to this to generalize and proselytize the use of electrostatic potential derived charges[23] by demonstrating its superiority in leading to relative strength and directionality of nucleic acid base associations.[24] The fact that our approach leads to very good relative and absolute free energies for neutral molecule solvation[25] and association[26] is further validation, as is the calculation of reasonable free energies for nucleic acid base association in water.[27,28] The bond length, angle, and dihedral terms have been used in the development of an OPLS force field for proteins.[29]

It is clear, on the basis of quantum mechanical calculations, why a molecular mechanical model using only electrostatics, exchange repulsion, and dispersion is capable of correctly calculating intermolecular interaction strengths and directionalities.[4] This is due partially to a fortuitous cancellation of angular dependency in the charge-transfer and exchange-repulsion energies. This also requires a correct reproduction of

(8) Gibson, K.; Scheraga, H. A. *J. Comput. Chem.* **1987**, *8*, 826.
(9) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. the conformational parameter in drug design: the active analog approach. In *Computer Assisted Drug Design*; Olsen, E. C., Christofferson, R., Eds.; ACS Symposium Series 112; American Chemical Society: Washington, DC, 1979; pp 205–226.
(10) Saunders, M. *J. Am. Chem. Soc.* **1987**, *109*, 3150.
(11) Lipton, M.; Still, W. C. *J. Comput. Chem.* **1988**, *9*, 343.
(12) Gerber, P.; Gubernator, K.; Muller, K. *Helv. Chim. Acta* **1988**, *71*, 1429.
(13) Wilson, S. R.; Cui, W.; Moskowitz, J.; Schmidt, K. *Tetrahedron Lett.* **1988**, 4373.
(14) Wilson, S.; Cui, W. Applications of Simulated Annealing to Peptides. *Biopolymers*, in press.
(15) Kaptein, R.; Zuiderweg, E.; Scheek, R.; Boelens, R.; Van Gunsteren, W. *J. Mol. Biol.* **1985**, *182*, 179.
(16) Brunger, A. T.; Clore, G. M.; Gronenborn, A. M.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3801–3805.
(17) Chiche, L.; Gaboriaud, C.; Hertz, A.; Mormon, J. P.; Castro, B.; Kollman, P. *Proteins* **1989**, *6*, 405.
(18) Daggett, V.; Kollman, P.; Kuntz, I. D. Molecular Dynamics Simulations on Small Peptides: Dependence on Deielctric Model and pH. *Biopolymers*, submitted.
(19) Hingerty, B.; Richie, R.; Ferrel, T.; Turner, J. *Biopolymers* **1985**, *24*, 427.
(20) Ramstein, J.; Lavery, R. *Proc. Natl. Acad. Sci., U.S.A.* **1988**, *85*, 7231.
(21) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
(22) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230–252.
(23) Singh, U. C.; Kollman, P. *J. Comput. Chem.* **1984**, *5*, 129.
(24) Rao, S.; Kollman, P. *Biopolymers* **1986**, *25*, 267.
(25) Bash, P.; Singh, U. C.; Langridge, R.; Kollman, P. *Science* **1987**, *236*, 564.
(26) Grootenhuis, P. D. J.; Kollman, P. *J. Am. Chem. Soc.* **1989**, *111*, 4046.
(27) Cieplak, P.; Kollman, P. *J. Am. Chem. Soc.* **1988**, *110*, 3734.
(28) Dang, L.; Kollman, P. *J. Am. Chem. Soc.* **1990**, *112*, 503.
(29) Tirado-Rives, J.; Jorgensen, W. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

the electrostatic multipole moments of the molecule, which is achievable with electrostatic potential derived charges. The Weiner et al. force fields used a combination of charges from ab initio calculations with different basis sets.[21,22] We are currently working[30] on the consistent use of 6-31G* electrostatic potential derived charges, because that basis set gives good multipole moments for molecules. This basis set was not used before because of computer limitations in 1984. These 6-31G* derived charges tend to overestimate dipole moments consistently by ~20%. The latter fact allows these charges to be "balanced" with a TIP3P[31] or SPC[32] water model for aqueous simulations using the effective two-body approximation.

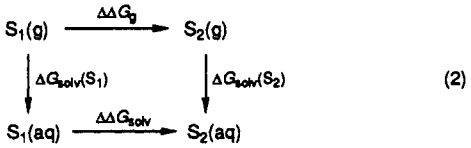
We are also developing approaches to include polarizability effects in the force fields in a general way, which is likely to be critical for an accurate representation of charged systems.[5,33,34] Other developments are the implementation of electrostatic potential based charges at the MNDO and AM1 levels.[35] It appears that 6-31G* quality charges might be derivable for very large molecules at modest computational cost, by employing a scaling factor.

Thus, progress is being made to refine further the already useful molecular mechanical representations of complex molecular systems including proteins and nucleic acids. However, quantum mechanical representations are the only physically correct description of atoms and molecules. Use of quantum mechanics is essential if one wants to model covalent processes such as chemical reactions or enzyme catalysis. Approaches that use quantum mechanics for the few atoms where bonding is changing and molecular mechanics energies for the rest may be applied to such systems. Warshel has developed a series of models coupling semiempirical and empirical quantum mechanical approaches with molecular mechanics dynamics.[36,37] Singh and Kollman[38] have coupled ab initio calculations with molecular mechanics/dynamics, and Bash et al.[39] have coupled MNDO/AM1 semiempirical methods with molecular mechanics/dynamics to study reactive processes. These combined methods promise to broaden the range of interesting biophysical phenomena amenable to computer simulation considerably; they have already led to new insights into enzyme catalysis.[40] In summary, much progress has been made in the last four years in developing and understanding approaches to represent as accurately as possible the energy of complex molecules using molecular mechanics or a combination of molecular mechanics and quantum mechanical methods. The future is bright for continued refinement of such representations, leading to new insight into biophysical phenomena.

Applications of Free Energy Perturbation Methods To Study Molecular Solvation, Ligand Binding, and Enzyme Catalysis and Sequence-Dependent Stabilities

In the last few years, the most exciting developments in applications of computational chemistry to complex molecules have been (1) X-ray and NMR refinement using molecular dynamics[15,16] and (2) free energy calculations. Our efforts have been focused in the latter area and were stimulated by earlier studies by Postma et al.,[41] Warshel,[42] McCammon,[43] and Jorgensen.[44] In 1985–1986, Chandra Singh implemented free energy perturbation and integration methods into the AMBER set of programs, and in early 1986, he, Paul Bash, and Frank Brown refined and tested the methodology and applied it to a wide variety of solvation calculations as well as to inhibitor binding to the protein thermolysin. The methodology used is summarized in ref 45, and an article by Beveridge nicely summarizes the different free energy calculations.[46] Below we attempt to summarize and give a sense of the excitment of applications of free energy methods to complex molecular systems.

Molecular Solvation. Bash et al.[25] and Singh et al.[45] applied the free energy perturbation approach to amino acid side chain solvation in the following context:



(2)

Experimentally, one determines the free energy of solvation of molecules S_1 and S_2, $\Delta G_{solv}(S_1)$, and $\Delta G_{solv}(S_2)$, and the thermodynamic cycle requires that

$$\Delta\Delta G_{solv}(aq) - \Delta\Delta G_g = \Delta G_{solv}(S_2) - \Delta G_{solv}(S_1) \quad (3)$$

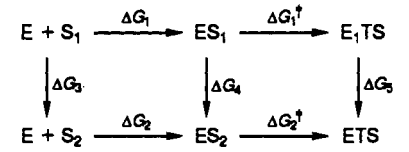
where $\Delta\Delta G_{solv}(aq)$ and $\Delta\Delta G_g$ are determined by "mutating" S_1 into S_2 in a periodic box of water molecules and in the gas phase, using either Monte Carlo or molecular dynamics to evaluate the ensemble average of the free energy differences. In Monte Carlo applications $\Delta\Delta G_g$ is neglected, and in our molecular dynamics applications this term is also neglected, assuming internal contributions to the free energy to be comparable in the gas phase and solution. The reasonableness of this is supported by the excellent agreement between theory and experiment in a wide variety of calculations[25] as well as detailed calculations by van Gunsteren.[47]

(30) Spellmeyer, D.; Merz, K.; Pearlman, D.; Kollman, P., work in progress.
(31) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926–935.
(32) Berendsen, H. J. C.; Postma, J.; van Gunsteren, W.; Hermans, J. *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordecht, 1981; p 331.
(33) Cieplak, P.; Lybrand, T.; Kollman, P. *J. Chem. Phys.* **1988**, *88*, 8017. Howard, A.; Singh, U. C.; Billeter, M.; Kollman, P. *J. Am. Chem. Soc.* **1988**, *110*, 6984. Lybrand, T.; Kollman, P. *J. Chem. Phys.* **1985**, *83*, 2923.
(34) Caldwell, J.; Dang, L. X.; Kollman, P. A. The Implementation of Non-Additive Potentials Using Molecular Dynamics: Development of a Water Potential and Water–Ion Cluster Interactions. *J. Am. Chem. Soc.*, submitted.
(35) Besler, B.; Merz, K.; Kollman, P. *J. Comput. Chem.* **1990**, *11*, 431.
(36) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(37) Warshel, A.; Weiss, R. *J. Am. Chem. Soc.* **1980**, *102*, 6218.
(38) Singh, U. C.; Kollman, P. *J. Comput. Chem.* **1986**, *7*, 718.
(39) Bash, P.; Field, M.; Karplus, M. *J. Am. Chem. Soc.* **1987**, *109*, 8092.
(40) Singh, U. C.; Pattabiraman, N.; Langridge, R.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6402–6406.
(41) Postma, J.; Berendsen, H.; Haak, J. *Faraday Symp. Chem. Soc. No.* **1981**, *17*, 55.
(42) Warshel, A. *J. Phys. Chem.* **1982**, *86*, 2218.
(43) Tembe, B.; McCammon, A. *Comput. Chem.* **1984**, *8*, 281.
(44) Jorgensen, W.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.
(45) Singh, U. C.; Brown, F. K.; Bash, P. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1987**, *109*, 1607.
(46) Beveridge, D.; DiCapua, F. Free Energy via Molecular Simulation: Applications to Chemical and Biomolecular Systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.
(47) W. van Gunsteren, unpublished.

The exciting result is that one is able to calculate the difference in free energy of solvation in reasonable agreement with experiment for a wide variety of polar and hydrophobic functional groups[48] as well as ground and excited states.[49] Furthermore, the standard deviation of such free energies is of the order of kilocalories/mole or less. If the simulations had been run significantly longer, as is now computationally feasible, the standard deviation could likely become ±0.3 kcal/mol. As such, these free energy calculations begin to become useful to calibrate and further refine molecular potential functions and to test out simpler approaches to calculate solvation free energies.[50] Given the level of agreement with experiment for amino acid side chains, we can *predict*[25] with some confidence the absolute free energy of solvation of nucleic acid bases which have not been determined experimentally. We can also use the free energy of solvation of hydrogen-bonded and stacked bases along with calculations of the free energy of association of bases in the gas phase to calculate the free energy of base pair association in water.[27,28] The stacked base pairs are more stable in water, and the hydrogen-bonded bases are more stable in the gas phase, consistent with experiments in water and CCl_4. The absolute free energies of association are in the range of $\Delta G = -1$ to -2 kcal/mol, consistent with experiment.[27] Although there are large error bars in the way we did the calculation, we have repeated it using potential of mean force free energy methods and have found similar results.[28] The agreement with experiment suggests that our potential function is well balanced and consistent with the TIP3P water model.[31]

One can combine state of the art quantum mechanics with free energy perturbation methods to calculate the relative stabilities of nucleic acid base tautomers in the gas phase and solution.[51] High-level quantum mechanics is required to calculate accurately the intrinsic gas-phase free energy difference, and these are in excellent agreement with experiments on three tautomeric equilibria including the amino ⇄ imino and enol ⇄ keto equilibria. In cytosine, ab initio calculations on the isolated molecule and experiments on the molecule in an inert matrix show the imino tautomer (which is complementary to A) to be as stable as the amino tautomer (complementary to G). However, the amino form is calculated to be solvated much more favorably. Thus, the accurate operation of the DNA genetic code as we known it clearly requires the bases to be in a highly polar environment, particularly during replication.

Relative Ligand Binding and Catalysis by Macromolecules. Thermodynamic cycle 4 can be used to analyze relative binding of different ligands S_1 and S_2 to an enzyme in either a noncovalent mode ES_i or a model for the transition state for catalysis ETS_i. The experimental free energies on binding, ΔG_1 and ΔG_2, and catalysis, $\Delta G_1^{\ddagger}$ and $\Delta G_2^{\ddagger}$, can be determined from the measured K_M and k_{cat} values. These horizontal processes are difficult to simulate because they involve



(4)

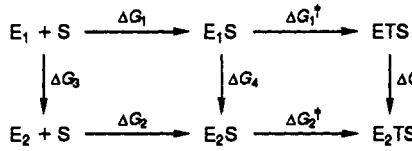
(ΔG_i) large charges and movements in water and substrate and ($\Delta G_i^{\ddagger}$) quantum mechanical effects. The vertical processes, ΔG_3, ΔG_4, and ΔG_5, are those that we have studied by free energy perturbation methods.

$$\Delta\Delta G_{bind} = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3 \quad (5)$$

$$\Delta\Delta G_{cat} = \Delta G_2^{\ddagger} - \Delta G_1^{\ddagger} = \Delta G_5 - \Delta G_4 \quad (6)$$

We note that ΔG_3 is identical with $\Delta\Delta G_{solv}$ since it corresponds to the relative solvation free energy in solution of S_2 and S_1.

Analogously, we can use cycle 7 to analyze the effect of site-specific mutations on the enzyme rather than the ligand on binding and catalysis. Equations 5 and 6 remain valid.



(7)

Our first application of this approach[52] to "drug design" involved comparing the free energy of binding of CBZ-Gly^P-X-Leu-Leu (S_1 = amidate, X = NH and S_2 = ester, X = O) inhibitors of the proteolytic enzyme thermolysin using cycle 4. Bartlett[53] had found for a variety of amidate (NH) and ester (O) inhibitors of thermolysin and $\Delta\Delta G_{bind} = 4.1 \pm 0.1$ kcal/mol, and the crystal structures by B. Matthews et al.[54] showed that the inhibitors bound nearly identically in the active site. The reason for this difference of 4 kcal/mol was attributed to the Ala 113 C=O hydrogen bond with the NH of the amidate. Our calculations on this system led to $\Delta\Delta G_{bind} = 4.2 \pm 0.5$ kcal/mol, with ΔG_4 (binding) = 7 kcal/mol and $\Delta\Delta G$ (solvation) = 3 kcal/mol. We thus suggested not only that differential solvation was important but also that the ester (O) inhibitor was held in place because of the leucines and the CBZ being in hydrophobic pockets and the phosphate interacting with the Zn^{2+} in what we called "forced repulsion". Thus, to relate the NH vs O free energy difference to the intrinsic strength of a hydrogen bond was, we felt, misleading.

In order to test this interpretation as well as use the theory in a *predictive* mode, which is a much more stringent test than reproducing experiment, we agreed with Paul Bartlett to *predict* the relative binding affinity of phosphinate inhibitors (X = CH_2). Before doing this, we studied the $\Delta\Delta G_{bind}$ (NH vs O) with different simulation protocols and molecular mechanical models and found $\Delta\Delta G$ values ranging from 3.3 to 5.9 kcal/mol. In addition, ΔG_3 varied from 0 to 3 kcal/mol, depending on models. It is clear that our estimated error of ±0.5 kcal/mol originally reported was optimistic, but also clear that all models were consistent with a large (>500) preference in binding affinity for

(48) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849.
(49) Debolt, S.; Kollman, P. A Theoretical Examination of Solvatochromism and Solute–Solvent Structuring in Simple Alkyl Carbonyl Compounds. Simulations Using Statistical Mechanical Free Energy Perturbation Methods. *J. Am. Chem. Soc.*, in press.
(50) Gilson, M.; Honig, B. *Proteins* **1988**, *4*, 7.
(51) Cieplak, P.; Bash, P.; Singh, U. C.; Kollman, P. *J. Am. Chem. Soc.* **1987**, *109*, 6283.
(52) Bash, P.; Singh, U. C.; Brown, F.; Langridge, R.; Kollman, P. *Science* **1987**, *235*, 574.
(53) Bartlett, P.; Marlowe, C. *Biochemistry* **1983**, *22*, 4618.
(54) Tronrud, D. E.; Holden, H. M.; Matthews, B. W. *Science* **1987**, *235*, 571.

X = NH vs X = O inhibitors. Two models were used in comparing X = NH to X = CH_2, and they showed, somewhat surprisingly, that both inhibitors were predicted to bind nearly identically to the enzyme ($\Delta\Delta G_{bind}$ = 0.0–0.3 kcal/mol favoring NH). A paper describing these results was submitted to *J. Am. Chem. Soc.* in September 1988[55] and a copy sent to Bartlett shortly thereafter. His group finished the synthesis of these inhibitors in October and tested a variety of amidate vs phosphinate inhibitors, finding a $\Delta\Delta G_{bind}$ averaging 0.1 kcal/mol,[56] in amazingly good agreement with the predictions. The calculations had indeed supported the importance of "forced repulsion", since the ΔG_4 was only 2 kcal/mol for NH → CH_2 and 5–7 kcal/mol for NH → O, as well as differential solvation, since the X = CH_2 inhibitor is 2 kcal/mol easier to desolvate than X = NH.

Two other systems have been studied in our lab with a focus on understanding and making predictions about different inhibitors binding to a common enzyme. In collaboration with Mark Murcko of Merck, we have studied binding of two different classes of carbonic anhydrase inhibitors to the enzyme. One set involved different hydrophilic groups that bound to zinc, and in the process a new type of inhibitor was predicted to be of interest to synthesize. The other set involved various hydrophobic tails on typical arenesulfonamide inhibitors. The first set required development of charge models for Zn^{2+}, which was a considerable challenge, and the other set involved testing models for large hydrophobic changes (hexyl → H) in the binding site and in water. Both of these systems were interesting and useful and gave results in reasonable agreement with various experiments.[57]

We have used free energy perturbation calculations to study the base sequence specificity of ribonuclease T_1–base interactions.[58] Ribonuclease T_1 (RT1) is known to cleave nucleic acids 3′ to G bases and to bind 2′GMP more strongly than other 2′-nucleotides, 2′GMP being more strongly bound ($\Delta\Delta G_{bind}$) than 2′AMP to the enzyme by 3.1 kcal/mol. A crystal structure is available for the RT1–2′GMP complex, so our first goal was to see if we could reproduce the $\Delta\Delta G$ for 2′GMP vs 2′AMP. We were successful, in that mutating 2′GMP to 2′AMP in solution led to a calculated ΔG_{solv} of 7.5 kcal/mol, quite consistent with the earlier study by Bash et al.,[25] in which only the 9-CH_3 bases were mutated into each other. In the enzyme active site, however, the 2′GMP → 2′AMP mutation was calculated to lead to a ΔG_{enz} = 10.3 kcal/mol, leading to a $\Delta\Delta G_{bind}$ = 2.8 kcal/mol, in very good agreement with experiment. The large ΔG in the enzyme site came about because the guanine base of 2′GMP forms seven H bonds to the enzyme, whereas the adenine base in 2′AMP only can form three. Only two H bonds are "conserved" in both inhibitors.

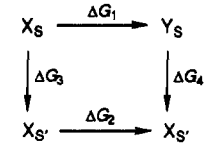
Another encouraging feature of the calculation is that, at the end of the mutation 2′GMP → 2′AMP → 2′GMP, the final structure is only 0.4 Å rms from the equilibrated initial structure and 0.8 Å rms from the 2′AMP complex structure. Thus, there is considerable structural reversibility here, although it is not perfect. The delicate balance between binding and solvation has been explored in other structural studies as well.[26,59–62]

Our first application of cycle 7 was to subtilisin, where Rao et al. predicted the effect on binding, $\Delta\Delta G_{bind} = \Delta G_4 - \Delta G_3$, and catalysis, $\Delta\Delta G_{cat} = \Delta G_5 - \Delta G_4$, for the Asn 155 → Ala mutation in subtilisin. The Genentech group had earlier studied the Asn155 → Thr and Asn 155 → His mutations and had found small $\Delta\Delta G_{bind}$ and large $\Delta\Delta G_{cat}$ values, but in any case the values for the Asn 155 → Ala mutation ($\Delta\Delta G_{bind}$ = 0.1 ± 0.8 (0.4 experimental) and $\Delta\Delta G_{cat}$ = 3.4 ± 1.0 (3.7 experimental) were in encouraging agreement.[63]

How does one construct the model for E_iTS or ETS_i necessary for the calculation of ΔG_5? We have used quantum mechanical calculations on stable tetrahedral adducts of CH_3O and amides to determine molecular mechanical parameters for tetrahedral species and then assumed that these adequately represented the transition state for amide hydrolysis in subtilisin. For this approach to work, it is essential that this tetrahedral intermediate resemble the transition state, that this transition state be similar in E_1 and E_2, and that the group being mutated not interact too strongly electronically with the atoms in the reactive part of the enzyme. The accuracy with which we calculate $\Delta\Delta G_{cat}$ suggests that this is so for this mutation in subtilisin. Hwang and Warshel[64] have also simulated this and other mutations in subtilisin and have been able to reproduce/predict $\Delta\Delta G_{bind}$ and $\Delta\Delta G_{cat}$ in Asn 155 mutants. Their approach, which uses an empirical valence bond method to simulate bond making/breaking, can in principle take into account electronic structure changes in E_1 and E_2 directly.

We have also used cycles 4 and 7, respectively, to study site-specific mutants in trypsin[65,66] and "site specifically mutated" substrates in α-lytic protease.[67] We were able to calculate $\Delta\Delta G_{bind}$ and $\Delta\Delta G_{cat}$ in these two enzymes.

Sequence-Dependent Stabilities. The free energy perturbation approach can be used to analyze sequence-dependent stabilities in proteins or nucleic acids. The analysis can be made in terms of thermodynamic cycle 8. The relative stability of sequence S



(8)

(55) Merz, K.; Kollman, P. *J. Am. Chem. Soc.* **1989**, *111*, 5649.
(56) Bartlett, P.; Morgan, B. P., manuscript in preparation.
(57) Merz, K.; Murcko, M.; Kollman, P. A. Free Energy Calculations on Carbonic Anhydrase Inhibitors. *J. Am. Chem. Soc.*, submitted.
(58) Hirono, S.; Kollman, P. Calculation of the Relative Binding Free Energy of 2′GMP and 2′AMP to Ribonuclease T1 Using Molecular Dynamics/Free Energy Perturbation Approaches. *J. Mol. Biol.* **1990**, *212*, 197.
(59) Lopez, M.; Kollman, P. *J. Am. Chem. Soc.* **1989**, *111*, 6212.
(60) Caldwell, J.; Kollman, P., unpublished results.
(61) Caldwell, J.; Kollman, P. *Biopolymers* **1986**, *25*, 249.
(62) Gough, C.; James, M.; Sielecki, A.; Kollman, P. Free Energy Calculations on Binding of Fluorostatines and Fluorostatones to Penicillopepsin. *J. Am. Chem. Soc.*, submitted.
(63) Rao, S.; Bash, P.; Singh, U. C.; Kollman, P. *Nature* **1987**, *328*, 551–554.
(64) Hwang, J. K.; Warshel, A. *Biochemistry* **1987**, *26*, 2669.
(65) Kollman, P.; Rao, S.; Brown, F.; Daggett, V.; Seibel, G.; Singh, U. C. *Protein Structure, Folding and Design 2*; Oxender, D., Ed.; A. Liss, Inc.: New York, 1987; pp 215–225.
(66) Seibel, G.; Kollman, P., manuscript in preparation.
(67) Caldwell, J.; Agard, D.; Bone, R.; Kollman, P. Free Energy Calculations on Alpha Lytic Protease Substrates. *Proteins*, accepted for publication.

in structures X and Y, ΔG_1, can be compared to the relative stability in sequence S', ΔG_2. However, the relative stabilities can be compared to the free energy changes upon mutating S → S' in structures X (ΔG_3) and Y (ΔG_4).

For example, we have applied this cycle to study the relative stability toward denaturation of native T4 lysozyme and its Thr 157 → Val mutant.[68] In this case, X = native protein, Y = denaturated protein, S = native sequence, and S' = mutated sequence, Thr 157 → Val.

$$\Delta\Delta G_{stab} = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3 \qquad (9)$$

Experimentally, one measures ΔG_2 and ΔG_1, but in the computer we can only simulate ΔG_4 and ΔG_3. ΔG_3 is straightforward to simulate, being analogous to ΔG_3 in cycle 7. What does one do to calculate ΔG_4, the mutational free energy for changing Thr 157 → Val in the denatured protein, given that one does not have a structure for the denaturated protein? Our approach was to use a tetrapeptide model of residues 156–158 as a model for the denatured protein, and when we did so, our calculated $\Delta\Delta G_{stab} = 1.9 \pm 1.1$ kcal/mol was in good agreement with the experimental value of 1.6 kcal/mol. Our calculations also suggested that $\Delta\Delta G_{stab}$ was 1.7 kcal/mol van der Waals interactions and only 0.2 kcal/mol electrostatic differential stabilization. This is presumably because the Thr O–H forms just as good H bonds to water in the denatured state as to protein groups in the native protein, but differential packing in the native structure (OH → CH_3) is more favorable for the smaller OH. We confirmed this result by carrying out a model calculation in which only the charges on C_β, O_γ, and HO_γ (which give all the H-bond stabilization) were mutated from their normal value to 0 on O_γ and HO_γ and a small value on C_β to insure charge neutrality. The free energy change upon this model mutation was within 0.1 kcal/mol of being identical in native and denatured protein. A further "model" mutation zeroed the charges on the backbone NH of Asp 159 in native and mutant proteins to assess the contribution of this group to differential stabilization. The resulting difference (0.6 kcal/mol) suggested that the H···O_γ interaction between the Asp NH H and the Thr 157 O contributed somewhat to the greater stability of the native (Thr 157) than mutant (Val 157) protein, but did not explain it all. Both of these model calculations emphasize the power of the free energy perturbation method to yield not only numbers but also useful insight.[65]

We have studied sequence-dependent DNA stabilities, where X = B DNA, Y = Z DNA, S = d-$(CGCGCG)_2$, and S' = $d(CGC^{5Me}GC^{5Me}G)_2$. Here we calculated $\Delta\Delta G_{stab}$ both in vacuo, with a distance-dependent dielectric and large hydrated counterions, and in solution, $\epsilon = 1$, with explicit water molecules and counterions, using periodic boundary conditions. The in vacuo calculations were done with a series of restraint weights, extrapolated to zero restraints, in order to insure that the structures remained in "canonical" B and Z structures. This was not required in the solution simulation.[69]

It is well-known that C → ^{5Me}C stabilizes Z DNA relative to B DNA, and Kollman et al.[70] suggested that this stabilization could be rationalized in terms of intramolecular van der Waals effects. On the other hand, Rich et al.[71] suggested that this differential stabilization could be rationalized through a hydrophilic effect, in that the 5-methyl group on cytosine was more buried in Z DNA than in B DNA.

The calculations[69] supported both points of view in that $\Delta\Delta G_{stab}$ for the in vacuo model was 0.9 kcal/mol, and for the solution model it was 1.7 kcal/mol. The latter divided by 4 (because four cytosines were mutated) was comparable to the experimental stabilization, which is in the range of 0.3–0.4 kcal/(mole·(base pair)).

Other calculations on DNA and RNA mutations by Tinoco's group in collaboration with us have also given results in good agreement with experiment.[72] We have also recently been able to reproduce the "Z phobicity" of AT base pairs by such methods,[73] and we have studied DNA daunomycin interactions.[74]

In summary, we have demonstrated the wide applicability of free energy perturbation methods to give both quantitative free energies and mechanistic insight into interactions involving biological molecules. At this point, the range of applicability of free energy perturbation approaches to macromolecular systems has yet to be fully established, but the perturbation of 2'GMP to 2'AMP in RNAse T_1, carried out over 40 ps in which six hydrogen bonds are broken/formed, suggests a wider applicability than is sometimes appreciated. However, one is often caught between the "Scylla" of wanting to sample longer and more accurately and the "Charybdis" of drift from the X-ray structure due to an imperfect representation of the system.

Other Systems and Qualitative Insights: TIM, Triostin A–DNA, Mitomycin–DNA, Papain, and the Neighbor Exclusion Principle

We now describe some other applications of computer modeling to biological systems, in which the problems/data available do not lend themselves to the application of quantitative free energy calculations.

We have been studying the mechanism of action of triose phosphate isomerase for some years. Our 1984 molecular mechanics study of His 95 → Gln mutation was, to our knowledge, the first such application of molecular mechanics to site-specific mutation.[75] Our calculations suggested a therefore unexpected reason why Gln 95 enzyme might be much less active than the His 95. A second study[76] used a quantum mechanical model for proton transfer to show how a Glu 165 → Asp mutation could be rationalized. If the transition state for rate-limiting proton transfer involved only a 0.3 Å longer O···C bond, a 10^3-fold rate decrease in proton transfer could be expected. Subsequent studies using free energy approaches have been aimed at examining

(68) Dang, L.; Merz, K., Jr.; Kollman, P. *J. Am. Chem. Soc.* **1989**, *111*, 8505.
(69) Pearlman, D.; Kollman, P. *Biopolymers* **1990**, *29*, 1193.
(70) Kollman, P.; Weiner, P.; Quigley, G.; Wang, A. *Biopolymers* **1982**, *21*, 1945–1969.
(71) Rich, A.; Nordheim, A.; Wang, A. H. J. *Annu. Rev. Biochem.* **1984**, *53*, 791.
(72) Ross, W.; Hardin, C.; Tinoco, I.; Rao, S.; Pearlman, D.; Kollman, P. *Biopolymers* **1989**, *28*, 1939.
(73) Dang, L. X.; Pearlman, D. A.; Kollman, P. A. Why Do AT Base Pairs Inhibit ZDNA Formation? *Proc. Natl. Acad. Sci. U.S.A.*, accepted for publication.
(74) Cieplak, P.; Rao, S.; Grootenhuis, P.; Kollman, P. *Biopolymers* **1990**, *29*, 717.
(75) Alagona, G.; Ghio, K.; Kollman, P. *J. Am. Chem. Soc.* **1984**, *106*, 3623.
(76) Alagona, G.; Ghio, C.; Kollman, P. A. *J. Mol. Biol.* **1986**, *191*, 23

the site-specific mutation Glu 165 → Asp 165 using free energy component analysis; such studies have suggested a potential role of Lys 13 in differentiating the Glu 165 and Asp mutant enzymes.[77]

Molecular dynamics methods have been usefully applied to simulate the functionally important loop closing event,[78] and we are now using it to compare a putative transition state for catalysis of wild type (Glu 165 Ser 96) and mutant (Glu 165 Pro 96, Asp 165 Ser 96, and Asp 165 Pro 96) enzymes. The results are intriguing and suggest that models of site-specific mutants.[79]

Triostin A binding to $d(CGTACG)_2$ in the crystal induces Hoogsteen base pair formation in the AT base pairs. Our molecular mechanics calculations[40] compared the energy of Hoogsteen and purely Watson–Crick base pairing in DNA and a triostin complex. Consistent with experiment, the double helix strongly prefers Watson–Crick base pairing because of reduced phosphate–phosphate repulsion. The presence of triostin significantly *stabilizes* the Hoogsteen pair, predominantly by sugar–peptide van der Waals interactions, beautifully illustrated in comparing van der Waals surface color slides of Hoogsteen and Watson–Crick models. The final energies of Hoogsteen and Watson–Crick structures are very close, consistent with NMR experiments which suggest, in the presence of triostin, pure Watson–Crick helices with some sequences and some Hoogsteen base pairs with others.[80]

Calculations were completed[81] on mitomycin interacting covalently in DNA double helices both with the O6 guanine attachment assumed to be correct at the time of the calculation and with an N2 attachment. It should be noted that we stressed that our calculations could not definitively predict which covalent linkage would be used, and this cautious attitude led to our calculations not becoming obsolete when the site of linkage was revised by subsequent experiments from O6 to N2.

Our calculations addressed the question of why DNA intercalations obey the neighbor exclusion rule, in which intercalation occurs only every other site. No unequivocal answer emerged,[82] but it is clear that simple molecular mechanics energies/structure cannot provide a definitive explanation. We have suggested vibrational entropy and counterion release as possible contributors to the unfavorable thermodynamics of neighbor intercalation binding.

Quantum mechanical calculations on SH^- attach on amides found, in contrast to similar calculations on OH^-/formamide, no local minimum for a tetrahedral adduct structure involving S.[83] Instead, only an ion–dipole complex minimum was found. This suggests that the precise mechanism of papain-catalyzed hydrolysis of peptides will be different from that of the serine proteases. We have examined some possibilities for concerted proton transfer in the sulfhydryl protease, but as yet have no definitive mechanism.[84] We think it extremely likely, on the basis of the quantum mechanical results, that simple XH^- attack on the amide bond followed by proton delivery to the leaving groups NRH, as is very likely for X = O, will not occur for X = S.

In summary, the appropriate use of quantum mechanics, molecular mechanics, molecular dynamics, and computer graphs has been shown in a number of cases to give useful qualitative insights into the structures and mechanisms of biological molecules.

P.A.K. is very grateful to all his collaborators, whose work is described in many of the references herein, and for research support from the NIH (GM-29072 and CA-25644) and the NSF (CHE 85-10066 and DMB-87-14775). The facilities of the UCSF Computer Graphics Lab, supported by NIH-RR-1081, R. Langridge, principal investigator, and the support of DARPA (under Contract N00014-86-K-0757 administered by the Office of Naval Research) are also gratefully acknowledged.

(77) Daggett, V.; Brown, F. K.; Kollman, P. A. *J. Am. Chem. Soc.* **1989**, *111*, 8247.
(78) Brown, F. K.; Kollman, P. A. *J. Mol. Biol.* **1987**, *198*, 533.
(79) Daggett, V.; Kollman, P. Molecular Dynamics Simulations of Mutants and Revertants of Triose Phosphate Isomerase: Rationalization of Relative Catalytic Potencies. *Protein Eng.*, in press.
(80) Gao, X.; Patel, D. S. *Biochemistry* **1988**, *27*, 1744.
(81) Rao, S. N.; Singh, U. C.; Kollman, P. *J. Am. Chem. Soc.* **1986**, *108*, 2058.
(82) Rao, S.; Kollman, P. *Proc. Natl. Acad. Sci., U.S.A.* **1987**, *84*, 5735.
(83) Howard, A.; Kollman, P. *J. Am. Chem. Soc.* **1988**, *110*, 7195.
(84) Arad, D.; Langridge, R.; Kollman, P. *J. Am. Chem. Soc.* **1990**, *112*, 491.