Title
Molecular mechanical and quantum mechanical simulations of enzyme catalysis

Author
Weiner, Scott Jay

Publication Date
1984

Peer reviewed|Thesis/dissertation

Molecular Mechanical and Quantum Mechanical Simulations of Enzyme Catalysis

by

Scott Jay Weiner

A.B., Occidental College 1980

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Chemistry

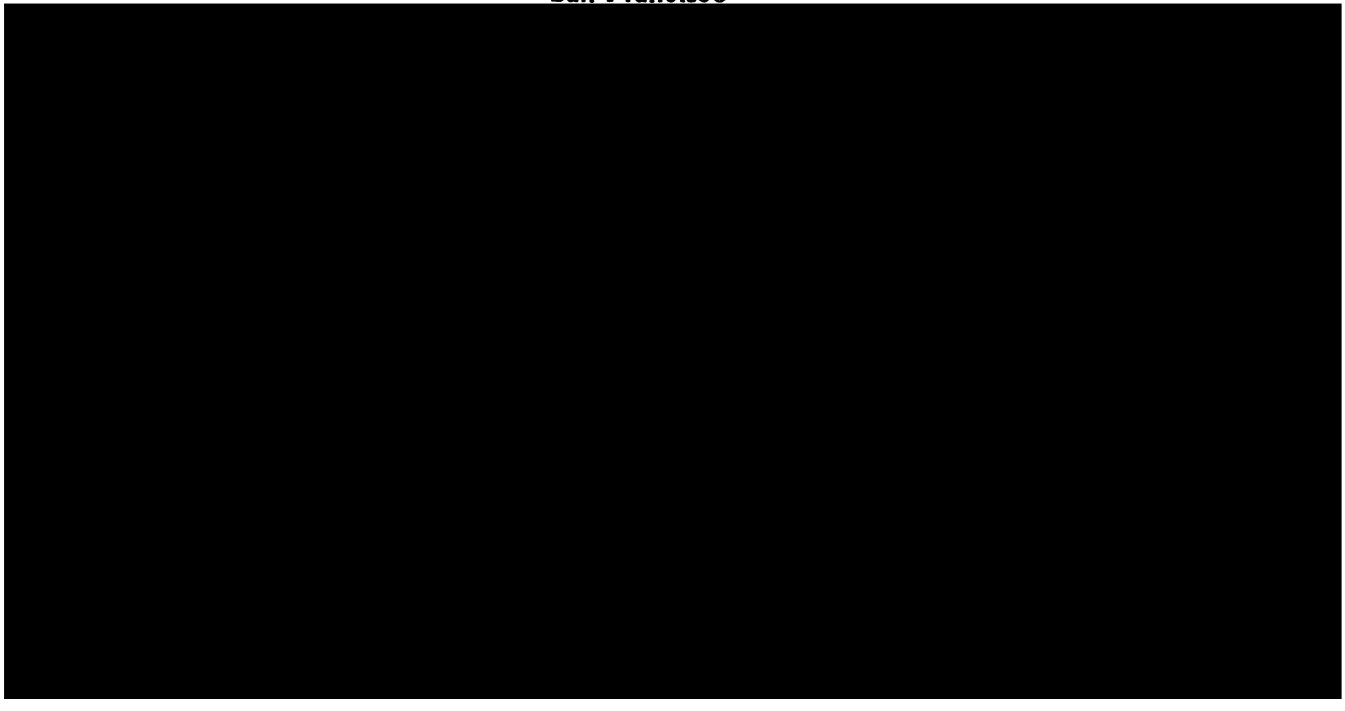
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



ACKNOWLEDGEMENTS

I would first like to acknowledge three people who have been extremely influential on me during my graduate study, Tack Kuntz, Ken Dill and Martin Shetlar. Each has, in his own way, been the catalyst for many exciting and productive conversations and collaborations. Both Tack and Ken served on my orals committee and were readers of this thesis; I sincerely thank them for that.

Moving to my peers, I acknowledge Joe Keepers for leaving U.C.S.F. so that I could finally get a desk of my own and for showing me that a strange link exists between chemistry and finance. Bob Tilton gets the credit for Bohemian Cigar Store and Happy Hour at Pier 39. Ron Verham was responsible for constantly reminding me that Southern California really is the Eden of this state. I thank John Taylor for showing me that your wildest dreams and aspirations really can be realized and for all of those stimulating conversations held over that wonderful Moffitt coffee! To Chandra Singh for teaching me about "those guys" and all of the other wonderful virtues of Singhlish. To John Thomason for helping me master Space Ace and showing me that there exists a graduate student out there who can actually spend a paycheck faster than I can. To George Seibel for thousands of hours of talk about anything and everything and to The Pub, chopping cars, Spot, unix heads, the emporium, a.d.c., hoser, the boss, triple D, wind tunnel and weirdness factor; remember, "Don't be a Marr!".

Finally, saving the best for last, I thank Peter Kollman for making my four years in graduate study exciting, challenging, enlightening and, most

importantly, enjoyable. He taught me the virtues of losing gracefully while still fighting strong to the end. He has shown me that it is possible to always stay young and energetic at heart regardless of what your biological time clock says. I will always remember him slaughtering me on the tennis courts, our arguments over nearly every scientific project, his pleasant demeanor in asking for "...just one more calculation" and for him always being the eternal optimist no matter how bleak the situation appeared. He has become one of my best friends and a small piece of his philosophy and outlook on life will always remain with me.

DEDICATION

To the two people who have always stood behind me and allowed me to pursue all of my dreams; I dedicate this thesis to Ron and Gloria Weiner.

TABLE OF CONTENTS

INTRODUCTION

Websters New Collegiate Dictionary defines **model** as: *a description or analogy used to help visualize something (as an atom) that cannot be directly observed.* In the most general sense, the focus of my graduate research has been to develop **models** for understanding both the structure and function of biological macromolecules. The only underlying hypothesis assumed in this thesis is that proteins exist in conformations (structure) and catalyze chemical reactions (function) in such a manner that the system tends toward a minimum in the free energy (ΔG). Thus, if we can evaluate the free energy (entropic (ΔS) and enthalpic (ΔH)) then we can address questions concerning the relative stabilities of various protein conformations or estimate barriers inherent in enzymatic reactions.

At the theoretical level, a detailed evaluation of all the important entropic contributions for a given process is beyond the scope of this thesis. Fortunately, for most of problems which are addressed here, the *relative* entropic energy tends to be very small compared to the enthalpic energy (when this is not the case then an attempt is made to approximate the entropic contribution and the limitations are pointed out). This leaves the internal energy, ΔE, which is nearly equal to ΔH at constant pressure in condensed phases as the energetic contribution which can be used by the theoretical biophysical chemist as a tool for better understanding protein structure and function.

The method used here for evaluating the internal energy of proteins and small molecules is the **molecular mechanics** approach. In molecular

mechanics, the internal energy of the system is defined by an empirical **potential energy function** $E(\vec{R})$, where $E(\vec{R})$ is a function of the spatial position of the atoms. For molecules, the energetic contributions to $E(\vec{R})$ can be represented by the predominant forces which are understood to be involved in determining molecular structure. The goal is then to minimize the potential energy as a function of the atomic positions to find the lowest energy structure for a given conformation (e.g. find R such that $E'(\vec{R}) = 0$).

The empirical energy function acts under a set of constants or parameters which will be referred to as a **force field**. Before any system can be energy refined, it is imperative to establish an accurate force field, which can be used with confidence, within the molecular mechanics framework. The first goal of this thesis (chapter 1) has been to develop a force field for simulation of proteins and nucleic acids. The emphasis in this chapter is on presenting the method used, its limitations and the actual parameter set derived. Once the force field has been developed, the focus shifts to applications of the force field in order to understand the structural aspects of proteins upon energy minimization.

The repeating nature of the polypeptide backbone makes N^{α}-acetyl-N-methylalaninamide (NANA) a representative model system for probing the structural and energetic conformations about protein Φ and Ψ torsion angles. Thus, the focus of the second chapter shifts to a full conformational analysis of alanyl dipeptide at both the molecular mechanical and *ab initio* quantum mechanical level. The results at both levels are compared with each other and with the best experimental results. Limitations in the theory are pointed out and the entropic contribution to the conformational energy is

estimated.

The second goal of this thesis is to establish a methodology for simulating chemical reactions in the gas phase and in aqueous solution. First the approach is elaborately described and applied to the hydrolysis of formamide by hydroxide ion (chapter 3). The energetics of bond making/breaking which occur in molecular reactions cannot be accurately calculated with simple molecular mechanics. To evaluate this important energetic contribution to the potential energy, we employ quantum mechanical techniques. An explicit solvent model consisting of 216 water molecules is added to the system at the molecular mechanics level. Solvation energies are calculated from molecular mechanics refined structures of intermediate steps along the proposed reaction pathway. Eight steps leading to product formation were analyzed and the gas phase vs. solution phase energetics for this small molecule reaction is discussed.

With the development of this method for simulating chemical reactions in solution, as well as completion of the parameter set for proteins, the time was ripe for combining the two to simulate an enzymatic reaction. Thus, the focus of the final chapter involves a simulation of the hydrolysis of a model tripeptide by trypsin. All of the predominant environmental and internal energies have been evaluated in a joint quantum mechanical/molecular mechanical framework. The energetics of the reaction pathway and a detailed structural analysis of the system is discussed.

Finally, the emphasis of this thesis is on understanding properties of biological macromolecules. It should be stressed that the ultimate goal is to develop and push the theory beyond its foreseeable limits so that the models

can be applied in areas ranging from drug design to protein modification.

(Chapters 1 and 2 have been published and appear as follows: S.J. Weiner *et al*, J. Amer. Chem. Soc., 1984, **105**, 765 and S.J. Weiner *et al*, J. Amer. Chem. Soc., 1984, **106**, 6243 respectively. At the time of this writing chapter 3 was in press: S.J. Weiner *et al*, "Simulation of Formamide Hydrolysis in the Gas Phase and in Aqueous Solution", J. Amer. Chem. Soc., 1984.)

CHAPTER 1

ABSTRACT

We present the development of a force field for simulation of nucleic acids and proteins. Our approach began by obtaining equilibrium bond lengths and angles from microwave, neutron diffraction and prior molecular mechanical calculations, torsional constants from microwave, NMR and molecular mechanical studies, nonbonded parameters from crystal packing calculations and atomic charges from the fit of a partial charge model to electrostatic potentials calculated by *ab initio* quantum mechanical theory. The parameters were then refined with molecular mechanical studies on the structures and energies of model compounds.

For nucleic acids, we focused on methylethylether, tetrahydrofuran, deoxyadenosine, dimethylphosphate, 9-methylguanine: 1-methylcytosine hydrogen bonded complex, 9-methyladenine: 1-methylthymine hydrogen bonded complex and 1,3-dimethyluracil base stacked dimer. Bond, angle, torsional, nonbonded and hydrogen bond parameters were varied to optimize the agreement between calculated and experimental values for sugar pucker energies and structures, vibrational frequencies of dimethylphosphate and tetrahydrofuran, and energies for base pairing and base stacking.

For proteins, we focused on Φ,Ψ maps of glycyl and alanyl dipeptides, hydrogen bonding interactions involving the various protein polar groups and energy refinement calculations on insulin. Unlike the models for hydrogen bonding involving nitrogen and oxygen electron donors, an adequate description of sulfur hydrogen bonding required explicit inclusion of lone pairs.

INTRODUCTION

There are two fundamental problems in simulating the structural and energetic properties of molecules: the first is how to choose an analytical function $E(\vec{R})$ which correctly describes the energy of the system in terms of its 3N degrees of freedom. The second is how the simulation can search or span conformational space ($\vec{R}$) in order to answer questions posed by the scientist interested in the properties of the system.

For complex systems, solutions to the first problem are an essential first step in attacking the second problem and, thus, considerable effort has been placed in developing analytical functions which are simple enough to allow one to simulate the properties of complex molecules yet accurate enough to obtain meaningful estimates for structures and energies.

In the case of the structures and thermodynamic stabilities of saturated hydrocarbons in inert solvents or the gas phase, the first problem has been essentially solved by molecular mechanics approaches of Allinger[1], Ermer and Lifson[2] and their coworkers. However, for polar and ionic molecules in condensed phases, unsolved questions remain as to the best form of the analytical function $E(\vec{R})$. In the area of proteins and peptides, seminal work has come from the Scheraga[3] and Lifson[4] schools. The Scheraga group has used both crystal packing (intermolecular) and conformational properties of peptides to arrive at force fields ECEPP, UNECEPP and EPEN for modeling structural and thermodynamic properties of peptides and proteins. Levitt, using the energy refinement software developed in the Lifson group, has proposed a force field for proteins based on calculations on lysozyme[5] and Gelin and Karplus have adapted this software along with many parameters from

the Scheraga studies to do molecular dynamics simulations of proteins[6]. Danber and Hagler[7] have clearly demonstrated the usefulness of crystal packing energies and structures in force field development. Hermans *et al*[8] have taken another approach, which combines the Scheraga nonbonded parameters with quadratic stretching and bending functions for use with x-ray data or in a stand alone mode to refine protein structures.

In the area of nucleic acids, the work of Sasisekharan[9], and Olson and Flory[10] was pioneering in the development of force fields but significant contributions have been made as well by Rein *et al*[11] and Pullman and Pullman[12]. Levitt has adapted his protein force field to nucleic acids and has carried out some important molecular mechanics and dynamics simulations on DNA fragments[13,14].

Our approach has been to use the powerful cartesian coordinate energy refinement of Lifson and Warshel[15] and to develop empirical force fields within this context. Our original parameter set was the first published nucleic acid force field in which complete geometry optimization of all atomic degrees of freedom could be carried out[16]. Our related force field for proteins[17] was similar but contained only modest modifications of the parameters used by Gelin and Karplus[6]. The most important changes concerned the explicit inclusion of H-bonding hydrogens and the use of partial charges taken from Mulliken populations of *ab initio* calculations.

Although many of the results of our simulations of proteins and nucleic acids with these force fields were encouraging, there were a few places where it was clear that improvements could be made[17,18]. The areas in most need of refinement involved the nonbonded (Lennard-Jones) and electrostatic

parameters. In view of the apparent power of our general approach for determining partial charges for complex molecules based on analysis of quantum mechanical electrostatic potentials[18], it seemed a propitious time to develop a second generation force field. Thus, in this paper, we present the development and, in the appendix, the results of a force field for proteins and nucleic acids.

GENERAL PERSPECTIVE

The basic equation for the force field is the same as that used earlier[16,17] with the addition of a weak 10-12 hydrogen bond term between hydrogen bonding hydrogens and H-bond acceptor atoms (equation 1). In the previous force field, the 10-12 coefficients C and D were set equal to zero for *hydrogen atom—hydrogen acceptor* interactions, following the results of Hagler *et al*[19]. This new force field contains a 10-12 function for two reasons. First, for strong H-bonds, it is clear that some repulsive term is required to prevent the occurrence of unrealistically short H-bonds[17] during energy refinement. Second, such a 10-12 function allows one to "fine tune" the H-bond distances and energies to desired values.

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

The bond stretching and bending functions are quadratic, which allows an adequate description of the structure and energies for relatively unstrained proteins and nucleic acids. A Fourier series approach to the torsional energy (i.e. more than one value of n may be used per dihedral angle in equation 1) allows rather accurate simulation of conformational preferences in simple

and complex molecules. For computational speed, a 6-12 function is used for the nonbonded parameters even though a 6-exp is likely to be be a better simple functional form[20]. Hagler *et al*[19] compared 6-12 and 6-9 nonbonded functions in crystal packing calculations and found neither to be clearly superior. As long as the interatomic distance is not well below the sum of the van der Waals radii, the 6-12 form should be adequate.

We retain the atom centered monopole approach to the electrostatic energies (with the exception of sulfur, where lone pairs are also included). This approach appears to do a satisfactory job in simulating molecular electrostatic interaction energies, provided the charges are chosen in a reasonable fashion. We feel that the fit of the potential charge model to quantum mechanically calculated electrostatic potentials is a superior method for determining the point charges.

We use a distance dependent dielectric, $\varepsilon=R_{ij}$, for the electrostatic energies; although we demonstrate in hydrogen bonding cases of model systems that results with a constant dielectric constant, $\varepsilon=1$, are very similar to those found with $\varepsilon=R_{ij}$. A rationale for using a distance dependent dielectric constant is that it mimics the polarization effect in attractive interactions, with closer interactions weighted more heavily. Second, it helps compensate for the lack of explicit solvation by implicitly damping longer range charge interactions more than shorter range ones. There is empirical and computational support for such a model[17], given that solvent (water) is not explicitly included in the calculation. However, when water is explicitly included, a constant dielectric constant is probably more appropriate, and, as noted below, this appears to be easy to achieve with the same set of charges and

small modifications in the 10-12 H-bond parameters.

In the "united atom" force field presented here, we include all atoms explicitly with the exception of hydrogens bonded to carbon. We should stress that this is merely for computational efficiency in simulations of large proteins and nucleic acids. Below we present all atom simulations on a sugar puckering model of nucleosides and on the Φ,Ψ maps of alanyl and glycyl dipeptides. In both cases, the united atom representation gives results quite similar to the all atom model. In a previous study[17], we used a united atom approximation for prealbumin but included the aromatic hydrogens of thyroxine explicitly, in order to correctly reproduce the Φ_1,Φ_2 conformational energies of thyroxine (these energies are strongly influenced by H---I non bonded interactions). Thus, there will be cases where a hybrid force field is appropriate, and, as noted below, this is straightforward to implement.

In the development of force field parameters, we used the following approach: we began with an initial set of parameters and then carried out simulations on a number of model systems, relevant to proteins and nucleic acids, to test these parameters or to determine some from scratch. One of the frustrating aspects of force field development is the dependence of the final results on "the pathway" or choice of model systems. It is thus incumbent on the developer to elucidate his pathway as clearly as possible, to insure that further work need not start from scratch. This methodology is carried out below using the AMBER molecular mechanics program[21].

FORCE FIELD DEVELOPMENT

ATOM TYPES

The basis of a force field is the choice of atom types, i.e. the selection of atoms which are enough alike, both chemically and physically, to be treated identically in the molecular mechanics refinement. In the case of a quantum mechanical calculation, one needs only a single atom "type" per atom, i.e. only the number of electrons is relevant. The decisions on atom types are inevitably compromises between possessing the most accurate representation of many molecules and having a manageable number of types. We list the types and their characteristics in Table I, so only a single comment is in order. The sp^3 atom types are fairly typical, but we have included more sp^2 types than earlier force fields to insure increased geometrical precision for such ring systems as purines, pyrimidines, indoles and imidazoles.

SOURCES OF PARAMETERS

(1) NON-BONDED PARAMETERS

The most difficult set of parameters to derive *a priori* are the nonbonded ones. We used as our starting point for sp^2 atoms the 6-12 and 6-9 parameters, derived by Hagler *et al*[19] from a fit of the lattice energies and crystal structures in amides. The significant difference between the 6-9 and 6-12 values of R* (van der Waals minimum) and ε* (van der Waals well depth) for a given atom caused us not to take these parameters directly. For example, the carbonyl carbon van der Waal radius *increased* from R*=1.81 Å, ε=0.184 kcal/mole in the 6-9 force field to R*=2.175 Å, ε=0.039 kcal/mole in the 6-12; whereas the aliphatic hydrogen *decreased* from R*=1.77 Å, ε=0.0025 kcal/mole in the 6-9 potential to R*=1.375 Å, ε=0.038 kcal/mole in the 6-12. In the 6-9 potential, oxygen and carbon had nearly the same size, with nitro-

gen 0.4 Å larger, whereas in the 6-12, the sizes varied in a smooth fashion from R*=1.6 Å (oxygen), 1.95 Å (nitrogen) and 2.2 Å (carbon).

To avoid these inconsistencies, we began with the Hagler *et al*[19] 6-12 oxygen parameters since oxygen (of C, N and O) has the most direct contact with neighboring molecules in the amides. This led us to take R*=1.6 Å and ε=0.20 kcal/mole for oxygen. We expect nitrogen to have a larger R* than oxygen, in the range of 0.1-0.2 Å larger, given the standard van der Waals radii of the atoms determined by observed atom-atom contacts in crystals[22,23]. The well depth of nitrogen was consistently 0.04-0.06 kcal/mole less than oxygen in the Hagler *et al* study, leading us to settle on the compromise parameters R*=1.75 Å, ε=0.16 kcal/mole. We then estimated the parameters for sp^2 carbons in an analogous fashion, obtaining R*=1.85 Å, ε=0.12 kcal/mole. Since all of these values are close to 0.2 Å larger than the "standard" van der Waals contact radii from crystal packing data[22,23], we also selected "larger" values for both phosphorus (R=2.10 Å, ε=0.20 kcal/mole) and sulfur (R*=2.00 Å, ε=0.20 kcal/mole) to be consistent within this framework. These P and S values are similar to those found in MM2[1], although a 6-exp is used there.

For aliphatic CH, CH_2 and CH_3 groups, (atom types CH, C2 and C3), there are two papers in the literature which suggest appropriate van der Waals 6-12 parameters for these extended atoms. Dunfield *et al*[24] have determined values based on crystal packing calculations of hydrocarbons and Jorgensen[25] has calculated the parameters from Monte Carlo liquid simulations of ethers and alcohols. These two parameter sets are very similar, suggesting appropriate values to use for united atoms. However, in our simulations of the conformational profile of methylethylether, *n*-butane and deoxy-

adenosine (and in our earlier study of base paired dinucleoside phosphates[16]), we found that the use of van der Waals parameters with R* as large as Jorgensen's or Dunfield's (for example, for a CH group R*=2.385 Å and ε=0.049 kcal/mole) gave a significantly poorer representation of intramolecular energies and structures than did smaller values. With methylethylether, *n*-butane and deoxyadenosine as model systems (described in detail below) we settled on compromise values of R*=2.00 Å, ε=0.15 kcal/mole for C3, R*=1.925 Å, ε=0.12 kcal/mole for C2, R*=1.85 Å, ε=0.09 kcal/mole for CH and R*=1.8 Å, ε=0.06 kcal/mole for CT (sp^3 carbon without hydrogens). This last value for CT was a compromise between our expectation that CT should be smaller than CH, but larger than an sp^2 carbon. In view of our expectation that the sp^3 atoms should be somewhat larger and less polarizable than sp^2, we used R*=1.85 Å, ε=0.12 kcal/mole for sp^3 nitrogen and R*=1.65 Å, ε=0.15 kcal/mole for sp^3 oxygens in alcohols and ethers. The aliphatic hydrogen parameter (R*=1.37 Å, ε=0.038 kcal/mole) was taken from the Hagler *et al* study[19]. We used a significantly smaller (R*=1.00 Å, ε=0.020 kcal/mole) value for potentially H-bonding hydrogens (N-H, O-H, S-H), in view of the fact that these atoms have significant parts of their density shifted to the heteroatom to which they are attached. Our lone pair van der Waal parameters (used only for sulfur) come directly from MM2[1].

In this force field, for interactions involving hydrogen bonding hydrogens and heteroatoms, we replace the 6-12 parameters with 10-12 parameters of well depth 0.5 kcal/mole except in one case noted below. The retention of the 6-12 parameters in these cases would lead to much too long H-bond distances while using no H-bond parameter, (as done by Hagler *et al*[18]) leads to H-bonds, in some cases, which are too short[17].

In the original Hagler *et al* study[19], all the nonbonded H---X (hydrogen bound to N or O; X=any atom) van der Waals parameters were taken to be the geometric mean of the H and X parameters. Since their study found that this approach gave good H-bond distances, they used A=0 and B=0 for these specific hydrogen nonbonded parameters. However, such an approach can lead to artifacts in which H---C distances become arbitrarily short. In our previous calculations[16,17], we had set 6-12 nonbonded parameters for *H atom—H bond acceptor* atoms equal to zero and used the geometrical mean values $A_{ij}=(A_iA_j)^{1/2}$ and $B_{ij}=(B_iB_j)^{1/2}$ for all other heteroatomic interactions involving these hydrogens. Here we employ the same approach, except for *H atom—H bond acceptor*, where a 10-12 function is used.

(2) ELECTROSTATIC PARAMETERS

We have used quantum mechanical calculations of the electrostatic potential to derive charges for atoms in salient molecules, as described in detail elsewhere[18]. This method uses quantum mechanically calculated electrostatic potentials to numerically fit atomic charge models. While we feel that this is a superior method for determining such charges, the process is still subject to three uncertainties. First, while the charges may depend upon the conformation of the molecule used, it is impractical to possess a separate set of charges for every conformation. Second, it is only practical to do quantum mechanical calculations for fragments of polymers and then to "patch" these together. Finally, the charges will differ depending on the basis set chosen. Elsewhere[18], we have analyzed the error due to the first problem by carrying out calculations on C3' *endo* and C2' *endo* conformations of a deoxyribose model. The polar group charges derived by a fit to the

potential in a C2' *endo* conformation were qualitatively similar to those found with C3' *endo* (within 5-10%). We have analyzed the second problem by comparing the charges derived by fitting the electrostatic potential of 1-methylcytosine, 1-aminodeoxyribose and dimethylphosphate, and "averaging" the charges at the linking atoms with the charges derived from a fit of the electrostatic potential for cytosine 3'-phosphate. Not only is the agreement between the derived charges generally quite good (within 0-10%), but the charges for atoms in the linker regions (C1' and O3' in the sugar) are similar (0.500 and -0.535 on the basis of the truncated models and 0.547 and -0.514 for cytosine 3'-phosphate). The final problem, the basis set dependence of the charges, is one that is crucial to deal with. We have followed the approach of Cox and Williams[26] (described below) and have checked the H-bond energies derived with these charges for consistency with experiment and/or accurate *ab initio* calculations. The use a 10-12 H-bond function, with a well depth of 0.5 kcal/mole, enabled us to "fine tune" hydrogen distances by varying the repulsive R^{-12} H-bond coefficient. In two cases, a small change in the point charges was necessary to insure accurate H-bond energies and geometries (see below).

(3) BOND LENGTH AND BOND ANGLE PARAMETERS

We took the parameters for equilibrium bond length, r_{eq}, and bond angle, ϑ_{eq}, from microwave and x-ray data on appropriate compounds. For example, ϑ_{eq}(C3-C2-C3) came from the microwave structure of propane[27], this being the most appropriate source for it. We made efforts to select the highest quality data on a reasonable reference compound, rather than less accurate data on a particular molecule which might more closely resemble

the fragment considered.

We were able to find suitable values for all r_{eq} parameters in the experimental literature. Many of the K_r came from normal mode calculations, in which the K_r values were varied to give the best fit to experimental frequencies of tetrahydrofuran, dimethylphosphate, N-methylacetamide, methanol, methanethiol, dimethylsulfide, dimethyl disulfide and benzene. We used a linear interpolation model for the remaining bond stretching force constants involving partial double bonded sp^2 atom---sp^2 atom (described in the next paragraph).

In accord with our harmonic approximation, we assume that any stretching force constant can be calculated via a direct linear interpolation between the "pure" C-C single bond (r_{eq}=1.507 Å, K_r=337 kcal/mole Å^2) and "pure" double bond (r_{eq}=1.336 Å, K_r=570 kcal/mole Å^2). The MM2 force constant for a *single* bond was taken as a fixed reference point. The pure *double* bond came from the analogous carbonyl stretching K_r (since C=O and C=C possess similar stretching frequencies) calculated from our normal mode analysis of N-methylacetamide (described below). The structural parameters were selected from microwave data on propane and propene respectively. This algorithm (Table II) was applied to all remaining carbon-carbon bonds, regardless of specific atom type, in building a consistent family of stretching force constants. Wherever applicable, we checked calculated force constants from our normal modes analysis with predicted scaled values. For example, our interpolation algorithm predicted a stretching force constant for benzene (r_{eq}=1.40 Å) of 475 kcal/mole Å^2 and the calculated value, which gave the best fit to the experimental frequencies, was 469 kcal/mole Å^2.

Similarly, an analogous interpolation scheme was employed for all carbon-nitrogen bonds. The "pure" single bond N-C equilibrium distance of 1.449 Å came from Benedetti's[28] structural parameter for the N-C_α bond while the K_r=317 kcal/mole Å^2 was taken from from MM2. For the pure double bond N=C force constant, we selected Harmony's[27] microwave data on methylene imine (r_{eq}=1.273 Å) and used the default value of 570 kcal/mole Å^2 for K_r. Our algorithm predicted K_r=490 kcal/mole Å^2 for the *partial* C=N bond of atom types found in the amide linkage, while the value which gave the best fit to experimental frequencies was 488 kcal/mole Å^2. These close correlations support the usefulness of our interpolation method for derivation of approximate stretching force constants. In this fashion we were able to derive K_r values for all the bonds in our force field.

We should note that this choice of 570 kcal/mole gives approximately 100-200 cm^{-1} too low frequency for pure C=O and C=C double bonds, in such systems as acetone and 2-butene, where a value of $K_r \approx$ 700 kcal/mole Å^2 is required to fit the vibrational frequencies. However, the use of such a pure C=C force constant gives a much poorer K_r for benzene. Since proteins and nucleic acids have more "aromatic" character than pure double bond C=C or C=N, we chose to use the 570 kcal/mole Å^2.

The development of bond angle parameters followed a similar route. We chose initial ϑ_{eq} values from experimental data on appropriate reference compounds; e.g., ϑ_{eq} for C3-C2-C3 (generally CX-C2-CX) came from the respective angle in propane. We note that such a choice for X-C2-X indirectly corrects for the absence of explicit hydrogens on C2. Initial values of K_ϑ for typical Xsp^3- Xsp^3-Xsp^3 values came from MM2, but both K_ϑ and ϑ_{eq} values

were altered in our model calculations on THF, described below. Our normal mode calculations also played a large role in our choice of K_ϑ values. For example, we made the assumption that all K_ϑ(C2-C2-X), where X=OS, OH, N* or other electronegative atom was the same as the K_ϑ(C2-C2-OS) derived for THF. In the above fashion we were able to derive reasonable and internally consistent values for all the K_r, K_ϑ, r_{eq} and ϑ_{eq} parameters in our force field.

(4) TORSIONAL PARAMETERS

Our torsional parameters initially came from experimental data on conformational equilibria of molecules. However, since the nonbonded and torsional terms are highly coupled, many were modified during the test case studies described below.

The torsional parameters we use are divided into three types: *general*, ***specific*** and ***improper***. Examples should suffice to illustrate each. The ***general*** torsional parameter for X-CH-CH-X has $V_3/2$=2.0 kcal/mole, # bonds = 4 and phase γ=0°. Recall that each CH atom has two non hydrogen attachments, implying that there are four X-CH-CH-X associated with a given bond. Each of these are assigned a torsional potential of magnitude 0.5 kcal/mole with γ=0°, to be consistent with $V_3/2$=2.0 kcal/mole (0.5 x 4 = 2.0). Such a phase leads to a preference for staggered over eclipsed bonds.

Any ***specific*** parameter, such as OS-CH-CH-OS, overrides any *general* parameter. For this example we use both the $V_3/2$=0.5 kcal/mole and a $V_2/2$=0.5 kcal/mole to insure the correct ***gauche*** tendency of O-C-C-O units[29,30]. In the case of the peptide bond, we employ a $V_1/2$=0.65 kcal/mole Fourier component to correctly reproduce the ***cis / trans*** energy difference for N-

methylacetamide of 2.1 kcal/mole[31].

Improper torsions such as X-X-N-H (e.g. C'-C_α-N-H for a peptide) possess four atoms ***not*** bonded in a successive fashion to one another and serve two purposes. The two fold terms insure the correct planar tendency of sp^2 atoms while the three fold terms (e.g. X-X-CH-C2) keep asymmetric centers, such as the C_α of amino acids, from racemizing when one uses the united atom approximation. The magnitude of the V_2 terms has been determined from normal mode calculations on N-methylacetamide (described below); $V_3/2$ is assigned an arbitrarily large (14 kcal/mole) value.

In the case of proteins, there are a number of cases where normal two fold torsional parameters were used, for example, in the amide bond, to reproduce a rotational barrier of 20.0 kcal/mole (ref. 3), in X-S-S-X bonds, to insure a gauche tendency and in tyrosine O-H, to maintain a barrier of 4.0 kcal/mole about the C-O torsion. We carried out model calculations on single amino acid systems, to insure that our V_2 values lead to net barriers in reasonable agreement with experiment. Our torsional parameters in those cases are very similar to the ones in ref. 3.

For the many torsions where no detailed experimental data exists, we initially employed a simple linear interpolation method to derive V_2(X-C_{sp^2}-C_{sp^2}-X) and V_2(X-C_{sp^2}-N_{sp^2}-X), under the assumption that the relative bond length was inversely proportional to V_2. We then compared the calculated values for benzene (r_{eq} = 1.397 Å) and N-methylacetamide (r_{eq} = 1.335 Å), and found that a simple linear interpolation produced a significantly higher V_2 value than that which gave a good fit to the out of plane vibrational frequency. For this reason, we employed a dual linear scaling method using a *pure* single, *pure*

double and *partial* double bond (i.e. NMA) as reference points (Table II). An example here will suffice to exhibit this algorithm. The *pure* double bond had $V_2/2$=30.0 kcal/mole from methylene imine, the *partial* double bond was $V_2/2$=10.0 kcal/mole as fit from our normal mode analysis of NMA and the barrier to rotation for *pure* single bond was assumed to be 0.0 kcal/mole. The structural parameters came from microwave experimental data on the NMA (CH_3-N), NMA (C=N) and methylene imine (C=N) for single, partial double and pure double bond character respectively. All C-N bond lengths between 1.335 Å and 1.449 Å were scaled between 10.0 kcal/mole and 0.0 kcal/mole, while C-N bonds between 1.273 Å and 1.335 Å were scaled between 30.0 kcal/mole and 10.0 kcal/mole. Our C-C torsional parameters were analyzed in an identical fashion (Table II). We found that the $V_2/2$ value which fit the lowest out of plane frequency for benzene was 5.5 kcal/mole. This value then represented the partial double bond character for C=C bonds and was used as an intermediate in the scaling algorithm for C-C torsions.

NUCLEIC ACID TEST CASES

(1) TETRAHYDROFURAN AND METHYLETHYLETHER

We began our analysis of model systems with a study of tetrahydrofuran (THF) and methylethylether (MEE). THF was selected to give us appropriate values for C-C-O, C-O-C and C-C-C bending force constants as well as C-C and C-O torsional potentials. The results of calculations on MEE suggest which C3---C3 van der Waals parameters to select.

THF, in the united atom approximation, is a five atom system OS-C2-C2-C2-C2. We began with microwave data on small models to determine r_{eq} and ϑ_{eq},

except for r_{eq}(C-O), where we used a standard value of 1.425 Å instead of the MM2 value of 1.41 Å (the length in methanol and closer to the average of the Cambridge crystal file average for phosphate ethers of 1.422 Å, rather than the 1.410 Å value for dimethylether). MM2[1] values were taken for the bond stretching parameters K_r for C-C and C-O bonds and bond bending parameters K_ϑ for C-C-C, C-O-C and C-C-O angles. The torsional parameters, $V_3/2$=1.5 kcal/mole for X-C-C-X and $V_3/2$=1.0 kcal/mole for X-C-O-X bonds roughly reproduce typical rotational barriers in alkanes and ethers. We set as our goal the reproduction of the following experimental data for THF (Table III): the energy difference between C_2 and C_S geometries of 0.1 kcal/mole, the barrier to planarity $\Delta E(C_{2v}-C_S)$=3.7 kcal/mole, the sugar pucker parameter $q \approx 0.4$ Å and the bond angles C-C-C, C-O-C and C-C-O for the C_2 and C_{2v} conformations.

To reach this goal, we varied K_ϑ and ϑ_{eq} as well as the torsional parameters. The MM2[1] values for the bending parameters are 32, 50 and 55 kcal/mole rad^2 for C2-C2-C2, C2-C2-OS and C2-OS-C2 respectively. We only altered the C2-C2-C2 value to 40 kcal/mole rad^2 to obtain a good fit with the various experimental data, with $V_3/2$(X-C2-C2-X)=1.45 kcal/mole, $V_3/2$(X-C2-OS-X)=1.05 kcal/mole and $V_2/2$(C2-C2-OS-C2)=0.1 kcal/mole (The latter torsional parameter is only a small perturbation used to refine the *gauche-trans* energy difference in methylethylether but since it makes physical sense, we retain it. One would expect from the work of Brunck and Weinhold[42] that the oxygen lone pairs would prefer to be *trans* to C-C rather than C-H bonds and thus C-O-C-C would have a small *gauche* tendency from electronic through bond effects). The increase from the MM2 K_ϑ(C-C-C) is also sensible since that value refers only to C-C-C bending while our force field should have

implicit contribution from H-C-H and H-C-C angle distortions. The results are summarized in Table III under the heading FF1 and indicate that we have done a reasonable job of reproducing the experimental data for THF noted above. After this had been completed, the normal mode calculational facility became available in AMBER[21,43] and we carried out such calculations on dimethylether and dimethylphosphate (described below).

We found that the bond bending force constants in the range of 40-55 kcal/mole rad^2 could not qualitatively reproduce the bending frequencies of THF. We confirmed that this was not an artifact of the united atom approximation by carrying out parallel calculations with an all atom model. Hence, we began to search for a set of bending parameters which could more closely reproduce the structures, energies and vibrational frequencies of THF. Table III contains the end results of our calculations. To reach the goal of reproducing the experimental frequencies, it was necessary to increase the bending force constants by 50-100%, which caused significant flattening of the ring. Increasing the torsional parameters compensated for this by raising the barrier to planarity, resulting in significantly larger but still qualitatively reasonable values of $V_3/2$(X-C2-C2-X)=2.0 kcal/mole and $V_3/2$(X-C2-OS-X)=1.45 kcal/mole. These changes in the torsional barrier worsened the agreement with experiment for the lowest non-pseudo rotation mode (out of plane torsion in table III), but the sum of the errors between calculated and observed frequencies, for the torsion and bending modes, is significantly reduced. It was clear from our vibrational analysis calculations on all atom and united atom THF that the separation of C-C and C-O models from H-C-C bending would be very difficult. This fact, plus the ambiguity in the assignments of C-C and C-O stretching modes from THF vibrational analyses in the

literature, caused us to use the K_r(C-C) and K_r(C-O) derived below for diethylphosphate.

In parallel with the THF calculations, we had been carrying out calculations on methylethylether using the bond, angle and torsion parameters from THF and varying the C3---C3 van der Waals parameter, as a means of ascertaining which values could best reproduce the experimental data on MEE. We found that, with the use of Dunfield's[24] or Jorgensen's[25] van der Waals parameters, we could not get qualitatively reasonable barriers of rotation about the C-O bond for methylethylether. We reverted to C3 van der Waals parameters of about the same R*, and somewhat smaller ε, as found in our previous force field[16]. We felt that the previous force field had possessed too deep van der Waals well depths and this had helped lead to excessively attractive energies for ligand-protein binding[17]. However, we were concerned that our small R* (compared to the Dunfield and Jorgensen values) might lead to a significant "collapse" during refinement. Thus, it was necessary to use as large a C3 R* as possible and still get reasonable properties for MEE.

We examined the effect of "scaling down" the 1-4 van der Waals (van der Waals interactions separated by only 3 bonds) by an empirical factor, as had precedent in the work of Dunfield[3]. Hagler *et al*[44], in studies of peptides, argued that if one allows bond angle relaxation, the necessity of reducing 1-4 van der Waals parameters disappears. However, there are a number of compelling reasons for choosing to scale down 1-4 van der Waals interactions. First, a R^{-12} repulsion term is too steep compared with the more correct exponential term and the error from this should be largest for the close 1-4 interactions. Second, there is likely to be more significant charge redistribution

during close 1-4 interactions than during corresponding intermolecular contacts, which will have the effect of reducing the net repulsion. Third, this reduction will enable us to use R* values closer to those of Dunfield's[24] or Jorgensen's[25] and still calculate reasonable properties for MEE. Below, we give further empirical examples of how reducing the 1-4 van der Waals interactions by a factor of 0.5 improves results of calculations on nucleosides and peptides. Such calculations led us to settle on an R* value of 2.00 Å and ε=0.15 kcal/mole for C3. This value was based on calculations with FF1 (Table III). When we derived FF2, the calculation was repeated on MEE. The agreement with experiment was better for the 120°-*trans* difference, not so favorable for the *gauche-trans* difference and significantly worse for the *cis-trans* energy difference. However, the *cis* conformation is sufficiently high in energy that, with a simple united atom force field, one cannot expect to reproduce the values as accurately as for low energy structures.

We wished to assess how much the effect of using this van der Waals parameter (plus the value of R*=1.65 Å, ε=0.15 kcal/mole for OS) would have on the calculated density and vaporization enthalpy of dimethylether (DME) previously studied by Jorgensen in Monte Carlo simulations[45]. Jorgensen had found a density 3% too low and an enthalpy of sublimation 5% too high for DME. Since our van der Waals parameters allow closer contacts, but have shallower well depths, we expected that a Monte Carlo simulation of DME with our van der Waals parameters would lead to errors in the opposite direction as those of Jorgensen. We carried out[46] such a simulation of DME using the same electrostatic parameters as his and found, indeed, that our density was 10% too high and the enthalpy of sublimation 10% too low.

We also tested these parameters on *n*-butane and calculated a *gauche-trans* energy difference of 0.91 kcal/mole, a dihedral angle in the *gauche* conformation of 66° and a C-C-C angle in the *gauche* conformation of 114.1°, all in quite good agreement with the experimental values of 0.97 kcal/mole[47], 67.5° and 113.5°[48] respectively. As in the case of MEE, we did, however, overestimate the *cis* barrier (calculated at 6.9 kcal/mole, with HFSCF + CI results of 4.5 - 4.7 kcal/mole)[48].

(2) DIMETHYLPHOSPHATE

The geometrical parameters r_{eq} and ϑ_{eq} for the phosphate group of dimethylphosphate (DMP) were taken from the older "standard" values for the $ROPO_2OR'$ group suggested by Newton[49]. A more recent critical analysis using the Cambridge crystal data bank[50] found parameters within a standard deviation of these values, leading us not to change the earlier parameters. We also left our torsional parameters as in the old force field, with both $V_2/2$ and $V_3/2$ terms at 0.75 kcal/mole. The results of calculations using our old and new parameters are compared with the best quantum mechanical values and the average experimental values in Table IV. The calculated energy as a function of conformation is qualitatively reasonable, although the gap between the molecular mechanical energies of various conformations is still significantly larger than that found quantum mechanically. However, the quantum mechanical values came from calculations with a minimal basis set and some of the OS-P-OS and O2-P-O2 angles calculated with such an approach are outside the range of experimental values[50,51]. Hence, we have not required that the molecular mechanical calculations give relative energies in precise agreement with the quantum mechanical results. It is clear

that one could make the relative molecular mechanical conformational energies for *g*,*g* and *g*,*t* agree more closely with the quantum mechanical difference of 0.2 kcal/mole by merely reducing the two-fold torsional parameter, $V_2/2$, for the C-O-P-O linkage.

There are two other interesting structural aspects of DMP that deserve comment. Gorenstein *et al*[51] noted that both quantum mechanical calculations and x-ray structures suggest that ϑ(O-P-O) is strongly dependent on phosphate conformation, with x-ray structure averages $\vartheta_{eq}=104.8°$ for *g*,*g* conformation and 99.3° for *g*,*t*. Second, ϑ_{eq}(O-P-O) angles differ by $\approx 5°$ when there are *gauche* ROPO linkages, since one anionic oxygen is *gauche* to the methyl group and the other is *trans* (ϑ_{eq} (OPO' *gauche*)=110° and ϑ_{eq} (OPO' *trans*)=105°). Both the new and old force fields show evidence of these two effects, although in FF2 they are an order of magnitude too small. In the old force field, the magnitudes of the effects are 20-60% of the observed ones, most likely due to the fact that the 1-4 van der Waals were not scaled by a factor of 0.5. Again, this is a situation where one must compromise; we deem the reproduction of the structure and energies of MEE and adenosine more important than reproducing the magnitude of the DMP angle differences. Our calculated ϑ(O-P-O) is very close to both the average of the angle for the *g*,*g* and *g*,*t* conformations as well as the x-ray crystallographic value. Thus we do not expect that large errors will result from the use of these average values.

We determined the stretching and bending force constants for the phosphate group by carrying out normal mode calculations on diethylphosphate and comparing these with results from a typical vibrational analysis calculation

by Brown and Peticolas[52] and with experiment (Table V). The fit between our calculations and the more complete force field calculations/ experiment is reasonable. The C-C stretching force constant of 260 kcal/mole $\AA^2$ is similar to that found by Karplus and Kuschick[54] on *n*-butane while the C-O K_r=310 kcal/mole $\AA^2$ fits the C-O stretching frequency in dimethylether[55] very well.

(3) CALCULATIONS ON DEOXYADENOSINE AND A MODEL DEOXYRIBOSE

When we turn to a more complex molecule such as deoxyadenosine or DNA in general, there are a large number of empirical parameters to be determined. The sources for the initial set of parameters for the nucleic acid bases adenine, guanine, cytosine, thymine and uracil were as follows. The bond length and angle equilibrium parameters for the nucleic acid bases were taken from x-ray structures. The torsional potentials were all assumed to be twofold, as in typical double or partially double bonded molecules. The interpolation approach (Table II) was used for determining the stretching force constants.

The sp^2 angle bending parameters remained undetermined and, in the absence of definitive values, we used analogies from our N-methylacetamide (NMA) normal mode calculations (described below). The default value of K_θ=70 kcal/mole rad^2 was used for the bases and this value was modified on the basis of normal mode calculations of NMA and benzene. For example, in NMA, the X-C=O bending parameter was 80 kcal/mole rad^2 and this value was used for such groupings in the nucleic acid bases. For NMA, the X-N-H value was somewhat smaller, 35 kcal/mole rad^2, causing us to use this X-N-H parameters in the bases. The Csp^2-Csp^2-Csp^2 value of 85 kcal/mole rad^2 came from the benzene normal mode calculations.

Some additional parameters needed to be determined for the sugar phosphate backbone of the nucleic acids. For all Csp^3-Csp^3-Csp^3 we used the same bending force constant as the value for C2-C2-C2 derived from THF. However, ϑ(X-CH-Y) (X,Y sp^3 carbons) should not be the same as ϑ(X-C2-Y), since propane is no longer an appropriate reference. The K_ϑ for Csp^3-OH-HO came from normal mode analysis calculations on methanol. The remaining parameters were taken from appropriate analogies (for example, we took all K_r and r_{eq} for Csp^3-Csp^3 from C2-C2).

The deoxyadenosine calculation was the first of these model calculations in which the electrostatic term has been included. The partial atomic charges employed here are listed in figure 1 in the appendix. Using a distance dependent dielectric constant, $\varepsilon=R_{ij}$, the C5'-O5' bond in the g,g range, and the H05'-O5'-C5'-C4' and C4'-C3'-O3'-HO3' in the *trans* ranges, complete energy refinement of deoxyadenosine was carried out. The torsional angle C4'-C3'-C2'-C1' was constrained to various values as a means of evaluating the energy as a function of sugar puckering. Only two local minima are found, with C2' *endo* and C3' *endo* conformations; the lowest barrier between them occurs in the O1' *endo* region. In good agreement with experiment, the C2' *endo* conformation is more stable than C3' *endo* by 0.6 kcal/mole, with the barrier between C2' *endo* and C3' *endo* being 1.3 kcal/mole (Table VI). The sugar pucker pseudorotation W values for the C3' *endo* and C2' *endo* conformations are 5° and 152° respectively. They occur near the middle of the range of observed values, although both are somewhat smaller than the center of the C3' *endo* and C2' *endo* ranges (18° and 162° respectively). The use of nonbonded parameters similar to Dunfield's[24] or Jorgensen's[25] significantly reduces the W value for the C2' *endo* minimum near the lower end of the

observed sugar pucker values (≈140°) and reduces the C2' *endo*/C3' *endo* energy difference to near 0.0 kcal/mole. This also occurs with a scale factor of 1.0 rather than 0.5 for 1-4 van der Waals interactions.

Olson[30] has recently carried out theoretical studies on a model for a deoxyribose ring, with substituents 1'-NH_2, 5'-CH_3 and 3'-OH, studying the pseudorotation profile of this model with a fixed out of plane q=0.38 Å and constrained bond lengths. The 1', 3', 5' substituents were treated as united atoms, but all H's on the ring were included in the calculation. Only endocyclic bond angles were energy refined at each W. We repeated these calculations (Table VI) with complete energy refinement using a standard bond stretching force constant of 300 kcal/mole Å^2. All bonds were assigned the Olson values for r_{eq}, while ϑ_{eq} (H-C-H) and ϑ_{eq}(C-C-H) were tetrahedral with a force constant of 40 kcal/mole rad^2. The agreement between the C2' *endo*/C3' *endo* energy difference and the W and q values for this model and our deoxyadenosine calculations was encouraging and supported the reasonableness of our calculated 1.3 kcal/mole C2' *endo* — C3' *endo* barrier, with a maximum near the O1' *endo* conformation. In the earlier study[30] and in our previous force field[16], a barrier of 2.0 kcal/mole had been found.

After our normal mode analysis of THF, we returned to the sugar conformational profile of deoxyadenosine. The results of that study are summarized in Table VII. The largest difference between FF1 (with sp^3 bending force constants ≈40-55 kcal/mole Å^2) and FF2 (bending force constants ≈60-100 kcal/mole Å^2) is found mainly in the C2' *endo*-O1' *endo* energy difference, which had been increased from 1.3 to 2.0 kcal/mole. FF2' differs from FF2 only in that the 1-4 electrostatic interactions have also been reduced by the

0.5 scale factor. This leads to a minimum energy W closer to the experimental value, but a somewhat smaller q (although still within the experimental range). We also studied conformations in the range of O1' *exo* ($W \approx 270°$). The barrier for the C2' *endo*--O1'*exo*--C3' *endo* transition was calculated to be 3.4 kcal/mole, with the minimum energy q=0.15 Å. The ring flattens considerably in this conformation, presumably to relieve the base---C5',O5' repulsions in the O1' *exo* conformation.

Model calculations on the ribonucleoside adenosine led to the same two local minima (C2' *endo* and C3' *endo*), with the O1' *endo* barrier in the range of 3.0 kcal/mole. However, the C2' *endo*/C3' *endo* energy difference depended on the electrostatic energy and the orientation of the 2'OH, which began for each refinement in a conformation O2'-HO2' eclipsing the C3'-C2' bond. Since our distance dependent dielectric model, $\varepsilon = R_{ij}$, still allows for rather strong intramolecular electrostatic interactions compared to what would occur for adenosine in aqueous solution (where presumably all the H-bonding sites would be occupied by H_2O molecules) we examined the effect of using a larger effective dielectric constant, $\varepsilon=4R_{ij}$. This is calculation labeled FF2'' in Table VII, in which the agreement with both experimental structures and energy differences is quite satisfactorily represented for deoxyadenosine and riboadenosine.

One of the major points in the Olson[30] work was the fact that sugar pseudorotations (between C2' *endo* and C3' *endo* conformations in furanose rings) was not "nearly free[44]" but required surmounting a 2.0 kcal/mole barrier at the O1' *endo* conformation. Our more complete refinement using the Olson parameters led to a barrier of 1.3 kcal/mole, both for her model and deoxy-

adenosine. Changing to larger K_ϕ, V_n raised the barrier to 2.0 kcal/mole, the value also found with our old force field[16]. Thus, the best available theoretical estimates for this barrier suggest it to be in the range of 1.3-2.0 kcal/mole.

(4) CALCULATIONS ON BASE PAIRING, STACKING AND SEQUENCE DEPENDENT STABILITIES

Recently, gas phase experiments have determined interaction energies for nucleic acid bases giving both hydrogen bonding and stacking energies. To provide a check of our van der Waals parameters for sp^2 atoms and on our method for deriving partial charges for atoms, we model built and energy refined Watson-Crick H-bonded complexes between 9-methylguanine and 1-methylcytosine, 9-methyladenine and 1-methylthymine and Hoogsteen base paired models between 9-methyladenine and 1-methylthymine. Stacking between two 1,3-dimethyluracil molecules were also studied. Such calculations were carried out with dielectric models ε=1 and $\varepsilon = R_{ij}$ and are compared with the more elaborate calculations by Langlet *et al*[58] and the gas phase mass spectrometric experiments of Yanson *et al*[59] (Table VIII). The agreement between the calculated and experimental values for H-bonding with either dielectric model is quite good; and our calculations, with $\varepsilon=R_{ij}$, agree with the findings of Langlet *et al* that the Hoogsteen base pairing for AT is better than Watson-Crick.

In the 1,3-dimethyluracil stacking calculations, Langlet *et al* have noted that one must extrapolate the experimental data to zero field (which was not done by Yanson *et al*, who found a base stacking enthalpy of -3.6 kcal/mole). Since the Langlet *et al* calculations find a similar enthalpy at the electric

field used by Yanson *et al*, we feel that their calculated values at zero field (ΔH=-9.1 kcal/mole) are a good estimate for the "true" experimental enthalpy. Our calculations are in satisfactory agreement with their results with a distance between base planes of 3.43 Å (In our previous force field, the minimum energy distance is calculated to be 3.30 Å). We also carried out such calculations in the all atom representation with the van der Waals parameters exactly as used by Hagler *et al*[19]; these results were similar to those we found in the united atom representation with our van der Waals parameters.

In a previous study of dinucleoside phosphates[16], we had found that observed sequence dependent stabilities in DNA melting could be rationalized with a simple dinucleoside model, leading us to carry out complete energy refinements on the ten base paired dinucleoside phosphates studied earlier with our previous force field (Table IX). The relative calculated energies (especially with $\varepsilon = R_{ij}$) are a significant improvement over the previous model calculations in comparing the homo and hetero polymers with a mononucleotide or dinucleotide repeat. In particular, the new force field qualitatively reproduces the relative melting temperature independent of dielectric model whereas the old force field was only successful in doing this with ε=1. In addition, the relative magnitudes of the energy differences with the new force field, $\varepsilon=R_{ij}$, are in the same order as the magnitudes of the Δt_m. Neither the old nor the new force field had been able to successfully calculate the relative energies for the trinucleotide repeat models, but these may require calculations on tetra, rather than dinucleosides.

The "improvement" of our new calculated relative energies over the previous ones is encouraging, but we must stress that the relation between our calculated relative energies and the experimental relative melting temperature is very indirect. First, we are assuming that the melting temperature differences are due mainly to differences in the energies of the double stranded forms. Second, we are assuming that our energy refined geometries for the base paired dinucleoside phosphates here and in ref. 16 are good representations of the double stranded geometries in longer DNA double helicies. Although we have concluded that such approximations are likely to be reasonable, we cannot prove this.

PROTEIN TEST CASES

(1) DETERMINATION OF THE PEPTIDE BACKBONE PARAMETERS

The first goal in developing the protein segment of our force field was to derive a consistent set of charges for the hydrogen, nitrogen, carbon and oxygen atoms contained in the amide segment of the peptide chain. Since it would be impractical to generate charges for various sizes of oligopeptide strands possessing all side chain combinations, it was necessary for us to select a model system which we felt could best represent the backbone. Our choice consisted of the dipeptides N-acetyl-N'-methyl glycyl amide and N-acetyl-N'-methyl alanyl amide for which numerous theoretical studies have previously been performed (Maigret *et al*[61], Schafer *et al*[62]). We hoped to develop a force field which best reproduced structures and energies of the dipeptides for the local minima of PCILO and all atom molecular mechanical calculations.

We initially generated the electrostatic potential surface at the 6-31G level for N-methylacetamide and fit the surface to a twelve point charge model, where all hydrogens were included. It must be pointed out that, at the time, our limited disk storage space made it impossible for us to calculate an initial charge set at the 6-31G basis level for even a glycyl dipeptide; hence the rationale for our choice of NMA.

Due to the inherent dependence of derived charges upon basis set, we felt it paramount to employ the most accurate charge model. Cox and Williams[28] found, in a study on small molecules, that electrostatic potential derived charges from a 6-31G basis set differed from the "optimal" 6-31G** by a ratio of 0.82:1. They also noted that a scaling of 0.91 is needed to adjust the 6-31G** calculated dipole moments to fit the experimental values. To be most consistent within this framework, we decided on a scaling factor of 0.75 (0.82 x 0.91) for our 6-31G derived NMA partial charges.

The bond and angle parameters for N-methylacetamide were taken from crystallographic data of the peptide linkage presented by Benedetti[28]. The twofold barrier to rotation about the C-N bond was $V_2/2$=10.0 kcal/mole (ref 3). To best represent the experimental energy difference of 2.1 kcal/mole between the *cis* and *trans* conformations of NMA[31], we included $V_1/2$=0.65 kcal/mole. Using this updated parameter set we energy refined the all atom glycyl and alanyl dipeptides using the 0.75 scaled charges. The dipeptides were constrained about their Φ and Ψ angles in 60° intervals and an energy map of these 36 regions was constructed. Each low energy area was searched for a local minimum by relaxing all degrees of freedom and further refining. Finally, "true" local minima were confirmed by using the Newton-

Raphson second derivative routine[43].

The next step was to best "fit" both the structures and energies obtained above for our ***all atom*** model dipeptides to their respective *united atom* representations. This time, our charges were derived by fitting the electrostatic potential points to a ***six*** point model of NMA. The structures were subjected to an identical grid search and local minima analysis as done above.

In the ***all atom*** model for alanyl dipeptide we located five local minima. However, if the 1-4 electrostatic energies were not scaled down in the *united* atom case, only three "true" local minima were found. Even with the initial Φ and Ψ torsions constrained to "force" the dipeptide into these regions, the extended structure and the highest energy conformation were not local minima along the potential energy surface. For this reason, and from results mentioned above on adenosine, we decided to empirically scale the 1-4 electrostatic interactions by a factor of 0.5. Unlike the united atom model, the all atom relative energies were not very sensitive to the inclusion of a scale factor for both 1-4 van der Waals and electrostatic energies.

The alanyl local minima geometries and energies were then recalculated with the appropriate scale factor for both the united and all atom models. For the alanyl dipeptide, we located five local minima. The two lowest in energy correspond to 1-7 hydrogen bonded structures (where the right handed system is 0.9 kcal/mole more stable than the left handed). An extended structure, forming 1-5 H-bonds, lies 3.2 kcal/mole above the global minimum. The right and left handed helicies, corresponding to 1-10 H-bonded conformations, occur at 3.6 and 4.5 kcal/mole above the most stable structure.

Five local minima are found for the glycyl dipeptide. Two isoenergetic global minima occur, as nonsuperimposable mirror images of themselves, for the 1-7 hydrogen bonded structures. Lying 3.1 kcal/mole above these systems is the extended conformation $\Phi=180°$ and $\Psi=180°$, with two more isoenergetic helical structures at 4.0 kcal/mole.

It was found that by scaling the 6-31G united atom charges by 0.81, we achieved good agreement for both the alanyl and glycyl dipeptides, with an average error less than 0.5 kcal/mole between the united and all atom representations. The full results appear in Table X. We have shown that an empirical scale factor of 0.5 in the 1-4 electrostatics is necessary to best fit the local minima structures and energies for alanyl dipeptide.

To best illustrate the low energy regions corresponding to these local minima structures, we constructed Φ, Ψ maps consisting of 1296 points derived from energy minimization with all geometric degrees of freedom relaxed (figure 2). The alanyl dipeptide maps of Ramachandran and Sasisekharan[63] and Brant *et al*[64] are qualitatively quite similar to our plot in the left half of the map (Φ between -180° and 0°). However, their plots fail to locate a low energy contour in the right half region centered about Φ = 60° and ranging from $\Psi\approx-90°$ to $\Psi\approx60°$. The appearance of this region is a manifestation of our methodology for generating the structures; in which we allowed all geometric degrees of freedom to relax during the minimization process. The occurrence of additional regions, in refinements employing relaxed geometries compared with constrained minimizations, has been shown by Gibson and Scheraga[65] in the alanyl dipeptide and Gelin Karplus[66] in a study on β-methylacetylcholine. Furthermore, infra-red data by Cung *et al*[67] and,

Avignon and Lascombe[68] support the existence of both the *axial* and *equatorial* C_7 structures ($\Phi\approx-80°$, $\Psi\approx70°$ and $\Phi\approx70°$, $\Psi\approx-60°$ respectively) for alanyl dipeptide. The 1-13 stabilizing hydrogen bonds which would be formed in the left handed alpha helical structure ($\Phi\approx60°$, $\Psi\approx60°$) cannot exist in a simple dipeptide model, causing this region to be 5 kcal/mole higher in energy in our map, compared to the global minimum. Thus, we would expect this area to become more stable relative to the right handed 1-7 system as the dipeptide model is extended to a tetrapeptide structure and we are currently addressing this question. Additionally, our map failed to exhibit the ubiquitous low energy "finger" contour extending down from the upper right quadrant. This left handed "finger" region ($\Phi\approx60°$, $90° < \Psi < 180°$) is present as a low energy region (< 5 kcal/mole) in the other Φ,Ψ maps[63,64,65] but is 7-8 kcal/mole higher in ours. We should point out that Richardson[69] has analyzed the conformation of 1000 non-glycine residues in globular proteins and has found none in this region.

Upon the implementation of software within AMBER[21,43] designed for generating a complete vibrational normal modes analysis, it became possible to "fine tune" both stretching and bending force constants for small molecules. The goal was to obtain a good fit of calculated normal mode frequencies with the best experimental data available for N-methylacetamide. Our initial starting parameter set consisted of the Benedetti structural terms, one and twofold rotational values as mentioned above and our 6-31G NMA charges scaled by the appropriate value of 0.81. The stretching and bending force constants were assigned standard default values taken from our original parameter set (e.g. 50 kcal/mole rad^2 for bending terms).

The methodology for our normal mode analysis employed an iterative process where initial calculations were run on an all atom model. Results from this simulation gave us insight into which modes were most coupled to hydrogen motions. The remaining non hydrogen modes were the ones used in our analysis of NMA.

A united atom model for NMA possesses six atomic centers which generate three out of plane and nine in plane motions. The out of plane bending modes are highly dependent upon the improper torsional parameters. With X-X-N-H=1.0 kcal/mole and X-X-C-O=10.5 kcal/mole, all three out of plane experimental frequencies were fit with an average error less than 4.0 cm^{-1} (see Table XI).

The highest frequency normal mode is almost entirely due to N-H stretching and was calculated by adjusting its respective force constant. The next two high energy modes, amide I and amide II, are strongly mixed with carbonyl stretches and were derived accordingly, giving us C=O K_r=570 kcal/mole $\AA^2$. Since the remaining in plane modes are highly coupled to each other, the six bond bending force constants were varied in a cyclic fashion until a reasonable fit with experiment was achieved (see Table XI). From our results obtained on glycyl dipeptide, alanyl dipeptide and N-methylacetamide, we now possessed a complete parameter set for the amide linkage segment of the protein backbone.

(2) CHARGE DERIVATION

At this point in the force field development we possessed a reasonable set of charges for the atoms of the peptide *backbone*. The next step was to gen-

erate a set of electrostatic potential derived charges for the side chains of peptides. The protein residues were broken down into two structural units, the ***bridge*** containing both the alpha and beta carbons (where applicable) and the ***chromophore*** possessing the remaining side chain atoms. Charges were computed at the STO-3G level for representative molecules of hydrogen bonding peptide side chains. For example, phenol and imidazole were used as the chromophores for tyrosine and histidine respectively. Again, due to basis set dependence of the calculated charges, we wished to represent the charge distribution in a manner consistent with the results of Cox and Williams[26]. Their study found that the STO-3G basis set derived charges could be best fit to the "optimal" 6-31G** by a ratio of 1.12:1. Due to this fact, coupled along with the 0.91 scaling needed to bring the 6-31G** charges in line with experimental dipoles, we decided upon a scaling factor of 1.0 (1.12 x 0.91) for all STO-3G derived charges.

The AMBER protein data base consists of the twenty amino acids, plus histidine protonated at both the delta and epsilon positions and a special residue for forming disulfide linkages. Each alpha and beta carbon, or bridge atom, can be thought of as a buffer for absorbing the remaining charge distribution from the chromophoric and backbone segments, needed to achieve neutrality or an ionic state. The overall charge for the backbone atoms is -0.246, while the chromophores are neutral, singly protonated or singly anionic. This excess charge was ratioed between the two carbon atoms by the same proportion as existed in our previous force field[16,17], where the charges came from Mulliken populations. The rationale for piecing together segments of molecules to form an overall partial charge distribution for larger molecular units appears above. For the special case of hydrocarbon

side chains where hydrogen bonding is not important, (e.g. alanine, valine, leucine, isoleucine and phenalalanine) no chromophore was used and the 0.246 was ratioed in direct accord with the Mulliken populations found in the previous force field.

As a final test for our charge selection, we examined water/chromophore hydrogen bonding systems for bonding energy, overall structure and hydrogen bonding distance. All molecular mechanical calculations were carried out with a distance dependent dielectric, $\varepsilon = R_{ij}$. The water charges were chosen to give us reasonable results for stabilization energy and H---O distance for the H_2O dimer system. Using the partial charges oxygen=-0.66 and hydrogen =0.33, we calculated a water dimer H---O distance of 1.79 Å and ΔE=-6.4 kcal/mole, in qualitative agreement with experimental values[72].

Next, we employed these water charges, with the electrostatic potential derived ones from the *chromophores* for the remaining dimer calculations (Table XII). The 10-12 hydrogen bonding term in our potential energy function allowed us another degree of freedom for "fine tuning" the hydrogen bond distance. We found that a standard well depth of 0.5 kcal/mole and an equilibrium distance of 1.95 Å produced hydrogen/heteroatom distances of the magnitude 1.80 Å (see Table XII). However, r_{eq}=1.95 Å produced a too long and not strong enough hydrogen bond for ammonium/water (which was selected as a model for our charged chromophores). We found that a more reasonable distance, 1.66 Å, could be achieved with a r_{eq}=1.85 Å and employed this value for all cationic and anionic/water interactions.

Finally, we addressed the more common hydrogen bonding situation in protein environments, with carbonyl acting as the proton acceptor and N-H as

the donor. For the model system N-methylacetamide dimer, we calculated an H-bond distance of 1.82 Å and a stabilization energy ΔE=-7.0 kcal/mole. As is shown in Table XII, rather similar H-bond energies are achieved, with the electrostatic potential derived charges, for both calculations where $\varepsilon=1$ and $\varepsilon=R_{ij}$. We used STO-3G basis set derived charges for all *chromophores* with the exception of two special cases where slight alterations in the charges were necessary to give good agreement with quantum mechanical calculations.

Our first special case involves the methanol/water hydrogen bonding model, which we used as our representation of both serine and threonine hydrogen bonding interactions. This system shows a direct basis set dependence upon overall stabilization energies for the two possible structures CH_3OH---OH_2 and HOH---$OHCH_3$. Work done by Del Bene[74] using an STO-3G basis set revealed that methanol as a proton donor was energetically more stable than water, while theoretical results by Tse *et al*[75], employing a 6-31G* basis set, showed methanol to be a better proton acceptor by 0.20 kcal/mole. To be consistent within our framework of using data from theoretical calculations employing "optimal" basis sets (6-31G*, 6-31G**), our goal was to best reproduce the relative energies found by Tse *et al*. We found that good agreement with quantum mechanical theory could be achieved by adding -0.058 units of charge to the oxygen, while placing the residual on the carbon, for an overall distribution of O=-0.550, H=0.310 and C=0.240. Aside from a reasonable correlation with theory, the altered charges imply a dipole moment of 1.81 D for methanol, as compared with the experimental value of 1.70 D.

The second special case which we considered was the energy and geometry of hydrogen bonded systems involving sulfur, which are of relevance in developing charges for methionine, cysteine and cystine. Our first set of model calculations were on the the H-bonded complexes HSH---OH_2 and H_2S---HOH. Specifically, we were interested in the angle above the plane (AAP) formed between the bisector of the sulfur hydrogens and the vector formed with the water hydrogen (involved in the H-bond) and the sulfur in the H_2S ---HOH complex. Theoretical calculations by Kollman *et al*[80] suggest that dimers formed by second row hydrides and H_2O possess much greater AAP than the corresponding first row hydrides. Reoptimizing this H_2S---HOH complex with a 4-31G basis set, we calculated an AAP of 78° and $\Delta E_{stabilization}$=-3.88 kcal/mole for the H_2S---HOH "linear" structure. With atom centered partial charges, the AMBER local minimum corresponds to a nearly "bifurcated" structure possessing an AAP of 15°. With a 4-31G basis set and the AMBER geometry, we calculated an energy 1.40 kcal/mole higher than the lowest energetic quantum mechanical structure. To place these values for second row hydride electron donors into a proper perspective, we should note that Umeyama and Morokuma[81] found *only* a 0.5 kcal/mole difference in energy, $\Delta E_{stabilization}$=-7.8 kcal/mole, for H_2O---HOH with AAP=45°, compared to the geometry possessing an AAP=0°. We concluded that to achieve reasonable qualitative agreement with the quantum mechanically calculated structures for second row elements, it was necessary to place explicit lone pairs on all sulfur atom types in our new force field. Our goal was to best fit the structures and energies of the water/H_2S system to theoretical calculations carried out with a 4-31G basis set and then to extrapolate these results to determine charges for sulfur containing amino acids.

The quantum mechanical electrostatic potential was calculated for the five atom centered model at the STO-3G level. To keep the lone pairs from "fusing" into the sulfur, we froze the sulfur charge and optimized the lone pair distances using a four point model. In this case, we found we had to vary both the charges and the 10-12 parameters to achieve optimal agreement with quantum mechanical theory. Our final molecular mechanical calculated structure had an AAP of 64° and possessed a quantum mechanical energy only 0.16 kcal/mole above the 4-31G optimized structure.

Proceeding with the concept of explicitly including lone pairs on sulfur atoms, we selected methylsulfide as our charge model for cysteine residues. An initial charge set was generated by fitting the quantum mechanical electrostatic potential at the STO-3G level. Our degrees of freedom consisted of the 10-12 H-bonded parameters, the partial charges and the LP-S-LP angle. We found that with R^*=3.0 Å and ε=0.1 kcal/mole and by empirically placing the lone pairs at ϑ(LP-S-LP)=160° about the sulfur we calculated an AAP of 79° and an O---S distance of 3.37 Å for CH_3HS---HOH. Consistent with the Kollman *et al*[80] results on hydrogen sulfide, we felt it necessary to fit the relative energies for H_2S as a proton donor and acceptor. To achieve very nearly isoenergetic states for this system, we varied the charges by adding 0.117 to hydrogen, -0.100 to the lone pairs, 0.090 on the sulfur and finally -0.007 to the carbon for neutrality. The final results giving $\Delta E_{stabilization}$=-3.1 kcal/mole for CH_3HS---HOH and $\Delta E_{stabilization}$=-3.2 kcal/mole for CH_3SH---OH_2. These methylsulfide charges are actually quite reasonable as they bring the dipole moment into a better agreement with the experimental value of 1.51 D. The new charges lead to a dipole moment of 1.82 D compared with 1.02 D if we used those directly fit to the electrostatic potential. Hence, we have

shown that empirically placing the lone pairs at 160° about the sulfur gives the proper directionality needed to be consistent with quantum mechanical results.

(3) INSULIN REFINEMENT

Thus far, we have parameterized the force field for small molecular subunits with the ultimate goal of applying the new parameters, with both generality and transferability, to larger systems of nucleic acids and proteins. Within the context of proteins, we have shown that the force field produced quite reasonable results for hydrogen bonding structures and energies of water/chromophore interactions, local energy minima of alanyl and glycyl dipeptides and vibrational frequencies of N-methylacetamide. As the first direct application of the new force field to an entire protein, we used our new parameter set to energy refine insulin by conjugate gradient optimization. We selected insulin for two reasons: first, its relatively small size, 500 atoms (404 crystallographically located heavy atoms and 96 added hydrogens), enabled us to refine the energy of the entire protein within a reasonable time frame. Second, due to the high resolution of the crystal structure (resolved to 1.5 Å), any large scale motions within the minimized structure, relative to the starting one, would reveal areas where reparameterization of our force field might be necessary. Specifically, we were interested in "quality" of the intramolecular hydrogen bonds, the overall degree of inward compactness of the entire minimized protein and the general question of how a distance cutoff for evaluating the non-bonded interactions effects the final structure.

The starting coordinates for porcine insulin by Dodson *et al*[82] were taken from the Brookhaven Protein Data Bank[83]. A subroutine within AMBER placed

all potential H-bonding hydrogens on each respective nitrogen, oxygen and sulfur with standard bond lengths and angles. Our first minimization employed a distance dependent dielectric, $\varepsilon=R_{ij}$, and a 9.0 Å cutoff distance for evaluating the non-bonded interactions. All non-bonded interactions between 8.0 Å and 9.0 Å were multiplied by a cubic equation whose value varies from 1 at 8.0 Å to 0 at 9.0 Å. Consistent with our rationale explained above, all 1-4 van der Waals and electrostatic interactions were scaled by 0.5.

After 1318 energy evaluations, the rms energy gradient reached 0.14 kcal/Å. An rms fit of the initial and final structures was 0.28 Å for the backbone atoms and 0.43 Å for all insulin atoms. Using the radius of gyration $((\sum_{i=1}^{n} R_i^2/n)^{1/2}$, where R_i is the distance of atom i to the center of mass and n is the number of atoms), we generated a sphere to represent the "volume" of insulin. Although many elaborate methods exist for generating the volume of a protein, a simple spherical representation will suffice to give a feel for the extent of inward compactness of insulin. In this specific case, we calculated an overall volume compaction of only 1.5%. This value is most likely an overestimation since our model does not explicitly include solvation which would tend to "pull" solvent facing side chains and backbone atoms outward from the center of mass.

The second test run used the identical criterion as our first with the exception that a nonbonded cutoff of 12.0 Å was used. It should be pointed out that since the calculation of the van der Waals interactions is the rate limiting computational step in all energy minimizations, it is important to find the smallest non-bonded cutoff which still gives reasonable results. After 3937 energy evaluations the rms gradient had reached 0.09 kcal/Å. It appears,

from calculations on insulin and our unpublished results on papain, that the overall number of energy evaluations and the rms gradient (after the system has reached a minimum by conjugate gradient techniques) is a function of nonbonded cutoff. We are currently addressing this question as a means of assessing the best methodology to undertake for subsequent energy refinements of proteins and nucleic acids.

We carried out five different refinements on insulin with the results appearing in Table XIII. It is interesting to note that nearly identical results are attained whether we use Jorgensen's or our new nonbonded parameters for the united atom carbons. Finally, the most encouraging result, from the standpoint of displaying the improvement of the new force field over the old one, appears for the refinement using our old parameter set (which possessed similar R*'s but smaller ε's for the nonbonded terms, as compared to the new force field) where we calculated 16% compaction for the backbone atoms and 1.01 Å rms movement for all insulin atoms.

Hartman *et al*[84] have shown that there is a definite decrease in the volume of the crystal structure of myoglobin at 80 K compared with 300 K. Our insulin energy refinements are equivalent to simulations carried out at 0 K, hence, the compaction which we observe is consistent within this framework. However, in view of the lack of inclusion of explicit water molecules or crystal symmetry in our minimization, more precise comparisons are not possible at this time. The nature and extent of protein compaction expected from an energy refinement is still unclear, and we are analyzing this question as a means of obtaining a greater understanding of the forces involved in protein structure.

DISCUSSION

The approach we have taken to developing the force field presented here has several unique aspects, but, in general, the various steps have precedent in the literature. The unique aspect in this case is that this is the first complete "consistent" force field developed for both proteins and nucleic acids. However, force fields are constantly evolving objects, and, although we may have reached a plateau, we anticipate further development in the future. Below we critically analyze the results of our study: intramolecular parameters (bond stretching, bending and torsion), nonbonded parameters (Lennard-Jones) and electrostatic charges.

We have used spectroscopic structural data to determine the K_r, r_{eq}, K_ϑ and ϑ_{eq} parameters for use in our basic energy expression (equation 1). After energy refinement of model systems, the calculated bond lengths, because of the generally large K_r values and the fact that our molecules are relatively unstrained, remain very near r_{eq}. This is also true for the bond angles for non cyclic parts of the structure. We note that our method of interpolating K_r for sp^2 atoms in planar rings, possessing bond lengths between pure single and double bond values, works adequately.

In the furanose ring or tetrahydrofuran, it is clear that the value of K_ϑ's derived from fit to energies and structure (40 -50 kcal/mole rad^2 here and in the MM2 force field) are significantly smaller than the values (60-100 kcal/mole rad^2) needed to reproduce angle bending frequencies in both cyclic and non-cyclic systems with an sp^3 carbon or oxygen at the apex of the bond. In order to reasonably fit frequencies, energy and the structure of THF, we needed to increase the torsional parameters V_φ for rotation around

the C-O and C-C bond. When we critically compare the results of calculations using these two choices of K_ϑ and V_φ in fitting the structures, energies and frequencies of THF and MEE, we find little to choose from, with both approaches giving fair agreement to experiment. However, since we have used vibrational frequencies of amides and other model systems to determine many K_ϑ and K_r, we slightly favor the parameter set containing the larger K_ϑ, V_φ values. It should be emphasized, however, that it is a simple matter to employ the smaller parameters by using the set noted in Table III(FF1).

The torsional parameters we have used come mainly from literature values for proteins and our previous nucleic acid force field. Our use of V_2 values for O-P-O-C, O-C-C-O and O-C-C-C groups allow us to "fine tune" conformational energy differences. We have used an interpolation method for V_2 values for partially double bonded sp^2 atom---sp^2 atom parameters which appears to give good agreement with experimental nucleic acid base frequencies[56]. A purely empirical fit of the out of plane "improper" V_2 torsional parameters for the C=O and N-H groups in amides has resulted in an excellent correlation to the out of plane experimental frequencies.

There are two straightforward ways in which these intramolecular aspects of our force field can be improved; the first is simply to abandon the united atom description of C-H groups, and the second is to use a more elaborate description of the intramolecular energy, including many more coupling and anharmonic terms in the energy expression. Our comparison of all atom vibrational calculations with the united atom ones suggest that we can adequately represent low frequency (< 700 cm^{-1}). However, extensive coupling

between various hydrogen including modes in the 800-1500 cm^{-1} precludes a more precise representation of these modes in this study.

The use of a more elaborate energy function for intramolecular interactions has much precedence in the literature[44], and, if our main goal was a more precise description of all the vibrational properties of macromolecules, we should certainly switch immediately to a more complex function. However, we have evidence that the low frequency vibrations which contribute most to thermodynamic properties are relatively insensitive to the inclusion of anharmonic bond/angle and coupling terms[85]. In addition, the analysis of the vibrational spectra and its fit to many more empirical parameters would need to be done for the large number of atom types in proteins and nucleic acids. This is a major undertaking and worth doing (we are currently engaged in such an analysis, still using equation 1, for the four nucleic acid bases)[86] but the other issues, discussed below, are much more critical for the development and understanding of the structures and energies of a macromolecular system.

One of the most difficult problems in the development of this force field was the choice of nonbonded parameters and the way to handle them for 1-4 interactions. For sp^2 atoms, we used values very similar to those of Hagler *et al*[19] and our model calculations on 1,3-dimethyluracil stacking (Table III) suggest that our stacking energy values are very similar to those calculated using the actual Hagler *et al* parameters. The base-base separation, interestingly enough, is somewhat larger (3.71 Å) with the Hagler *et al* values than with ours (3.43 Å), but this likely reflects the fact that we explicitly include the C-H hydrogens in the former case. However, this difference in

structure may be significant in its relevance to our choice of nonbonded parameters for sp^3 atoms. Use of nonbonded parameters, as have been determined to give good crystal packing parameters and energies for hydrocarbons by Dunfield *et al*[24], and a very similar set of values which gives good agreement with liquid state energies and densities for ethers[25], results in a much poorer representation of sugar conformational properties than if smaller values are used in our force field. Above, we have argued why the use of a scale factor (completely empirical) for the 1-4 nonbonded interactions makes some sense, particularly because a 6-12 function would be too repulsive for shorter nonbonded interactions, compared to the more realistic 6-exponential form.

A 1-4 nonbonded scale factor also allows us to use van der Waals radii for CH, C2 and C3 atoms somewhat closer to Jorgensen's[25]. Above we have shown that such parameters for C3 lead to 10% errors in energy and density, compared to errors of 3-5% found by Jorgensen. However, since proteins and nucleic acids contain a considerable fraction of sp^2 atoms, our van der Waals parameters for CH, C2 and C3 should not lead to very large errors. The refinement of insulin suggests that our nonbonded parameters are reasonable.

It is possible that all these "problems" could be simply solved by abandoning the united atom representation for CH, C2, C3 (and corresponding sp^2) groups. However, one should not do this "lightly", given the sensitivity of simulations on large molecules to nonbonded cutoffs (above discussions on insulin and our unpublished refinements of papain) and the fact that nonbonded function evaluation is the rate limiting step in such simulations. (C-H

hydrogens can make up to 50% of the atoms in proteins and 30% for nucleic acids.) However, such a step may be appropriate in some cases (our simulations on thyroxine and its protein binding)[17] and it would not be difficult to extend our current force field in such a way, since all of our calculations for partial charges have been done both in the united and all atom representations. Switching the nonbonded parameters to use a 6-exp rather than 6-12 functional form is simple enough to do, but also would slow the most time consuming part of the calculation. However, as computing power increases, this also seems a likely refinement for the near future.

One of the most useful results from this study has been the generalization of the initial studies by Momany *et al*[87], Smit *et al*[88] and Cox and Williams[26] to using electrostatic potentials for determining the appropriate atomic partial charges used in evaluating the electrostatic term in equation 1. Together with 10-12 parameters, these charges lead to H-bond energies and structures in reasonable agreement with available *ab initio* calculations and experiments. The two cases where the charges were altered from the electrostatic potential determined values are instructive in this regard: polarization effects clearly play a role in whether CH_3OH---OH_2 or HOH---$OHCH_3$ is the lower energy structure, and these effects are also of greater importance for the hydrogen bonding involving sulfur, rather than oxygen, as an electron donor. Thus, our simple representation of H-bonding in equation 1 is clearly an area for future improvement of the force field. Ultimately, a more complete description of the charge distribution and its effect on hydrogen bonding can be considered by a modification of equation 1 (this is subject to the time consuming nature of evaluating nonbonded interactions).

We have developed this force field using a distance dependent dielectric constant ($\varepsilon=R_{ij}$) since such an approach is a way to qualitatively simulate the fact that the system is in water and the intramolecular electrostatic interactions should die off more rapidly with distance than in the gas phase. However, we have shown for H-bonding in nucleic acid and protein models that the use of $\varepsilon=1$ gives quite similar H-bond energies and structures as that with $\varepsilon=R_{ij}$ Thus, our force field may also be well suited to simulations with explicit inclusion of water and we are currently developing approaches to do just that. However, we emphasize that the use of larger distance dependence ($\varepsilon=4R_{ij}$) was required to reproduce the relative C2' *endo*/C3' *endo* energy in adenosine, where the intramolecular electrostatic effects would be expected to be much more damped in solution, than for example, stacked base pairs. Also, the decision (purely empirical) to scale the 1-4 electrostatic interactions might be alleviated by the more complex electrostatic energy function, but it is not completely clear how to proceed along such lines. We emphasize that the single most crude aspect in the application of the force field is the way solvation effects are modeled and this is the area which deserves the most effort for refinement in the near future.

At this point, it is worthwhile to make a brief comparison of our parameter set with other force fields. First we will consider DNA. Elsewhere, we have shown that our electrostatic charges for the nucleic acid bases gave more accurate gas phase energies than others[18]. Our nucleic acid backbone charges are somewhat larger in magnitude but qualitatively similar to others in the literature[9,11,16]. For simulations without explicit inclusion of water, it is not clear that our more accurately determined charges will be an advantage over the others but, if water is included, they should be able to give an

accurate representation of relative water-water and water-solute interactions. The question of the magnitude of the phosphate charge is appropriate to note here, since some other force fields[9,89] use less than a unit negative charge. Again, without explicit counter-ions in the calculation, this seems reasonable, but we have shown that complete neutralization appears too extreme an approximation[90]. A correct representation of intra- and interstrand phosphate repulsion is important and, at this point, it is not clear how best to handle this, since the net charges and the dielectric constant used in equation 1 are so interdependent. Further work is needed to sort out this point, since it may be that a range of net phosphate charges and effective dielectric constants would be capable of reasonably representing the hydrodynamics as well as the local conformational energies of nucleic acids[10,91].

The van der Waals parameters employed here are similar to those in the recent literature[10], in that the radius is typically ≈ 0.2Å larger than the standard crystallographic van der Waals radius.

Our intramolecular force field (FF2, Table III) has larger K_ϑ, V_φ than earlier force fields because of our desire to better fit bending vibrational frequencies. However, sugar puckering profiles and conformational energies are similar to both FF1 (Table III) and other force fields. The take home message here is that all molecular mechanics studies on deoxyfuranoses have led to two local minima structures (C2' *endo* and C3' *endo*), with a smaller O1' *endo* than O1' *exo* barrier connecting them. All the calculations, except the Levitt and Warshel[92], suggest an O1' barrier of 1.3-2.0 kcal/mole, significantly larger than thermal energies. It appears that the reason for the discrepancy is merely an inappropriate choice for ϑ_{eq}(C-O-C) in the Levitt and Warshel

paper (where ϑ_0=120°). However, FF1, FF2 and the Olson study[30] found C2' *endo* more stable, whereas our earlier force field (FF0)[16] and that of Hingerty and Broyde[89] found C3' *endo* more stable. NMR data supports the greater stability of C2' *endo*. Olson attributed C2' *endo* stabilization to the prescence of an O-C-C-O *gauche* torsional term. The fact that even the force fields which find C3' *endo* more stable contain a *gauche* O-C-C-O term, suggests that the reason why deoxyfuranose prefer C2' *endo* to C3' *endo* may be more subtle.

We disagree with the suggestion by Olson[93] that force field parameters appropriate for proteins and hydrocarbons are not *per se* transferable to nucleic acids. Her argument was based on two facts. First, she pointed out the inadequacy of very early protein van der Waals parameters in representing base stacking. As we have demonstrated here, the Hagler *et al*[19] amide parameters do not have this flaw. Second, she cited the poor representation of the O1' *endo* sugar puckering barrier by Levitt and Warshel[92]. As noted above, this was probably due to an incorrect choice of parameters. Of course, one could imagine that "fine tuning" parameters separately for proteins or nucleic acids might lead to quantitatively better individual force fields then compromise efforts like this one, but we suggest that there will be no major flaws.

At this point, it is worth briefly comparing our force field with some of the other protein force fields in the liteature. It is difficult to compare parameters in detail due to the different methodologies employed in the application of the force fields. For example, our force field uses united atoms only for C-H's but considers complete energy refinement, Gelin and Karplus[6] use

united atoms for all hydrogens with complete energy refinement, and Momany's ECEPP[3] uses rigid bond lengths and angles but includes all hydrogen explicitly. Although there are some non-trivial differences in the relevant bond length, angle and dihedral parameters, it appears that these will not result in great differences in predicted conformational energies and structures. To our knowledge, ours is the first presentation in the literature in which the derivation of K_r, R_{eq}, K_ϑ, ϑ_{eq}, V_n and γ have been given in detail, although Momany *et al*[3] have analyzed in some detail their derivation of V_n and γ[3]. Although our use of vibrational calculations and a scaling algorithm has enabled a reasonably consistent set of parameters, we have noted the inherent limitations of the united atom (C-H) approximation and the simple harmonic potential function (eqn 1) in deriving more quantitative agreement with experimental frequencies.

The most important difference in the force fields resides in the nonbonded (electrostatic and van der Waals) parameters. It is likely that our electrostatic potential derived charges are a more accurate representation of the nature of electrostatic interactions than the CNDO/2 Mulliken charges used in ref. 3 and 6, thus allowing smaller well depths for our 10-12 parameters than theirs. Our van der Waals radii are similar to those of Gelin and Karplus[6], but our well depths are somewhat smaller. As discussed above, our van der Waals radii are smaller than those in the CH united atom force field of Dunfield[25]. Our representation of the electrostatic term also differs from those used in the extensively parameterized force fields of Allinger[1] and Oie[94], where a bond dipole model of electrostatics is used. We feel our approach is the more general, particularly since the ionic systems considered here would be difficult to adequately represent with bond moments.

However, we stress that one could equally well use our methodology for deriving empirical electrostatic models from quantum mechanical calculations[18] to fit to either bond dipoles or partial charge models, or a combination of the two.

CONCLUSIONS

We have presented an approach and the results of the development of a molecular mechanical force field. To our knowledge, this is the first time that such a general force field has been developed in a consistent way for both proteins and nucleic acids. Although we have done only a limited number of detailed calculations to test the parameter set, the results of calculations on furanose sugar puckering, base stacking and hydrogen bonding, base paired dinucleoside phosphate refinement, Φ, Ψ energy contours for dipeptide models, H-bonding calculations on protein polar and ionic groups and refinement of insulin all suggest that the model contains no major flaws. However, we have also delineated areas for future improvement of such force fields, and we feel that the results presented here are a reasonable starting point for such development.

ACKNOWLEDGEMENTS

We are pleased to acknowledge the support of the NIH (GM-29072 and CA-25644) in this study. DAC is an Alfred P. Sloan Foundation Fellow. Our development of the torsional second derivative part of AMBER was based upon a formulation by B. Brooks and M. Karplus of Harvard University. The curve plotting software was provided to us by N. Pattabiramin at U.C.S.F.. Unpublished normal modes analyses for the nucleic acids were done by D.

Nguyen of U.C. Davis. A special thanks to Shoshana Wodak for bringing the question of protein compaction in insulin refinements to our attention and Peter Murray-Rust for carrying out the Cambridge crystal file search for phosphate geometries.

Table I
List of Atom Types

United[a]	Type
C2	sp^3 carbon with two hydrogens
C3	sp^3 carbon with three hydrogens
CD	sp^2 aromatic carbon in 6 membered ring with one hydrogen
CE[c]	sp^2 aromatic carbon in five membered ring between two nitrogens with one hydrogen (in purines)
CF	sp^2 aromatic carbon in five membered ring next to a nitrogen without a hydrogen (e.g. C_δ-N_ε=C_ε in histidine)
CG	sp^2 aromatic carbon in five membered ring next to a N-H (e.g. C_δ-N_ε=C_ε in histidine)
CH	sp^3 carbon with one hydrogen
CI	sp^2 carbon in six membered ring of purines between two "NC" nitrogens
CJ	sp^2 carbon in pyrimidines at positions 5 & 6 (more pure double bond than aromatic) with one hydrogen
CP[c]	sp^2 aromatic carbon in five membered ring between two nitrogens with one hydrogen (in His)
All Atom[b]	**Type**
C	sp^2 carbonyl carbon and aromatic carbon with hydroxyl substituent in tyrosine
C*	sp^2 aromatic carbon in five membered ring with one substituent (e.g. CE_γ in Trp)
CA	sp^2 aromatic carbon in 6 membered ring with one substituent
CB[d]	sp^2 aromatic carbon at junction between five and six membered rings (e.g. CE_δ in Trp, C4 and C5 in purines)
CC	sp^2 aromatic carbon in five membered ring with one substituent and next to a nitrogen group (e.g. C_γ in His)
CM	sp^2 same as CJ but with one substituent
CN[d]	sp^2 aromatic junction carbon in between five and six membered rings (e.g. C_ε in Trp)
CT	sp^3 carbon with four explicit substituents

a United atom carbons with implicit inclusion of hydrogens.

b Non hydrogen containing carbons.

c Structural differences in the internal angles of the five membered rings are the reason why these atoms, which appear in the same environment by definition, are assigned different atom types.

d Neutron diffraction studies on tryptophan show that $\vartheta_{eq}(C_\varepsilon', C_\delta', C_\delta)=116.2°$ while $\vartheta_{eq}(C_\zeta, C_\varepsilon, C_\delta')=122.7°$. Due to this structural variation, we opted for two atom types at the junction carbons in Trp.

Table I (continued)
List of Atom Types

Nitrogens	Type
NC	sp^2 nitrogen in six membered ring between with lone pairs (e.g. N3 in adenine)
NA	sp^2 nitrogen in five membered ring with hydrogen attached (e.g. protonated His)
NB	sp^2 nitrogen in five membered ring between with lone pairs (e.g. N7 in purines)
N*	sp^2 nitrogen in purines and pyrimidines with alkyl group attached (N9 in purines, N1 in pyrimidines)
N	sp^2 nitrogen in amide groups
N2	sp^2 nitrogen in base NH_2 groups and arginine NH_2
N3	sp^3 nitrogen with four substituents (e.g. Lys N_ζ)
NT	sp^3 nitrogen with three substituents (e.g. unprotonated amines)
Oxygens	**Type**
O	carbonyl oxygen
O2	carboxyl and phosphate non-bonded oxygens
OS	ether and ester oxygens
OH	alcohol oxygens
Hydrogens	**Type**
H3	hydrogens of lysine and arginine (positively charged)
H2	amino hydrogens from NH_2 in purines and pyrimidines
HC	explicit hydrogen attached to carbon
H	amide and imino hydrogens
HO	hydrogen on hydroxyl or water oxygen
HS	hydrogen attached to sulfur
Sulfurs	**Type**
S	sulfurs in disulfide linkages and methionine
SH	sulfur in cystine
Phosphorus	**Type**
P	phosphorus in phosphate groups
Lone Pair	**Type**
LP	lone pairs

Table II
Standardized Parameters for Scaling Algorithms

Bond	r_{eq}^{a}	K_r^{b}
Pure C-C	1.507[c]	317[d]
Pure C=C	1.336[e]	570[f]
Pure C-N	1.449[g]	337[h]
Pure C=N	1.273[i]	570[j]
Torsion	r_{eq}^{a}	$V_2/2^{k}$
Pure X-C-C-X	1.507[c]	0.0[l]
Partial X-C=C-X	1.397[m]	5.5[n]
Pure X-C=C-X	1.336[e]	30.0[o]
Pure X-C-N-X	1.449[g]	0.0[p]
Partial X-C=N-X	1.335[q]	10.0[r]
Pure X-C=N-X	1.273[i]	30.0[s]

a In Å

b In kcal/mole Å^2.

c Microwave data from acetone ref. 27.

d Value taken from MM2 ref 1.

e Microwave data from propene ref 26.

f Default from NMA normal mode analysis for carbonyl force constant.

g Benedetti structural data ref. 28.

h Value derived from normal mode analysis on NMA.

i Microwave data from methylene imine ref. 27.

j Default value.

k In kcal/mole.

l Assumed free rotation about pure C-C single bond.

m Structural data from benzene ref. 27.

n From normal modes analysis of benzene.

o Approximate rotational barrier of ethtylene is ≈60 kcal/mole, see ref. 32.

p Assumed free rotation about a pure single C-N bond.

q Benedetti structural data ref. 28.

r Ref. 3.

s Calculated rotational barrier in methylene imine is 57.5 kcal/mole, see ref. 33.

Table III
Comparison of Two Model Force Fields for THF and MEE

Parameter	FF1[a]	FF2[b]	Experiment[c]
K_ϑ(C-C-O)[d]	50	100	-
K_ϑ(C-O-C)[d]	55	80	-
K_ϑ(C-C-C)[d]	40	63	-
$V_3/2$(C-C)[e]	1.45	2.0	-
$V_3/2$(C-O)[e]	1.05	1.45	-
Results for Tetrahydrofuran			
ΔE (C_2-C_s)[f]	0.1	0.2	$\approx$0.1[h]
ΔE (C_2-C_{2v})[g]	3.44	3.7	3.52[h]
Structural Parameters			
C_2 conformation			
q[i]	0.43	0.40	0.38[j]
$\vartheta(C-O-C)$	109.7	109.9	110.5[j]
$\vartheta(C-C-O)$	105.1	105.2	106.5[j]
$\vartheta(C-C-C)$	100.6	101.3	101.8[j]
C_s conformation			
q[i]	0.40	0.37	0.364[k],0.38[k]
$\vartheta(C-O-C)$	106.0	107.0	106.2[k]
$\vartheta(C-C-O)$	103.4	103.7	105.0[k]
$\vartheta(C-C-C)$	104.2	104.2	104.1[k]
Vibrational Frequencies[l]			
B; pseudorotation	43	39	($\approx$0)[m]
A; out of plane torsion	301	356	286[n]
B; ring bending	441	543	581[n]
A; ring bending	509	620	655[n]
A; ring stretching	840	839	888,913[n]
B; ring stretching	955	976	909,964[n]
A; ring stretching	970	991	918,1071[n]
A; ring stretching	1069	1116	1030,–[n]
B; ring stretching	1130	1148	1070,1241[n]
Results for Methylethylether			
$\Delta E(g-t)$	1.4	1.6	1.4$\pm$0.2[o]
$\Delta E(c-t)$	7.7	9.4	(5.9)[p]
$\Delta E(120°-t)$	2.1	2.8	(2.9)[p]
Structural Parameters			
Gauche			
Φ	81	77	72,85[q]
ϑ(C-O-C)	112.9	113.2	(113.2)[p]
ϑ(C-C-C)	112.2	111.4	(112.2)[p]
Cis			
ϑ(C-O-C)	119.5	117.6	(116.3)[q]
ϑ(C-C-C)	119.7	118.1	(117.3)[q]

a Force field developed initially.

b Force field developed with larger bond angle bending terms.

c Experimental values.

d Bond bending force constant in kcal/mole $Å^2$.

e Torsional parameter in kcal/mole.

f Difference in energy between C_2 and C_s conformations of THF (kcal/mole).

g Difference in energy between C_2 and planar C_{2v} conformations of THF (kcal/mole).

h ref. 34

i Mean out of plane distance of ring, as defined in ref. 35.

j ref. 36.

k ref. 37.

l Vibrational frequencies of THF in cm^{-1}. The symmetry is C_2.

m Pseudorotation mode, see ref. 34.

n Experimental frequencies from ref. 38a and 38b. The two sets of assignments in the two references are given when they are not in agreement.

o ref. 39,40,41.

p MM2 calculations ref. 39.

q See ref. 41 for discussions on these parameters.

Table IV
Dimethylphosphate Energies and Geometries[a]

	FF2[b]	FF0[c]	QM[d]	X-Ray[e]
Relative Energies of Conformations[f]				
g,*g*[g]	0	0	0	45
g,*t*[h]	0.8	0.9	0.14	6
t,*t*[i]	1.7	1.9	0.88	
Geometrical Parameters (*g*,*g*)[j]				
Φ_1^k	70	63	68	73
Φ_2^i	70	63	68	73
ϑ(C-O-P)[m]	120.6	123.0	112.4	121.7
ϑ(O-P-O)[n]	103.0	103.6	98.8	104.8
ϑ(O'-P-O')[o]	119.8	118.6	125.7	119.7
ϑ(O'-P-O)[p]	108.3	109.5	-	110.6
ϑ(O'-P-O)[p]	108.1	107.4	-	105.6
Geometrical Parameters (*gt*)[q]				
Φ_1^k	69	63	75	74
Φ_2^i	179	179	179	169
ϑ(O-P-O)[n]	102.8	102.4	94.9	99.3
ϑ(O'-P-O')[o]	119.8	119.2	124.3	—[e]
ϑ(O'-P-O)[p]	108.4	109.6	-	—[e]
ϑ(O'-P-O)[p]	108.0	106.6	-	—[e]
Geometrical Parameters (*tt*)[r]				
ϑ(O-P-O)[n]	102.6	101.2	91.0	—[e]
ϑ(O'-P-O')[o]	119.9	119.7	123.1	—[e]
ϑ(O'-P-O)[p]	108.3	108.6	-	—[e]

a Energies in kcal/mole, angles in degrees.

b This study, FF2 as described in Table III, with scale factor of 0.5 for 1-4 van der Waals and electrostatic interactions.

c Previous force field, see ref. 16.

d Gorenstein et al, ref. 51.

e Unpublished results by Peter Murray-Rust using the Cambridge Crystal data file. The geometrical parameters are averages taken from the 10 structures which have R < 9% and no atom larger than Br. Of these, 7 are *g*,*g* and 3 are *g*,*t*. Except for ϑ(O-P-O), there is no statistically significant difference between the ϑ's, so the values reported under the *g*,*g* average are the average for all 10 structures.

f For x-ray structures, the numbers reported are the number of structures of type R-O-P-O_2'-O-R'$^-$ with each conformation.

g *Gauche*,*gauche* conformation.

h *Gauche,trans* conformation.

i Trans-trans conformation.

j Selected geometrical parameters for the *g,g* conformation.

k Φ_1 = dihedral angle $C_1-O_1-P-O_2$.

l Φ_2 = dihedral angle $O_1-P-O_2-C_2$.

m C-O-P angle (or its average).

n O-P-O angle (or its average).

o O'-P-O' angle (involving anionic oxygens).

p O'-P-O angle; there are four such angles and we report only the largest and smallest.

q Selected geometrical parameters for the *g,t* conformation.

r Selected geometrical parameters for the *t,t* conformation.

Table V
Calculated and Observed Frequencies for Diethyl Phosphate (cm^{-1})

Mode	This Work	Brown and Peticolas[a]	Exp.[b]
ν_7	54	-	-
ν_8	105	-	-
ν_9	107	-	-
ν_{10}	136	-	-
ν_{11}	200	187	195
ν_{12}	211	201	210
ν_{13}	287	290	321
ν_{14}	290	329	345
ν_{15}	367	333	357
ν_{16}	418	398	393
ν_{17}	458	492	503
ν_{18}	524	565	551
ν_{19}	580	578	569
ν_{20}	741	775	763
ν_{21}	841	814	812
ν_{22}	983	941	945
ν_{23}	983	954	945
ν_{24}	1036	1051	1053
ν_{25}	1088	1064	1053
ν_{26}	1102	1080	1077
ν_{27}	1235	1225	1215

a ref. 52.

b ref. 53.

Table VI
Comparison of Calculations on Deoxydenosine and a Simplified Model

	FF1[a]	Olson Model[b]
E(C2' *endo*)[c]	0.0	0.0
W^d	152	159
q^e	0.40	0.41
E(C3' *endo*)[c]	0.56	0.46
W^d	5	18
q^e	0.40	0.42
E(O1' *endo*)[c]	1.29	1.27
W^d	75	86
q^e	0.40	0.42

a This work, with force constants as in Table III ; calculations on deoxyadenosine with base, 5' and 3' substituents all included (CH united atoms for CH carbons). Scale factor of 1.0 for 1-4 non-bonded and electrostatic interactions, $\varepsilon = R_{ij}$.

b Using the standard model of Olson (ref. 30), with explicit hydrogens on the sugar, but united atoms at the 5' (CH_3), 3' (OH) and 1' (NH_2) positions and using her torsional, bending, van der Waals and electrostatic parameters with complete minimization (all K_τ=300 kcal/mole Å^2, ε=4).

c Relative energy in kcal/mole of given conformation.

d Pseudorotation angle (see ref. 35) of C2' *endo* or C3' *endo* conformation; value at the top of the potential curve for O1' *endo*.

e Mean out of plane distance for furanose atoms (in Å), see ref. 35.

Table VII
Calculations on Adenosine and Deoxyadenosine

Deoxyadenosine	FF1[a]	FF2[b]	FF2'[c]	FF2"[d]	Exp[e].
ΔE(C3'*endo*-C2'*endo*)[f]	0.56	0.53	0.66	0.52	0.66
ΔE(O1'*endo*-C2'*endo*)[g]	1.29	2.00	1.80	1.80	-
W(C2'*endo*)[h]	152	152	170	155	165
q(C2'*endo*)[i]	0.40	0.38	0.36	0.38	(0.35-0.41)
W(C3'*endo*)[h]	5	3	1	5	(2-20)
q(C3'*endo*)[i]	0.40	0.37	0.37	0.38	(0.35-0.41)
Adenosine		FF2[b]	FF2'[c]	FF2"[d]	Exp.
ΔE(C3'*endo*-C2'*endo*)[f]		-0.68	1.46	0.21	(0.19-0.42)
W(C2'*endo*)[h]		174	178	17	(150-170)
q(C2'*endo*)[i]		0.37	0.35	0.37	(0.35-0.41)
W(C3'*endo*)[h]		3	359	3	(2-20)
q(C3'*endo*)[i]		0.39	0.34	0.38	(0.35-0.41)

a Same force field as in Table III.

b Same force field as in Table III; V_ϕ are larger than force field 1 (FF1).

c Same as FF2 with 1-4 electrostatic tems scaled by 0.5.

d Same as FF2' with $\varepsilon=4R_{ij}$ rather than $\varepsilon=R_{ij}$.

e Experimental data from Davies (ref. 56) and, Altona and Sundaralingham (ref. 57).

f Energy difference between energy minimized C3' *endo* and C2' *endo* conformations.

g Energy difference between O1' *endo* and C2' *endo* conformations.

h Energy refined pseudorotation angle (ref. 35) for given conformation.

i Energy refined mean out of plane sugar distance (see ref. 35) for sugar ring.

Table VIII
Hydrogen Bonding and Stacking for Base Pairs

Complex	$\Delta E(\varepsilon=1)^a$	$\Delta E(\varepsilon=R)^b$	$\Delta E(\text{Langlet})^c$	$\Delta H(\text{expt.})^d$
GC Watson-Crick[e]	-21.2	-21.6	-23.7	-21.0
AT Watson-Crick[f]	-11.3	-12.9	-12.9	-13.0
AT Hoogsteen[g]	-9.8	-13.5	-13.6	-13.0
1,3-Dimethyluracil Stack[h]	-9.8[i]	-9.3	-9.1	(-9.1)

a Energy of complex formation with $\varepsilon=1$ in kcal/mole.

b Energy of complex formation with $\varepsilon=R_{ij}$ in kcal/mole.

c Energy calculated by Langlet et al (ref. 58).

d Experimental value for associations inferred from the experiments by Yanson et al (ref. 59). In the case of the 1,3-dimethyluracil stacking, the value in parentheses is the value calculated by Langlet *et al* (ref. 58), since these authors showed that there was an important electric field dependence in the experiments by Yanson.

e Watson and Crick H-bonded structure of 9-methylguanine and 1-methylcytosine. Model built using computer graphics and then energy refined.

f Watson and Crick H-bonded structure of 9-methyladenine and 1-methylthymine. Model built using computer graphics and then energy refined.

g Hoogsteen H-bonded structure of 9-methyladenine and 1-methylthymine. Model built using computer graphics and then energy refined.

h Stacked complex of 1,3-dimethyluracil model built using figure 11 A1 in the paper by Langlet *et al* (ref. 58) and energy refined, base-base minimum energy distance = 3.43 Å.

i Calculation using explicit representation of C-H groups and the Hagler *et al* non-bonded parameters (ref. 19) base-base minimum energy distance ≈3.71Å.

Table IX
Calculations on sequence specificity of melting T of nucleotides

DNA polymer	$-\Delta t_m^a$	$\Delta E(\varepsilon=1)^b$	$\Delta E(\varepsilon=R_{ij})^c$
1. poly[d(G-C)]·poly[d(G-C)]	-	0.0	0.0
2. poly d(G)·poly d(C)	13	5.4	2.2
3. poly[d(A-T)]·poly[d(A-T)]	-	0.0	0.0
4. poly d(A)·poly d(T)	-9	-0.9	-1.2
5. poly[d(T-G)]·poly[d(C-A)]	-	0.0	0.0
6. poly[d(T-C)]·poly[d(G-A)]	7	1.7	0.8
7. poly[d(A-T-C)]·poly[d(G-A-T)]	-	1.1	0.9
8. poly[d(T-T-G)]·poly[d(C-A-A)]	2	0.0	0.0
9. poly[d(T-A-C)]·poly[d(G-T-A)]	5	0.5	0.5
10. poly[d(T-T-C)]·poly[d(G-A-A)]	8	1.2	0.5

a Difference in melting temperature between isomers in °C; a positive value in the case of 1 vs. 2; 3 vs. 4; 5 vs. 6; means the hetero-polymer (1, 3 or 5) melts higher (is more stable). In the case of polymers 7-10; the highest melting polymer is 7, followed in order by 8, 9, 10; see ref. 16 for discussion; experimental data from ref. 60.

b Difference in calculated energy (kcal/mole) between polymers. For example, in comparing 1 vs. 2, we compare the energy of $d(CG)_2$ and $d(GC)_2$ to the energy of $d(G_2)\cdot d(G_2)$.

c Same as b for calculations with dielectric constant $\varepsilon=R_{ij}$.

Table X
Geometries and Local Minima for Glycyl and Alanyl Dipeptides

Local Minima	Φ^a	Ψ^b	ΔE^c
	Glycyl Dipeptide All Atom Representation		
1-7 H-Bonded	76°	-65°	0.0
1-7 H-Bonded	-76°	65°	0.0
Extended	180°	180°	3.1
1-10 H-Bonded	60°	39°	4.0
1-10 H-Bonded	-60°	-39°	4.0
	Glycyl Dipeptide United Atom Representation		
1-7 H-Bonded	77°	-64°	0.0
1-7 H-Bonded	-77°	64°	0.0
Extended	180°	180°	3.2
1-10 H-Bonded	66°	35°	4.1
1-10 H-Bonded	-66°	-35°	4.1
	Alanyl Dipeptide All Atom Representation		
1-7 H-Bonded	-76°	66°	0.0
1-7 H-Bonded	69°	-64°	0.9
Extended	-161°	169°	3.2
1-10 H-Bonded	-62°	-40°	3.6
1-10 H-Bonded	53°	43°	4.5
	Alanyl Dipeptide United Atom Representation		
1-7 H-Bonded	-79°	69°	0.0
1-7 H-Bonded	68°	-58°	0.8
Extended	-150°	154°	2.3
1-10 H-Bonded	-69°	-29°	3.0
1-10 H-Bonded	55°	35°	4.6

a Φ convention appears in ref. 69.

b Ψ convention appears in ref. 69.

c Relative energy in kcal/mole.

Table XI
Normal Modes Analysis

Assignment	Experimental ν^b	Calculated ν^c
	Benzene[d]	
	Out of Plane	
E_{2u}	410	415
B_{2g}	703	703
	In Plane	
E_{2g}	606	616
A_{1g}	992	949
B_{1u}	1010	972
E_{1u}	1038/1486	1182
B_{2u}	1310	1596
E_{2g}	1596	1544
	N-methylacetamide[e]	
	A'' Out of Plane	
peptide torsion	192	194
C=O out of plane wag	600	598
N-H out of plane wag	725	718
	A' In Plane	
C-N-CH_3 bend	289	315
CH_3 bend	439	452
amide IV	628	591
CH_3-C stretch	883	838
C-N stretch	1120	1023
amide III	1300	1295
amide II	1569	1588
amide I	1660	1667
N-H stretch	3306	3304
	Methanol[f]	
C-O-H bend	1033	1040
C-O stretch	1345	1300
O-H stretch	3681	3709
	Methanethiol[g]	
C-S-H bend	708	701
C-S stretch	803	813
S-H stretch	2573	2571
	Dimethylsulfide[h]	
C-S-C Bend	282	284
C-S Stretch	691	705
C-S Stretch	741	733
	Dimethyl Disulfide[i]	
C-S-S-C Torsion	102	104
C-S-S Bend	239	234
C-S-S Bend	272	272
S-S Stretch	509	509
C-S Stretch	689	718

a Relative assignment number.

b In cm^{-1}

c Calculated with the second derivative routine in Amber, in cm^{-1}

d Ref. 54.

e Ref. 69.

f Ref. 54.

g Ref. 70.

h Ref. 70.

i Ref. 70.

Table XII
Water/Chromophore Hydrogen Bonding Distances and Energies

Chromophore	H_2O as Proton Acceptor Distance[a]	ΔE[b]	H_2O as Proton Donor Distance[a]	ΔE[b]
	$\varepsilon = R_{ij}$			
NMA[c]	1.82	-5.8	1.79	-7.1
Methanol[d]	1.80	-5.8	1.80	-6.0
Imidazole[e]	1.78	-7.6	1.81	-6.5
Phenol[f]	1.77	-7.5	1.85	-4.5
Imidazolium[g]	1.72	-14.0		
Me-Guanidinium[g]	1.68	-11.8		
Me-Ammonium[h]	1.66	-14.0		
Indole[i]	1.79	-6.9		
Water[j]	1.79	-6.4		
Acetate[k]			1.73	-19.8
	$\varepsilon = 1$			
NMA	1.87	-5.2	1.83	-6.7
Methanol	1.85	-4.6	1.86	-5.0
Imidazole	1.83	-6.7	1.87	-5.8
Phenol	1.83	-6.1	1.89	-4.1
Imidazolium	1.77	-15.3		
Me-Guanidinium	1.73	-13.5		
Me-Ammonium	1.70	-17.0		
Indole	1.84	-5.8		
Water	1.84	-5.1		
Acetate			1.80	-19.6

a In Å

b In kcal/mole.

c *ab initio* 4-31G calculations (ref. 73) find formamide---HOH to give a ΔE=-9.2 kcal/mole and formamide---OH_2 to give a ΔE=-6.8 kcal/mole. Given the usual overestimation of H-bond energies by 4-31G, these *ab initio* energies are probably upper bounds.

d ref. 74,75.

e Del Bene finds ΔE=-5.6 kcal/mole for Im---HOH and a ΔE=-9.0 kcal/mole for Im---OH_2 ref. 76.

f No experimental data but should be a better proton donor than H_2O and a worse proton acceptor.

g No experimental data but the right order of magnitude ref. 77.

h Kebarle suggests $\Delta H \approx$-16.0 kcal/mole, ref. 77.

i No experimental data but ΔE should be similar to imidazole.

j Experimental data (ref. 78) suggests a ΔE=-5.5 kcal/mole.

k Very accurate 6-31G**/MP2 calculations (ref. 79) on $HCOO^-$---H_2O suggest a ΔE=-21 kcal/mole.

Table XIII
Refinement of Insulin

	Energy Evaluations	RMS[a] Gradient	RMS[b] Backbone	RMS[c] All Atoms	Compaction/ Expansion[d]
Cutoff[e]=9.0Å, $\varepsilon=R_{ij}$	1318	0.14	0.28	0.43	-1.6%
Cutoff=9.0Å, $\varepsilon=4R_{ij}$	3642	0.30	0.51	0.66	-9.4%
Cutoff=9.0Å, $\varepsilon=R_{ij}$ (Jorgensen Nonbonded)[f]	1080	0.20	0.26	0.41	+0.2%
Cutoff=9.0Å, $\varepsilon=R_{ij}$ (Old Parameter Set)[g]	1100	0.06	0.79	1.01	-16.0%
Cutoff=12.0Å, $\varepsilon=R_{ij}$	3937	0.09	0.56	0.72	-7.0%

a RMS gradient, in units of kcal/Å, calculated at end of the energy refinement.

b Root mean square fit in Å for the minimized insulin backbone atoms, compared with coordinates from the starting crystal structure.

c Root mean square fit in Å for all the minimized insulin atoms, compared with coordinates from the starting crystal structure.

d These values represent the ratio of the minimized volume to initial volume. The specific volumes were generated using the radius of gyration calculated from all insulin backbone atoms.

e Cutoff is the distance from which all nonbonded interactions will be evaluated.

f Ref. 24.

g This run incorporated our old parameter set (ref. 17) except that we used the new hydrogen bond 10-12 potentials. The 10-12 parameters were necessary to keep atoms from "fusing", which we found occurred in our initial run.

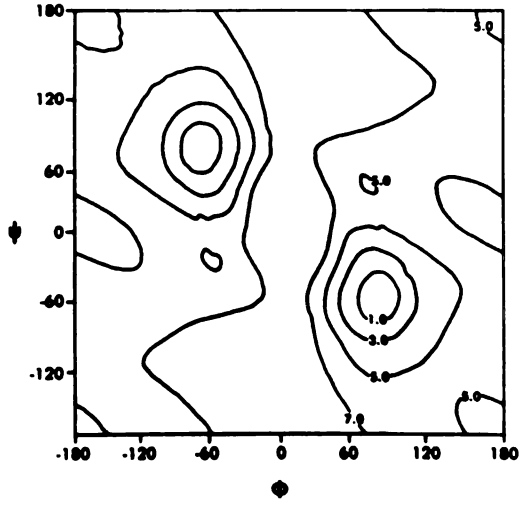


Figure 1. Energy contour map for *N*-acetyl-*N'*-methylglycinamide. Contours are in kcal/mol. The usual IUPAC convention for Φ, Ψ is used.

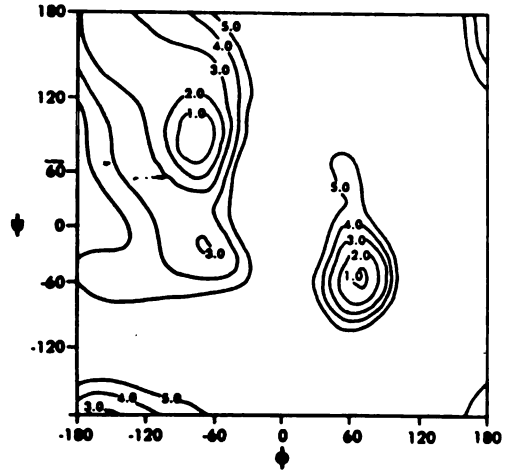


Figure 2. Energy contour map for *N*-acetyl-*N'*-methylalaninamide. Contours are in kcal/mol.

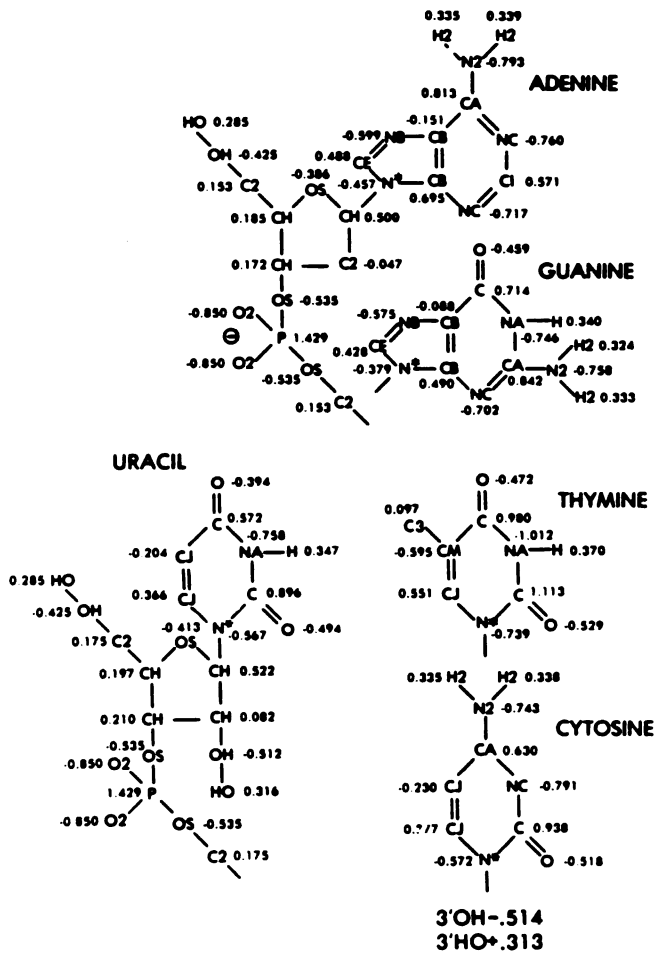


Figure 3. Net atomic charges for the nucleic acids in units of electron charge.

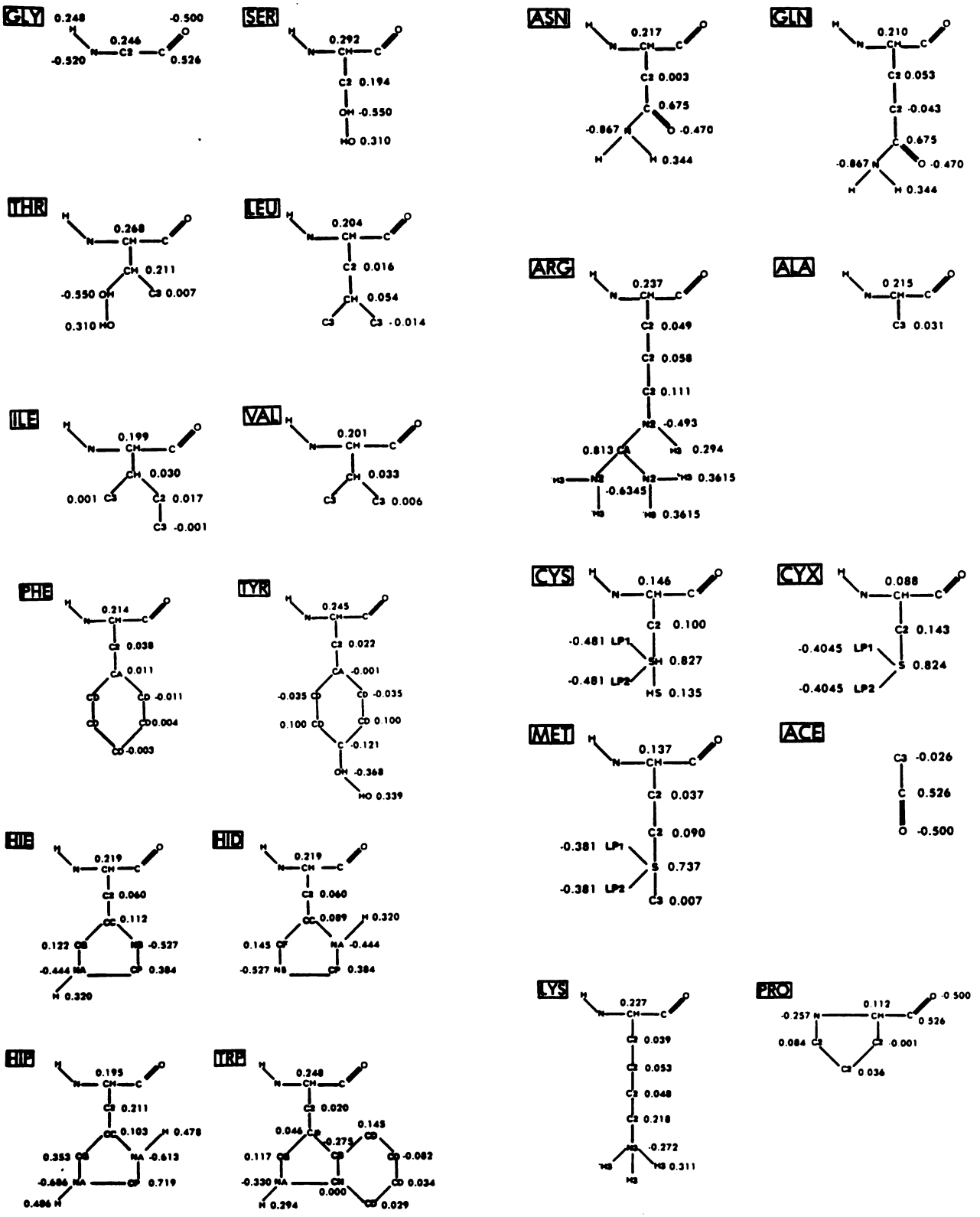


Figure 4. Net atomic charges for the amino acid side chains and backbone atoms in units of electron charge.

(1) N. Allinger, J. Amer. Chem. Soc., 1977, **99**, 8127.

(2) O. Ermer and S. Lifson, J. Amer. Chem. Soc., 1973, **95**, 4121.

(3) F. Momany, R. McGuire, A. Burgess and H. Scheraga, J. Phys. Chem., 1975, **79**, 2361.

(4) A. Warshel, M. Levitt and S. Lifson, J. Mol. Spect., 1970, **33**, 84.

(5) M. Levitt, J. Mol. Biol., 1974, **82**, 93.

(6) B. Gelin and M. Karplus, Biochem, 1979, **18**, 1256.

(7) P. Danber and A. Hagler, Acct. Chem. Res., 1980, **13**, 105.

(8) J. Hermans, D. Ferro, J. McQueen and S. Wei in "Environmental Effects on Molecular Structure and Properties", 1976, B. Pullman ed., Dordrecht, Holland.

(9) V. Sasisekharan in "Conformation of Biological Molecules and Polymers",1973, E. Bergman and B. Pullman eds., Jerusalem and A. Lakshminarayanan and V. Sasisekharan, Biopolymers, 1969, **8**, 475, 489, 505.

(10) W. Olson and P. Flory, Biopolymers, 1972, **11**, 25.

(11) R. Rein, N. Goel, N. Fukida, M. Pollack and P. Claverie, Ann. N.Y. Acad. Sci., 1969, **153**, 805-814; R. Ornstein and R. Rein, Biopolymers, 1979, 18, 2821-2847.

(12) B. Pullman and A. Pullman, Prog. Nucl. Acid Res., 1969, **9**, 327.

(13) M. Levitt, Proc. Nat. Acad. Sci., 1978, **75**, 640.

(14) M. Levitt, Cold Spring Harbor Symposium on DNA Structure, 1982.

(15) S. Lifson and A. Warshel, J. Chem. Phys., 1968, **49**, 5116.

(16) P. Kollman, P. Weiner and A. Dearing, Biopolymers, 1981, 20, 2583.

(17) J. Blaney, P. Weiner, A. Dearing, P. Kollman, E. Jorgensen, S. Oatley, J. Burnridge and C. Blake, J. Amer. Chem. Soc., 1982, **104**, 6424 and G. Wipff, A. Dearing, P. Weiner, J. Blaney and P. Kollman, J. Amer. Chem. Soc., 1983, **105**, 997.

(18) U.C. Singh and P. Kollman, "An Approach to Computing Electrostatic Charges for Molecules", J. Comp. Chem., 1983 (accepted for publication).

(19) A. Hagler, E. Euler and S. Lifson, J. Amer. Chem. Soc., 1974, 96, 5319.

(20) H. Margenam and N. Kestner in "The Theory of Intermolecular Forces", Pergammon Press, Oxford, 1970.

(21) P. Weiner and P. Kollman, J. Comp. Chem., 1981, **2**, 287,.

(22) L. Pauling in "The Nature of the Chemical Bond", 1960, Third Edition, Ithaca, New York.

(23) A. Bondi, J. Phys. Chem., 1964, **68**, 441.

(24) L. Dunfield, A. Burgess and H. Scheraga, J. Phys. Chem., 1978, 82, 2609.

(25) W. Jorgensen, J. Amer. Chem. Soc., 1981, **103**, 335.

(26) S. Cox and D. Williams, J. Comp. Chem., 1981, **2**, 304.

(27) M. Harmony, V. Laurie, R. Kuezkowski, R. Schwendeman, D. Ramsay, F. Lovas, W. Lafferty and A. Maki, J. Phys. Chem. Ref. Data, 1979, v. 8, #3, 619.

(28) E. Benedetti in "Peptides-Proceeds of the 5th American Peptide Symposium", M. Goodman and J. Meienhofer, eds., New York, 1977, p. 257.

(29) D. Hayes, S. Rothenberg and P. Kollman, J. Amer. Chem. Soc., 1977, 99, 2150; F. Marsh, P. Weiner, J. Douglas, P. Kollman, G. Kenyon and J. Gerlt, J. Amer. Chem. Soc., 1980, **102**, 1660; G. Govil, Biopolymers, 1976, 15, 2303.

(30) W. Olson, J. Amer. Chem. Soc., 1982, **104**, 278.

(31) R. Barker and G. Boudreaux, Spec. Acta, 1967, **23A**, 727.

(32) J. Douglas, B. Rabinovich and F. Looney, J. Chem. Phys., 1955, 23, 315.

(33) J. Lehn, Ther. Chim. Acta., 1970, **16**, 351.

(34) G. Engelsholm, A. Luntz, W. Gwinn and D. Harris, J. Chem. Phys., 1969, 50, 2446.

(35) D. Cremer and J. Pople, J. Amer. Chem. Soc., 1975, **97**, 1354.

(36) A. Almenningen, H. Seip and T. Walladsen, Acta. Chim. Scand., 1969, 23, 2748.

(37) H. Geise, W. Adams and L. Bartell, Tetrahedron, 1969, **25**, 3045.

(38) a) J. Eyster and E. Prohofsky, Spec. Acta, 1974, **30A**, 2041.
b) J. Deroualt, M. Forel and P. Maraval, Can. J. Spect., 1978, 23, 67.

(39) S. Profeta, Jr. and P. Kollman, unpublished MM2 results on methylethylether.

(40) T. Kitayawa and T. Miyazawa, Bull. Chem. Soc. Japan, 1968, 41, 1976.

(41) W. Jorgensen and M. Ibrahim, J. Amer. Chem. Soc., 1981, **103**, 3976.

(42) T. Brunck and F. Weinhold, J. Amer. Chem. Soc., 1979, **100**, 1700.

(43) The code to derive the analytical second derivatives for torsional energies which avoided singularities was provided to us by B. Brooks and M. Karplus of Harvard University.

(44) A. Hagler, P. Stern, R. Sharon, J. Becker and F. Naider, J. Amer. Chem. Soc., 1979, **101**, 6842.

(45) W. Jorgensen and M. Ibrahim, J. Amer. Chem. Soc., 1981, **103**, 3976.

(46) We first reproduced values for internal energy and volume reported in ref. 41 and then repeated the simulation with the altered parameters.

(47) A. Verma, W. Murphy and H. Bernstein, J. Chem. Phys., 1974, 60, 1540; K. Kuchitsu, Bull. Chem. Soc. Japan, 1959, **32**, 748.

(48) This is an MM2 value from N. Allinger and S. Profeta, Jr., J. Comp. Chem., 1980, **1**, 181.

(49) M. Newton, J. Amer. Chem. Soc., 1973, **95**, 256.

(50) Unpublished analysis of $ROPO_2R^-$ structures by Peter Murray-Rust.

(51) D. Gorenstein, J. Findlay, B. Luxon and D. Kar, J. Amer. Chem. Soc., 1977, **99**, 3473.

(52) E. Brown and W. Peticolas, Biopolymers, 1975, **14**, 1259.

(53) ref. 58 and T. Shimanouchi, M. Tsuboi and Y. Kyogoku, Adv. Chem. Phys., 1964, Duchesne, J., ed., London, Interscience, vol VII, p. 435.

(54) M. Karplus and J. Kushick, Macromolecules, 1981, **14**, 325.

(55) T. Shimanouchi in "Tables of Molecular Vibrational Frequencies", parts 1-3, National Standard Reference Data Series - National Bureau of Standards, Washington, D.C., 1967.

(56) D. Davies, Prog. Nucl. Mag. Res. Spect., 1978, **12**, 135.

(57) C. Altona and M. Sundaralingham, J. Amer. Chem. Soc., 1972, 94, 8205.

(58) J. Langlet, P. Claverie and F. Caron in "Intermolecular Forces", 14th Jerusalem Symposium, 1981, B. Pullman ed., Reidel, Dordrecht.

(59) I. Yanson, A. Teplitsky and L. Sukhodur, Biopolymers, 1979, 18, 1149.

(60) R. Wells, J. Larson, R. Grant, B. Shortle and C. Cantor, J. Mol. Biol., 1970, **54**, 465.

(61) B. Maigret, B. Pullman and D. Perahia, J. Theor. Biol., 1971, **31**, 269.

(62) L. Schafer, C. Alsenoy and J. Scarsdale, J. Chem. Phys., 1982, 73, 1439.

(63) G. Ramachandran and V. Sasisekharan, Adv. Prot. Chem., 1968, 23, 283.

(64) D. Brant, W. Miller and P. Flory, J.M.B., 1967, **23**, 47.

(65) K. Gibson and H. Scheraga, Biopolymers, 1966, **4**, 709.

(66) B. Gelin and M. Karplus, J. Amer. Chem. Soc., 1975, **97**, 6996.

(67) M. Cung, M. Marraud and J. Neel in "The Jerusalem Symposia on Quantum Chemistry and Biochemistry - Conformation of Biological Molecules and Polymers", E. Bergmann and B. Pullman eds., 1973, Jerusalem, p.69.

(68) M. Avignon and J. Lascombe, ibid, p.97.

(69) J. Richardson, Adv. Prot. Chem., 1981, **34**, 167.

(70) H. Rey-Lafon and M.T. Forel, Spec. Acta., 1973, **29A**, 471.

(71) D. Scott and J. McCullough, J. Amer. Chem. Soc., 1958, **80**, 3554.

(72) T. Dyke and J. Muenter, J. Chem. Phys., 1974, **60**, 2929; L. Curtiss, D. Frurip and M. Blander, 1979, J. Chem. Phys., 71, 2703.

(73) A. Johansson, P. Kollman, S. Rothenberg and J. McKelvey, J. Amer. Chem. Soc., 1974, **96**, 3794.

(74) J. Del Bene, J. Chem. Phys., 1971, **55**, 4633.

(75) Y. C. Tse, M. Newton and L. Allen, Chem. Phys. Lett., 1980, **75**, 350.

(76) J. Del Bene, J. Amer. Chem. Soc., 1978, **100**, 5285.

(77) P. Kebarle, Environmental Effects on Molecular Structure and Properties, B. Pullman, ed., D. Reidel, Dordrecht, Holland, 1976, p. 81.

(78) L. Curtis, D. Frurip and M. Blander, J. Amer. Chem. Soc., 1978, **71**, 2703.

(79) G. Alagona, C. Ghio and P. Kollman, "Bifurcated vs. Linear H-Bonds", J. Amer. Chem. Soc., (in press).

(80) P. Kollman, J. McKelvey, A. Johansson and S. Rothenberg, J. Amer. Chem. Soc., 1975, **97**, 955.

(81) H. Umeyama and K. Morokuma, J. Amer. Chem. Soc., 1977, **99**, 1316.

(82) G. Dodson, E. Dodson, D. Hodgkin and C. Reynolds, Can. J. Biochem., 1979, **57**, 469.

(83) Protein Data Bank, Chemistry Dept., Brookhaven National Laboratory, Upton, New York, 11973.

(84) H. Hartmann, F. Parak, W. Steigemann, G. Petsko, D. Ponzi and H. Frauenfelder, Proc. Natl. Acad. Sci., 1982, **79**, 4967.

(85) P. Kollman, D. Case, S. Profeta, Jr. and Peter Murray-Rust, unpublished normal mode calculations on 17-OH progesterone.

(86) D. Case and D. Nguyen unpublished results.

(87) F. Momany, J. Phys. Chem., 1978, **82**, 592.

(88) P. Smit, J. Derissen and F. B. van Duijneveldt, Mol. Phys., 1979, 37, 521.

(89) B. Hingerty and S. Broyde, Biochem., 1982, **21**, 3243.

(90) R. Tilton, P. Weiner and P. Kollman, Biopolymers, 1983, **22**, 969.

(91) W. Olson, Biopolymers, 1975, **14**, 1775,1797.

(92) M. Levitt and A. Warshel, J. Amer. Chem. Soc., 1978, **100**, 2607.

(93) W. Olson in Topics in Nucleic Acid Structure, vol. II, S. Neidle, ed., McMullian, London, 1982.

(94) T. Oie, G. Maggiora and R. Christoffersen, Int. Jour. Quant. Chem.: Quan. Biol. Sym., 1981, p. 1.

Appendix

Bond Parameters

Bond	K_r	r_{eq}
C -C2	317.	1.522
C -C3	317.	1.522
C -CB	447.	1.419
C -CD	469.	1.40
C -CH	317.	1.522
C -CJ	410.	1.444
C -CM	410.	1.444
C -CT	317.	1.522
C -N	490.	1.335
C -N*	424.	1.383
C -NA	418.	1.388
C -NC	457.	1.358
C -O	570.	1.229
C -O2	656.	1.25
C -OH	450.	1.364
C2-C*	317.	1.495
C2-C2	260.	1.526
C2-C3	260.	1.526
C2-N	337.	1.449
C2-N2	337.	1.463
C2-N3	367.	1.471
C2-NT	367.	1.471
C2-OH	386.	1.425
C2-OS	320.	1.425
C2-S	222.	1.81
C2-SH	222.	1.81
C3-CM	317.	1.51
C3-N	337.	1.449
C3-N*	337.	1.475
C3-N2	337.	1.463
C3-N3	367.	1.471
C3-OS	320.	1.425
C3-S	222.	1.81
CA-C2	317.	1.51
CA-CB	469.	1.404
CA-CD	469.	1.40
CA-CJ	427.	1.433
CA-N2	481.	1.340
CA-NA	427.	1.381
CA-NC	483.	1.339
CB-C*	388.	1.459
CB-CB	520.	1.370
CB-CD	469.	1.40
CB-CN	447.	1.419
CB-N*	436.	1.374

CB-NB	414.	1.391
CB-NC	461.	1.354
CC-C2	317.	1.504
CC-CF	512.	1.375
CC-CG	518.	1.371
CC-NA	422.	1.385
CC-NB	410.	1.394
CD-CD	469.	1.40
CD-CN	469.	1.40
CE-N*	440.	1.371
CE-NB	529.	1.304
CF-NB	410.	1.394
CG-C*	546.	1.352
CG-NA	427.	1.381
CH-C2	260.	1.526
CH-C3	260.	1.526
CH-CH	260.	1.526
CH-N	337.	1.449
CH-N*	337.	1.475
CH-NT	367.	1.471
CH-OH	386.	1.425
CH-OS	320.	1.425
CI-NC	502.	1.324
CJ-CJ	549.	1.350
CJ-CM	560.	1.343
CJ-N*	448.	1.365
CN-NA	428.	1.38
CP-NA	477.	1.343
CP-NB	488.	1.335
CT-CT	310.	1.526
CT-HC	331.	1.09
CT-N	337.	1.449
H -N	434.	1.01
H -N2	434.	1.01
H -NA	434.	1.01
H2-N	434.	1.01
H2-N2	434.	1.01
H2-NT	434.	1.01
H3-N2	434.	1.01
H3-N3	434.	1.01
HO-OH	553.	0.96
HO-OS	553.	0.96
HS-SH	274.	1.336
LP-S	600.	0.679
LP-SH	600.	0.679
O2-P	525.	1.48
OH-P	230.	1.61
OS-P	230.	1.61
S -S	166.	2.038

Appendix

Angle Parameters

Angle	K_ϑ	ϑ_{eq}
C*-CG-NA	70.0	108.7
C -C2-C2	63.0	112.4
C -C2-CH	63.0	112.4
C -C2-N	80.0	110.3
C -CB-CB	85.0	119.2
C -CB-NB	70.0	130.0
C -CH-C2	63.0	111.1
C -CH-C3	63.0	111.1
C -CH-CH	63.0	111.1
C -CH-N	63.0	110.1
C -CJ-CJ	85.0	120.7
C -CM-C3	85.0	119.7
C -CM-CJ	85.0	120.7
C -N*-CH	70.0	117.6
C -N*-CJ	70.0	121.6
C -N -C2	50.0	121.9
C -N -C3	50.0	121.9
C -N -CH	50.0	121.9
C -N -CT	50.0	121.9
C -N -H	35.0	119.8
C -N -H2	35.0	120.0
C -NA-C	70.0	126.4
C -NA-CA	70.0	125.2
C -NA-H	35.0	116.8
C -NC-CA	70.0	120.5
C -OH-HO	35.0	113.0
C2-C -N	70.0	116.6
C2-C -O	80.0	120.4
C2-C -O2	70.0	117.0
C2-C2-C2	63.0	112.4
C2-C2-N	80.0	111.2
C2-C2-N2	80.0	111.2
C2-C2-N3	80.0	111.2
C2-C2-OS	80.0	109.5
C2-C2-S	50.0	114.7
C2-CC-CF	70.0	131.9
C2-CC-CG	70.0	129.05
C2-CC-NA	70.0	122.2
C2-CC-NB	70.0	121.05
C2-CH-C3	63.0	111.5
C2-CH-N*	80.0	109.5
C2-CH-OH	80.0	109.5
C2-CH-OS	80.0	109.5
C2-N -H	38.0	118.4
C2-N2-H2	35.0	118.4

C2-N2-H3	35.0	118.4
C2-N3-H3	35.0	109.5
C2-NT-H2	35.0	109.5
C2-OH-HO	55.0	108.5
C2-OS-C2	100.0	111.8
C2-OS-C3	100.0	111.8
C2-OS-HO	55.0	108.5
C2-OS-P	100.0	120.5
C2-S -C3	62.0	98.9
C2-S -S	68.0	103.7
C2-SH-HS	44.0	96.0
C3-C -N	70.0	116.6
C3-C -O	80.0	120.4
C3-C -O2	70.0	117.0
C3-C2-OS	80.0	109.5
C3-CH-C3	63.0	111.5
C3-CH-N	80.0	109.5
C3-CH-OH	80.0	109.5
C3-N -H	38.0	118.4
C3-N2-H2	35.0	118.4
C3-N3-H3	35.0	109.5
C3-OH-HO	55.0	108.5
C3-OS-P	100.0	120.5
C3-S -S	68.0	103.7
C3-SH-HS	44.0	96.0
CA-CB-CB	85.0	117.3
CA-CB-NB	70.0	132.4
CA-CD-CD	85.0	120.0
CA-CJ-CJ	85.0	117.0
CA-N2-C2	50.0	123.2
CA-N2-C3	50.0	123.2
CA-N2-H	35.0	120.0
CA-N2-H2	35.0	120.0
CA-N2-H3	35.0	120.0
CA-NA-H	35.0	118.0
CA-NC-CB	70.0	112.2
CA-NC-CI	70.0	118.6
CB-C*-C2	70.0	128.6
CB-C*-CG	85.0	106.4
CB-C -NA	70.0	111.3
CB-C -O	80.0	128.8
CB-CA-N2	70.0	123.5
CB-CA-NC	70.0	117.3
CB-CB-N*	70.0	106.2
CB-CB-NB	70.0	110.4
CB-CB-NC	70.0	127.7
CB-CD-CD	85.0	120.0
CB-CN-CD	85.0	122.7
CB-CN-NA	70.0	104.4
CB-N*-C3	70.0	125.8
CB-N*-CE	70.0	105.4

CB-N*-CH	70.0	125.8
CB-NB-CE	70.0	103.8
CB-NC-CI	70.0	111.0
CC-C2-CH	63.0	113.1
CC-CF-NB	70.0	109.9
CC-CG-NA	70.0	105.9
CC-NA-CP	70.0	107.3
CC-NA-H	35.0	126.35
CC-NB-CP	70.0	105.3
CD-C -CD	85.0	120.0
CD-C -OH	70.0	120.0
CD-CA-C2	70.0	120.0
CD-CA-CD	85.0	120.0
CD-CB-C*	85.0	134.9
CD-CB-CN	85.0	116.2
CD-CD-C	85.0	120.0
CD-CD-CD	85.0	120.0
CD-CD-CN	85.0	120.0
CD-CN-NA	70.0	132.8
CE-N*-C3	70.0	128.8
CE-N*-CH	70.0	128.8
CF-CC-NA	70.0	105.9
CF-NB-CP	70.0	105.3
CG-C*-C2	70.0	125.0
CG-CC-NA	70.0	108.75
CG-CC-NB	70.0	109.9
CG-NA-H	35.0	126.35
CH-C -N	70.0	116.6
CH-C -O	80.0	120.4
CH-C -O2	65.0	117.0
CH-C -OH	70.0	115.0
CH-C2-C*	63.0	115.6
CH-C2-C2	63.0	112.4
CH-C2-C3	63.0	112.4
CH-C2-CA	63.0	114.0
CH-C2-CH	63.0	112.4
CH-C2-OH	80.0	109.5
CH-C2-OS	80.0	109.5
CH-C2-S	50.0	114.7
CH-C2-SH	50.0	108.6
CH-CH-C2	63.0	111.5
CH-CH-C3	63.0	111.5
CH-CH-CH	63.0	111.5
CH-CH-N*	80.0	109.5
CH-CH-OH	80.0	109.5
CH-CH-OS	80.0	109.5
CH-N -C2	50.0	118.0
CH-N -H	38.0	118.4
CH-NT-H2	35.0	109.5
CH-OH-HO	55.0	108.5
CH-OS-CH	100.0	111.8

CH-OS-HO	55.0	108.5
CH-OS-P	100.0	120.5
CJ-C -NA	70.0	114.1
CJ-C -O	80.0	125.3
CJ-CA-N2	70.0	120.1
CJ-CA-NC	70.0	121.5
CJ-CJ-N*	70.0	121.2
CJ-CM-C3	85.0	119.7
CJ-N*-CH	70.0	121.2
CM-C -NA	70.0	114.1
CM-C -O	80.0	125.3
CM-CJ-N*	70.0	121.2
CN-CB-C*	85.0	108.8
CN-NA-CG	70.0	111.6
CN-NA-H	35.0	124.2
CP-NA-CG	70.0	107.3
CP-NA-H	35.0	126.35
CT-C -N	70.0	116.6
CT-C -O	80.0	120.4
CT-N -H	38.0	118.4
H -N -H	35.0	120.0
H2-N2-H2	35.0	120.0
H3-N2-H3	35.0	120.0
H2-NT-H2	35.0	109.5
H3-N -H3	35.0	120.0
H3-N3-H3	35.0	109.5
HO-OH-HO	47.0	104.5
HO-OH-P	45.0	108.5
LP-S -C2	600.0	96.7
LP-S -C3	600.0	96.7
LP-S -LP	600.0	160.0
LP-S -S	600.0	96.7
LP-SH-C3	600.0	96.7
LP-SH-HS	600.0	96.7
LP-SH-LP	600.0	160.0
N*-C -O	80.0	120.9
N*-CE-NB	70.0	113.9
N*-CH-OS	80.0	109.5
N -C -O	80.0	122.9
N -CH-C2	80.0	109.7
N -CH-CH	80.0	109.7
N2-CA-N2	70.0	120.0
NA-C -N*	70.0	115.4
NA-C -O	80.0	120.6
NA-CA-N2	70.0	116.0
NA-CA-NC	70.0	123.3
NA-CP-NA	70.0	110.75
NB-CP-NA	70.0	111.6
NC-C -N*	70.0	118.6
NC-C -O	80.0	122.5
NC-CA-N2	70.0	119.8

NC-CB-N*	70.0	126.0
NC-CI-NC	70.0	129.1
NT-C2-C	80.0	111.2
NT-C2-C2	80.0	111.2
NT-CH-C	80.0	109.7
NT-CH-C2	80.0	109.7
NT-CH-CH	80.0	109.7
O2-C -O2	80.0	126.0
O2-P -O2	140.0	119.9
OH-P -O2	45.0	108.2
OS-P -O2	100.0	108.2
OS-P -OH	45.0	102.6
OS-P -OS	45.0	102.6

Appendix

Torsional Parameters

Torsion	$V_n/2$	γ	n
X-C*-C2-X	0.0	0	2
X-C*-CB-X	2.4	180	2
X-C*-CG-X	23.6	180	2
X-C -CB-X	4.4	180	2
X-C -CD-X	5.3	180	2
X-C -CH-X	0.0	0	2
X-C -CJ-X	3.1	180	2
X-C -CM-X	3.1	180	2
X-C -CT-X	0.0	0	2
X-C -N*-X	5.8	180	2
X-C -N -X	10.0	180	2
X-C -NA-X	5.4	180	2
X-C -NC-X	8.0	180	2
X-C -OH-X	1.8	180	2
X-C2-C -X	0.0	180	3
X-C2-C2-X	2.0	0	3
X-C2-CC-X	0.0	0	2
X-C2-N -X	0.0	0	3
X-C2-N2-X	0.0	0	3
X-C2-N3-X	1.4	0	3
X-C2-OH-X	0.5	0	3
X-C2-OS-X	1.45	0	3
X-C2-S -X	1.0	0	3
X-C2-SH-X	0.75	0	3
X-CA-C2-X	0.0	0	2
X-CA-CB-X	5.1	180	2
X-CA-CD-X	5.3	180	2
X-CA-CJ-X	3.7	180	2
X-CA-N2-X	6.8	180	2
X-CA-NA-X	6.0	180	2
X-CA-NC-X	9.6	180	2
X-CB-CB-X	16.3	180	2
X-CB-CD-X	5.3	180	2
X-CB-CN-X	4.4	180	2
X-CB-N*-X	6.6	180	2
X-CB-NB-X	5.1	180	2
X-CB-NC-X	8.3	180	2
X-CC-CF-X	14.3	180	2
X-CC-CG-X	15.9	180	2
X-CC-NA-X	5.6	180	2
X-CC-NB-X	4.8	180	2
X-CD-CD-X	5.3	180	2
X-CD-CN-X	5.3	180	2
X-CE-NB-X	20.0	180	2
X-CF-NB-X	4.8	180	2

X-CG-NA-X	6.0	180	2
X-CH-C2-X	2.0	0	3
X-CH-CH-X	2.0	0	3
X-CH-N*-X	0.0	0	2
X-CH-N -X	0.0	0	3
X-CH-OH-X	0.5	0	3
X-CH-OS-X	1.45	0	3
X-CI-NC-X	13.5	180	2
X-CJ-CJ-X	24.4	180	2
X-CJ-CM-X	27.2	180	2
X-CN-NA-X	6.1	180	2
X-CP-NA-X	9.3	180	2
X-CP-NB-X	10.0	180	2
X-CT-CT-X	2.0	0	3
X-CT-N -X	0.0	0	3
X-N*-CE-X	6.7	180	2
X-N*-CJ-X	7.4	180	2
X-NT-C2-X	1.0	0	3
X-OH-P -X	0.75	0	3
X-OS-P -X	0.75	0	3
C2-C2-S -LP	0.0	0	3
C2-OS-C2-C2	0.10	0	2
C2-OS-C2-C2	1.45	0	3
C2-OS-C2-C3	0.1	0	2
C2-OS-C2-C3	0.725	0	3
C2-OS-CH-C2	0.1	0	2
C2-OS-CH-C2	0.725	0	3
C2-OS-CH-C3	0.1	0	2
C2-OS-CH-C3	0.725	0	3
C3-OS-C2-C3	0.10	0	2
C3-OS-C2-C3	1.45	0	3
C3-OS-CH-C3	0.1	0	2
C3-OS-CH-C3	0.725	0	3
C2-S -S -C2	0.6	0	3
C2-S -S -C2	3.5	0	2
CH-C2-SH-LP	0.0	0	3
CH-OS-CH-C2	0.1	0	2
CH-OS-CH-C2	0.725	0	3
CH-OS-CH-CH	0.1	0	2
CH-OS-CH-CH	0.725	0	3
CH-OS-CH-N*	0.0	0	2
CH-OS-CH-N*	0.725	0	3
CT-CT-C -O	0.067	180	3
H-N -C -O	0.65	0	1
H-N -C -O	2.5	180	2
HC-CT-C -O	0.067	180	3
LP-S -S -C2	0.0	0	3
LP-S -S -LP	0.0	0	3
N-CT-C -O	0.067	180	3
O-C -C2-N	0.2	180	3
O-C -CH-C2	0.1	180	3

O-C -CH-CH	0.1	180	3
O-C -CH-N	0.1	180	3
OH-C2-C2-OH	0.5	0	2
OH-C2-C2-OH	2.0	0	3
OH-C2-CH-OH	0.5	0	2
OH-C2-CH-OH	1.0	0	3
OH-CH-CH-OH	0.5	0	3
OH-CH-CH-OH	0.5	0	2
OH-P -OS-C2	0.25	0	3
OH-P -OS-C2	0.75	0	2
OH-P -OS-C3	0.25	0	3
OH-P -OS-C3	0.75	0	2
OH-P -OS-CH	0.25	0	3
OH-P -OS-CH	0.75	0	2
OS-C2-C2-OH	0.5	0	2
OS-C2-C2-OH	2.0	0	3
OS-C2-C2-OS	0.5	0	2
OS-C2-C2-OS	2.0	0	3
OS-C2-CH-OH	0.5	0	2
OS-C2-CH-OH	1.0	0	3
OS-C2-CH-OS	0.5	0	2
OS-C2-CH-OS	1.0	0	3
OS-CH-C2-OH	0.5	0	2
OS-CH-C2-OH	1.0	0	3
OS-CH-CH-OH	0.5	0	3
OS-CH-CH-OH	0.5	0	2
OS-CH-CH-OS	0.5	0	3
OS-CH-CH-OS	0.5	0	2
OS-P -OS-C2	0.25	0	3
OS-P -OS-C2	0.75	0	2
OS-P -OS-C3	0.25	0	3
OS-P -OS-C3	0.75	0	2
OS-P -OS-CH	0.25	0	3
OS-P -OS-CH	0.75	0	2

Appendix

Improper Torsional Parameters

Torsion	$V_n/2$	γ	n
X -X -C -O	10.5	180.0	2.
X -X -N -H	1.0	180.0	2.
X -X -NA-H	1.0	180.0	2.
X -C2-CH-X	14.0	180.0	3.
X -CH-CH-X	14.0	180.0	3.
X -CH-N -C2	1.0	180.0	2.
X -CH-N -C	14.0	180.0	3.
X -H2-N -H2	1.0	180.0	2.
X -N2-CA-N2	10.5	180.0	2.
X -O2-C -O2	10.5	180.0	2.
C2-CH-C -N3	7.0	180.0	3.
C3-CH-CA-C3	7.0	180.0	3.
CH-CH-C -N3	7.0	180.0	3.

Appendix

Hydrogen Bond Parameters

Acceptor	Donor	C	D
H	NB	7557	2385
H	NC	10238	3071
H	O2	4019	1409
H	O	7557	2385
H	OH	7557	2385
H	S	265720	35029
H	SH	265720	35029
HO	NB	7557	2385
HO	O2	4019	1409
HO	O	7557	2385
HO	OH	7557	2385
HO	S	265720	35029
HO	SH	265720	35029
H2	NB	4019	1409
H2	O2	4019	1409
H2	O	10238	3071
H2	OH	4019	1409
H2	S	7557	2385
H2	SH	7557	2385
H3	NB	4019	1409
H3	O2	4019	1409
H3	O	4019	1409
H3	OH	4019	1409
H3	S	7557	2385
H3	SH	7557	2385
HS	NB	14184	3082
HS	O2	4019	1409
HS	O	14184	3082
HS	OH	14184	3082
HS	S	265720	35029
HS	SH	265720	35029

Appendix

Non-Bonded Parameters

Atom	R*	ε
C	1.85	0.12
C*	1.85	0.12
C2	1.925	0.12
C3	2.00	0.15
CA	1.85	0.12
CB	1.85	0.12
CC	1.85	0.12
CD	1.85	0.12
CE	1.85	0.12
CF	1.85	0.12
CG	1.85	0.12
CH	1.85	0.09
CI	1.85	0.12
CJ	1.85	0.12
CM	1.85	0.12
CN	1.85	0.12
CP	1.85	0.12
CT	1.80	0.06
H	1.00	0.02
H2	1.00	0.02
H3	1.00	0.02
HC	1.375	0.038
HO	1.00	0.02
HS	1.00	0.02
LP	1.20	0.016
N	1.75	0.16
N*	1.75	0.16
N2	1.75	0.16
N3	1.85	0.08
NA	1.75	0.16
NB	1.75	0.16
NC	1.75	0.16
NT	1.85	0.12
O	1.60	0.20
O2	1.60	0.20
OH	1.65	0.15
OS	1.65	0.15
P	2.10	0.20
S	2.00	0.20
SH	2.00	0.20

CHAPTER 2

ABSTRACT

Molecular mechanical and *ab initio* (4-31G) calculations on N^{α}-acetyl-N-methylalaninamide have been carried out. At the molecular mechanical level, five local energy minima have been found and the free energies have also been determined for these five structures. Addition of a dispersion energy term to the SCF quantum mechanical energies has been shown to effect the relative energies of these local minima and to improve the agreement between the quantum mechanical and molecular mechanical relative energies. The structural properties of the five conformers calculated at the molecular mechanical and quantum mechanical level are in generally good agreement.

INTRODUCTION

Many theoretical[1-7] and experimental[8-11] studies have been performed on the potential energy surface of alanyl dipeptide. Most recently, Scarsdale *et al*[3] have performed *ab initio* quantum mechanical (4-21G basis set) geometry optimization on seven N^a-acetyl-N-methylalaninamide (NANA) structures and have compared their results to ECEPP molecular mechanical calculations. In our opinion, their work has given a somewhat incomplete view of the "state of the art" of molecular mechanical calculations on this molecule. Thus, the goals of this paper are threefold: First, to suggest that "state of the art" molecular mechanical and quantum mechanical calculations are in *qualitative* agreement with each other, provided that one relaxes all geometrical degrees of freedom and considers dispersion attraction; secondly, to evaluate basis set differences at the *ab initio* level on the relative energies of some of the conformations of NANA; and, finally, to present relative free energies of the local minima on the NANA surface.

METHODS

To explore the NANA potential energy surface, we used our molecular mechanics program AMBER[12] and the technique of adiabatic mapping; where the Φ, Ψ dihedrals were constrained to fixed values and the remaining degrees of freedom allowed to relax. We have only considered trans peptide units in this study. The potential energy function, parameter set and Φ, Ψ plot appear in reference 6. All low energy structures were subjected to further refinement using a second derivative Newton-Raphson algorithm to assess whether they were local minima. The alanyl dipeptide unit appears in figure 1. Two local minima conformations correspond to 1-7 hydrogen bonded systems between O6 and H18. These are $\Phi \approx -75°$, $\Psi \approx 65°$ (C_7^{eq}) and

$\Phi \approx 70°$, $\Psi \approx -65°$ (C_7^{ax}). A third local minima conformation, with a 1-5 hydrogen bond between O16 and H8, (C_5) has $\Phi \approx -160°$, $\Psi \approx 170°$. Two other local minima found are characterized by $\Phi \approx -60°$, $\Psi \approx -40°$ (C_{10}^R) and $\Phi \approx 55°$, $\Psi \approx 45°$ (C_{10}^L). The geometries of the three lowest energetic local minima (C_7^{eq}, C_7^{ax} and C_5) were taken as starting structures for quantum mechanical *ab initio* (4-31G basis set) gradient optimization using the program Gaussian 80 U.C.S.F.[13]. To evaluate the dispersion correction (DC) term for the quantum mechanical results, a $\frac{-B}{R^6}$ term was used (B values appear in reference 6). We have reduced the 1-4 dispersion interactions by a factor of 2, consistent with our molecular mechanical model; this reduction of short distance dispersion attraction also has precedence in ref. 14. All calculations were performed on VAX 11-780's, with each quantum mechanical calculation taking ≈ 20 hours for SCF + gradient evaluation.

RESULTS AND DISCUSSION

The results of our molecular mechanical and quantum mechanical calculations appear in Tables 1-3. Since *ab initio* SCF calculations do not contain dispersion energy[15], which could lead to differences in the conformational energies, we have amended the Scarsdale *et al* [3] results with a dispersion energy correction (DC) (Table 1). The DC for the 4-21G alanyl structures is -10.8, -11.5, -9.7, -9.6, and -11.2 kcal/mole for C_7^{eq}, C_7^{ax}, C_5, C_{10}^R and C_{10}^L respectively (summing over all nonbonded interaction pairs).

The addition of these DC values to the 4-21G *ab initio* energies significantly improves the agreement between the quantum mechanical and molecular mechanical relative energies. First, the ordering of the C_7^{ax} and C_5 structures, relative to C_7^{eq}, is reversed (Table 1). Stern *et al*[5] and our gas phase

molecular mechanical models (ε = 1) find C_7^{eq} to be more stable than C_7^{ax} by 1.1-1.3 kcal/mole, in reasonable agreement with (4-21G) QM + DC of 1.7 kcal/mole. The $C_7^{eq} - C_5$ energy difference found in our molecular mechanical model is not as close to the QM + DC value (4.0 vs. 2.5 kcal/mole)[16]. Finally, the other two local minima found here are reasonably close in energy to the corresponding *ab initio* ones, although for C_{10}^R, the *ab initio* structure differs somewhat from the molecular mechanical one, likely due to the shallowness of the potential surface in this region[3].

It is clear that the dispersion correction is an important addition to the *ab initio* relative energies. Other evidence for the importance of DC in evaluating relative energies comes from a study of the corresponding glycyl dipeptide C_7 and C_5 structures at the *ab initio* 4-21G level, where an energy difference of 0.8 kcal/mole was found. For the glycyl dipeptide the DC is -8.5 kcal/mole for C_7 and -6.7 kcal/mole for C_5. When the DC is added to these structures, the energy difference becomes 2.5 kcal/mole, in closer agreement with the molecular mechanics (ε = 1) difference of 3.9 kcal/mole. While we cannot claim that our dispersion correction is quantitatively accurate[14], it is likely that the C_7^{ax} and C_7^{eq} structures will be stabilized relative to C_5 no matter what the dispersion coefficients. Further support for the reasonableness of the approach taken here is found in the work of Prissette and Kochanski[17]. They have shown that for different configurations of $(Cl_2)_2$, a simple atom centered $\frac{-B}{R^6}$ dispersion term does an excellent job of reproducing a more complete quantum mechanical calculation of dispersion attraction between the two chlorine molecules.

Scarsdale *et al*[3] pointed out that the molecular mechanical (ECEPP) calculations of Zimmerman *et al*[2] were inconsistent with their quantum mechanical calculations, because ECEPP finds C_7^{ax} to be 8.8 kcal/mole higher in energy than C_7^{eq}, compared with only 1.7 kcal/mole for the (4-21G) QM + DC results. This difference is caused by ECEPP's use of fixed bond lengths and angles, since molecular mechanical results which allow full relaxation of these parameters are consistent with the quantum mechanical values[4-6]. The rigid geometry *ab initio* STO-3G energies of Peters and Peters[7] also differ significantly from the relaxed geometry *ab initio* and molecular mechanical values.

We had begun a 4-31G gradient optimization on the C_7^{eq}, C_7^{ax} and C_5 structures of NANA, beginning with the molecular mechanics refined geometries, before the Scarsdale *et al*[3] work appeared (Table 1). The relative energies of C_7^{eq} and C_5 before dispersion correction are surprisingly different (0.2 kcal/mole 4-31G, 1.4 kcal/mole 4-21G), whereas the $C_7^{eq} - C_7^{ax}$ differences are nearly identical (2.4 kcal/mole 4-31G, 2.6 kcal/mole 4-21G). We confirmed that this $C_7^{eq} - C_5$ difference was *not* mainly due to geometric differences by carrying out 4-31G single point *ab initio* calculations on C_5 and C_7^{eq} at the 4-21G geometries, finding $\Delta E_{C_7^{eq}-C_5}$ = 0.6 kcal/mole. Thus, the remaining energy difference of 0.8 kcal/mole between such similar basis sets remains a mystery[18]. The 4-31G *ab initio* energy difference between C_5 and C_7^{ax} is large enough that even dispersion correction does not reverse the order of stabilities, even though it brings them closer.

We have compared the bond lengths, bond angles and dihedral angles found in the molecular mechanical and 4-31G and 4-21G quantum mechanical cal-

culations on the various conformations of NANA. The calculated bond lengths are insensitive to conformation and are very similar in the three sets of calculations, with standard deviations among the calculated values for given bonds ranging from 0.003 - 0.01 . For the bond angles (looking now only at the three conformations C_7^{ax}, C_7^{eq} and C_5 for each angle; this corresponds to 36 angles for each conformation for a total of 108 total bond angles) 77 are found with standard deviations less than 1.0°, 30 between 1.0° and 2.0° and 1 with a standard deviation of 2.5°. The largest conformational dependent angle differences are around the C_α carbon, and, thus, we report these in detail in Table 2.

Focusing on the C_α angles for the three lowest energy structures (Table2), we see that the agreement between the quantum mechanical and molecular mechanical calculated values is good for those angles not involving hydrogens (N-C_α-C_β, N-C_α-C and C_β-C_α-C) with the order of the angles identical for both calculations and the values of the angles differing on average by 1.0° (largest discrepancy being 2°). There are larger differences between the calculated angles involving hydrogens (N-C_α-H_α, C_β-C_α-H_α and C-C_α-H_α), with an average difference of 1.7° and the largest difference 4.1°. Given that the force field reported in ref. 6 has been optimized for united atoms, and not as yet at the all atom level, the agreement is still reasonable. It is likely that the quantum mechanical calculations are more quantitatively accurate in the calculations of these angles, but it is not clear that the extra accuracy is worth the computational price here, given the 3-4 order of magnitude difference in computer time involved in the two types of calculations.

The Φ, Ψ and peptide (ω) dihedral angles are compared in Table 3. The Φ, Ψ values are qualitatively similar for a given conformation with the exception of C_{10}^{R}. As Scarsdale *et al*[3] point out, however, the potential surface as a function of these dihedral angles is very shallow and the 4-21G and 4-31G structures are not necessarily at true local minima. The two ω dihedral angles are similar and differ by at most 6° between 4-21G and molecular mechanics, but the two methods do not always agree on the sign of the deviation from planarity.

A final factor which must be considered in relating the calculated results to experimental confomer populations are entropy differences. We can approximate the relative gas phase entropies and free energies for the molecular mechanics models within the rigid rotor, harmonic oscillator approximations for various conformations[20]. The entropic contribution at 298K helps stabilize the C_5 conformation, relative to either C_7 by ≈ 1 kcal/mole (Table 1).

DISCUSSION

It is clear that there is no major qualitative discrepancy between the dispersion corrected *ab initio* results and the molecular mechanical for gas phase NANA. The gas phase (ε = 1) molecular mechanical results reported by us and Stern *et al*[5] and the 4-21G (and 4-31G) QM + DC find C_7^{eq} lowest in energy, C_7^{ax} higher by 1-2 kcal/mole, C_5 higher by 2-4 kcal/mole and C_{10}^{R}, C_{10}^{L} higher by 5-7 kcal/mole. The largest quantitative discrepancy between the relative molecular mechanical and quantum mechanical energies is the magnitude of the energy difference between C_7^{eq} and C_5 which is 1.9-2.5 kcal/mole at the *ab initio* level and 3-4 kcal/mole at the molecular mechanical. The quantum mechanical calculations still suffer from basis set dependence and too simple

dispersion correction, so they are not yet a definitive reference point. Experimental data on NANA in non-polar solvents is most consistent with a non negligible fraction of C_5, suggesting that the quantum mechanical values are closer to correct. On the other hand, we have demonstrated that both entropy effects and dielectric constant changes profoundly effect the magnitude of the $C_7^{gg} - C_5$ energy difference at the molecular mechanical level, with the C_7^{gg} and C_5 of approximately equal free energies with $\varepsilon = 4$. This, combined with the uncertainty in the experimental data in non-polar solvents, precludes a definitive evaluation of how far from the truth these various calculated energy differences are. Thus, it is important to perform further gas phase experiments on this most important molecule to definitively establish its conformational equilibrium.

ACKNOWLEDGEMENTS: PAK would like to thank the NIH (GM-29072) for research support.

Table 1
Structures and Energies for Alanyl Dipeptide

	FF2[a] $\varepsilon = 1$	FF2[a] $\varepsilon = 4$	FF2[a] $\varepsilon = R_{i,j}$	4–31G[b] QM+DC	4–21G[c] QM+DC	ECEPP[d] $\varepsilon = 2$	*Ref.* 5[e] $\varepsilon = 1$	PCILO[f]
Φ^g	-75.6	-80.4	-75.6	-81.1	-84.6	-84	-80	-78
Ψ^g	68.5	69.0	65.6	66.3	73.0	79	80	40
ΔE^h	0.0	0.0	0.0	0.0(0.0)	0.0(0.0)	0.0	0.0	0.3
ΔG^h	0.0	0.0	0.0	-	-	-	-	-
Φ	68.6	72.1	69.0	73.6	74.6	78	70	75
Ψ	-67.4	-68.2	-64.4	-60.4	-62.0	-64	-80	-40
ΔE	1.1	0.7	0.6	2.2(2.4)	1.7(2.6)	8.8	1.3	0.0
ΔG	1.2	0.9	0.7	-	-	-	-	-
Φ	-157.6	-159.3	-160.7	-161.8	-165.7	-154	-	-171
Ψ	169.1	166.1	168.6	167.3	167.3	153	-	164
ΔE	4.0	1.2	3.2	1.9(0.2)	2.5(1.4)	0.4	-	1.6
ΔG	2.9	0.3	2.1	-	-	-	-	-
Φ	-60.3	-65.2	-60.7	-	-91.9	-74	-60	-29
Ψ	-34.2	-50.1	-40.7	-	-5.5	-45	-40	-59
ΔE	4.7	1.2	3.6	-	6.1(4.9)	1.1	4.2	2.4
ΔG	3.6	0.3	2.6	-	-	-	-	-
Φ	53.5	55.8	53.9	-	60.8	54	60	-
Ψ	36.5	50.8	41.6	-	40.6	57	40	-
ΔE	5.7	1.8	4.3	-	6.3(6.7)	2.3	7.3	-
ΔG	4.9	1.3	3.6	-	-	-	-	-

a Molecular mechanical simulations using the program AMBER (ref. 12) and second derivative techniques to assure that each of the five structures were "true" local minima. FF2 parameters appear in ref. 6. The dielectric constant used appears in each column as ε.

b This study, gradient optimized using Gaussian 80 U.C.S.F. (ref. 13). The total energy for C_7^{gg} = -492.133696 a.u. The largest gradient component for the three structures C_7^{gg}, C_7^{gx} and C_5 are 0.0015, 0.0068 and 0.0013 a.u. respectively. Over the last two optimization cycles the energy was reduced 0.04, 0.04 and 0.003 kcal/mole for C_7^{gg}, C_7^{gx} and C_5 respectively. The ΔE = QM + DC in kcal/mol. The values in parenthesis correspond to the QM values.

c Structures are from Scarsdale *et al* (ref. 3).

d Empirical energy calculations by Zimmerman *et al* using fixed bond lengths and angles (ref. 2).

e Molecular mechanics refined structures by Stern *et al*. Each of the Φ, Ψ values were constrained in 10° increments with all other degrees of freedom allowed to relax (ref. 5).

f PCILO results by Pullman *et al* (ref. 1).

g The specific dihedral angles are illustrated in fig. 1. For the proper convention see ref. 6.

h All energies ΔG and ΔE are in kcal/mole.

Table 2
Bond Angles Around C_α in NANA (°)

Struct.	$N-C_\alpha-C_\beta$	$N-C_\alpha-C$	$N-C_\alpha-H_\alpha$	$C_\beta-C_\alpha-C$	$C_\beta-C_\alpha-H_\alpha$	$H_\alpha-C_\alpha-C$
Molecular Mechanics ($\varepsilon=1$)[a]						
C_7^{gg}	110.5	109.8	109.0	112.6	107.0	107.9
C_7^{ax}	112.9	111.1	106.2	113.2	106.2	106.7
C_5	111.1	107.7	108.9	111.9	108.8	108.3
C_{10}^R	110.7	111.4	108.1	111.3	107.7	107.5
C_{10}^L	111.1	111.8	107.1	112.8	107.0	106.6
4-21G[b]						
C_7^{gg}	110.1	109.5	106.8	110.6	111.1	108.6
C_7^{ax}	112.4	112.7	106.0	111.8	109.1	104.2
C_5	111.6	106.4	109.3	110.5	109.3	109.8
C_{10}^R	110.6	114.0	108.0	108.4	109.8	106.0
C_{10}^L	113.3	111.1	105.8	110.6	109.7	106.0
4-31G[a]						
C_7^{gg}	109.7	110.5	107.2	111.3	110.1	108.0
C_7^{ax}	113.3	112.8	105.3	111.9	108.6	104.3
C_5	111.8	107.2	108.8	110.9	108.8	109.3

a This study.

b Ref. 3.

Table 3
Representative Dihedral Angles for NANA (°)

Struct.	CH1–C–N–C_α	C–N–C_α–C	N–C_α–C–N	C_α–C–N–CH2
Molecular Mechanics ($\varepsilon=1$)[a]				
C_7^{eq}	178.4	-75.5	68.5	-179.9
C_7^{ax}	-177.8	68.6	-67.5	-179.8
C_5	177.0	-157.6	169.0	179.6
C_{10}^R	-179.2	-60.3	-34.4	179.5
C_{10}^L	178.8	53.7	36.6	-178.9
4-21G[b]				
C_7^{eq}	-177.0	-84.6	73.0	-179.0
C_7^{ax}	176.0	74.7	-62.0	-178.9
C_5	178.7	-165.7	167.3	178.0
C_{10}^R	-173.1	-91.9	-5.5	179.0
C_{10}^L	174.5	60.8	40.6	178.4
4-31G[a]				
C_7^{eq}	-178.6	-81.1	66.3	-178.7
C_7^{ax}	178.2	73.6	-60.4	-179.8
C_5	178.7	-161.8	167.3	178.8

a This study.

b Ref. 3.

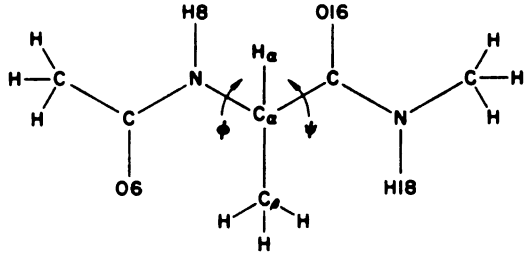


Figure 1.

REFERENCES

(1) B. Pullman and A. Pullman, Adv. Prot. Chem., 1974, **28**, 348.

(2) S. Zimmerman, M. Pottle, G. Nemethy and H. Scheraga, Macromolecules, 1977, **10**, 1.

(3) J. Scarsdale, C. Van Alsenoy, V. Klimkowski, L. Schafer and F. Momany, J. Amer. Chem. Soc., 1983, **105**, 3438.

(4) B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan and M. Karplus, J. Comp. Chem., 1983, **4**, 187.

(5) P. Stern, M. Chorev, M. Goodman and A. Hagler, Biopoly., 1983, **22**, 1885.

(6) S. Weiner, P. Kollman, U. Singh, D. Case, C. Ghio, G. Alagona, S. Profeta and P.Weiner, J. Amer. Chem. Soc., 1984, **106**, 765.

(7) D. Peters and J. Peters, J. Mol. Struct., 1981, **85**, 107.

(8) M. Cung, M. Marraud and J. Neel in "The Jerusalem Symposia on Quantum Chemistry and Biochemistry - Conformation of Biological Molecules and Polymers", E. Bergmann and B. Pullman eds., 1973, Jerusalem, p.69.

(9) M. Avignon and J. Lascombe, ibid, p.97.

(10) M. Avignon and P. Van Huong, Biopoly., 1970, **9**, 427.

(11) A. Burgess and H. Scheraga, ibid., 1973, **12**, 2177.

(12) P. Weiner and P. Kollman, J. Comp. Chem., 1981, **2**, 287.

(13) U.C. Singh and P. Kollman, QCPE Program #446, QCPE Bull., 1982, **2**, 17.

(14) See C. Douketis, G. Scoles, S. Marchetti, M. Zen and A. Thakkar, J. Chem. Phys., 1982, **76**, 3057 for a leading reference to calculate dispersion energy as a series of R^{-6}, R^{-8} and R^{-10} terms.

(15) H. Margenau and N. Kestner, "The Theory of Intermolecular Forces", (1970), Pergamon Press, Oxford.

(16) There is no local minimum in the Hagler *et al* surface for the *c* 5 structure, although such a structure appears to be ≈ 3-4 kcal/mole higher in energy (estimated from their contour plot).

(17) J. Prissette and E. Kochanski, J. Amer. Chem. Soc., 1978, **100**, 6609.

(18) Single point *ab initio* calculations on the Schafer *et al* (ref. 19) glycyl dipeptide structures also give a 0.8 kcal/mole stabilization of *c* 5 relative to C_7 (4-31G basis set compared to the 4-21G results), such that at the 4-31G level with no dispersion correction, both C_7 and *c* 5 are approximately equoenergetic.

(19) L. Schafer, C. Alsenoy and J. Scarsdale, J. Chem. Phys., 1982, **76**, 1439.

(20) A. Hagler, P. Stern, R. Sharon, J. Becker and F. Naider, J. Amer. Chem. Soc., 1979, **101**, 6842.

CHAPTER 3

ABSTRACT

We present the results of a new approach for simulating chemical reactions using quantum mechanical and molecular mechanical methods. This approach is applied to the hydrolysis of formamide by hydroxide ion. In the gas phase, tetrahedral complex (TC) formation is calculated to proceed with no barrier and TC breakdown involves a small barrier (12 kcal/mol). In solution, we calculate a 22 kcal/mol barrier for *formation* of the TC with a second, smaller, barrier occurring for TC breakdown. The calculated reaction energies and activation energies are in quite good agreement with available experimental data.

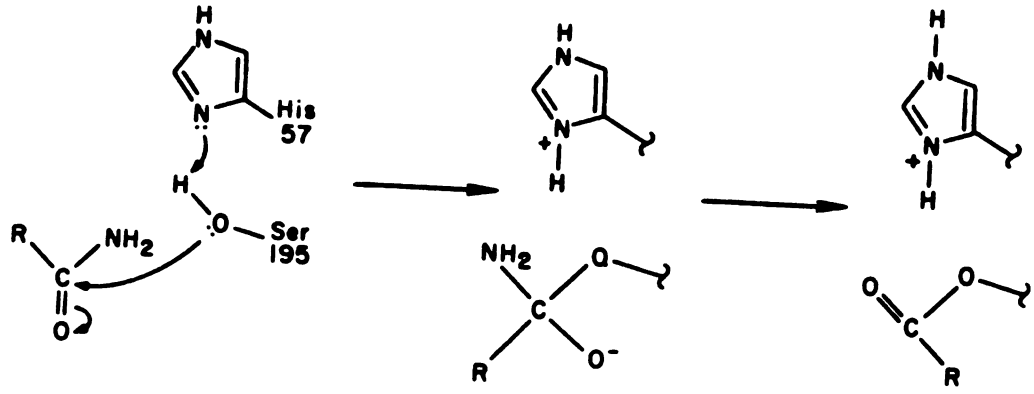
INTRODUCTION

The mechanisms by which enzymes catalyze chemical reactions have intrigued theoretical chemists and biochemists for years[1-4]. Warshel and Levitt's pioneering approach to simulating enzymatic reactions[5], and the application of this approach to lysozyme cleavage of saccharide linkages, was the first study which combined the environmental and internal strain factors using a molecular mechanical model with semi-empirical quantum mechanical techniques to evaluate the energetics of bond breaking. The results of their calculations were encouraging and showed the dramatic effects which electrostatic interactions have in stabilizing the intermediate carbonium ion in this reaction. Although their method has much merit, we feel that recent developments in *ab initio* quantum mechanical theory[6] and accurate potentials for liquid water[7] make it a propitious time to develop another approach for simulating enzymatic reactions.

With this in mind, we present a method for simulating non catalyzed, as well as enzymatic reactions, in aqueous solution. This method can best be broken down into two very general steps: the use of *ab initio* **quantum mechanics** to evaluate bond breaking energies and **molecular mechanics** for calculating the remaining energies, dominated by strain and non-covalent interactions. The solute(s) are completely surrounded by explicit water molecules, taken from a Monte Carlo simulation on liquid water[8], and allowed to energy refine using molecular mechanics. As our first application of this approach, we have chosen to focus on the base catalyzed hydrolysis of formamide:

$$OH^- + H_2NCHO \longrightarrow HCOO^- + NH_3 \qquad (1)$$

This reaction was selected for its close analogy with amide hydrolysis catalyzed by the serine proteases:



The first step in our approach uses *ab initio* quantum mechanical techniques for evaluating the structural, energetic and electronic properties of various "snapshots" along the pathway of formamide hydrolysis in the gas phase.

There have been several previous *ab initio* studies on nucleophilic attack of carbonyl carbon centers. Alagona *et al*[9] have performed *ab initio* minimal basis set (STO-3G) calculations on formamide bond cleavage by hydroxide in

the gas phase and noted the importance of adding solvent to the calculation.

A thorough quantum mechanical analysis of essentially the reverse, nonionic, reaction has been performed by Oie *et al*[10]:

$$NH_3 + HCOOH \longrightarrow NH_2CHO + H_2O \qquad (3)$$

Their study focused on characterizing intermediates and transition states along the reaction pathway and showed the importance of including electronic correlation energy in the analysis of the reaction energetics.

Williams *et al*[11] have extended the theory one step further by including a few ancillary solvent (water) molecules in the quantum mechanical treatment of the reaction:

$$H_2O + CH_2O \longrightarrow HOCH_2OH \qquad (4)$$

The gas phase mechanism was shown to be concerted and catalyzed by a single water molecule. They have shown the importance of including even a few solvent molecules into the calculation and the dramatic stabilizing effect which they can have.

These calculations, and others[12-13], have been valuable in illustrating the power and utility of *ab initio* methods for studying gas phase chemical reactions. However, realistic solvation energies can only be achieved by adding solvent into the calculation on a much larger scale. At this point there is no general agreement on the best way to proceed which is both accurate and computationally practical[14-19].

Placing a few water molecules about the solute cannot lead to a proper evaluation of solvation energies. In particular, the disruption of water-water energies upon incorporation of ionic solutes has been shown to be very important in representing aqueous solvation; this effect is not represented in such "supermolecule" approaches. Monte Carlo[20] and molecular dynamics[21] have been applied quite successfully to studying the properties of bulk water and solvation energies of small molecules. However, these treatments are computationally very expensive for small solutes, let alone a system containing protein and substrate. It is clear that a method is needed which is capable of reasonably evaluating quantitative solvation energies, yet can be applied to enzymatic systems without using enormous amounts of computer time.

In view of the considerations mentioned above, we have combined *ab initio* theory with explicit solvation calculated by a molecular mechanical approach. We have performed *ab initio* calculations on eight "snapshot" structures along the reaction coordinate of hydroxide attack on formamide and subsequent water catalyzed breakdown of the tetrahedral complex (TC). These gas phase structures were placed in a solvent "bath" and energy refined using a molecular mechanical approach. The solvation energy of the system was then determined from these molecular mechanical calculations.

We find that in the gas phase, tetrahedral complex (TC) formation is calculated to be a "downhill" process; with the TC 26 kcal/mol lower than the reactants. A barrier of 12 kcal/mol is found for H_2O catalyzed breakdown of the tetrahedral complex, with the products, ammonia and formate, lying 49 kcal/mol lower than reactants. When solvent is included, a dramatic change

occurred in the reaction profile. The aqueous phase reaction is found to proceed through a solvent induced barrier of 22 kcal/mol to TC formation; with the energetics for H_2O proton donation giving rise to a second barrier. These calculations are in qualitative agreement with experimental results[22] for hydrolysis of amides in basic solutions.

METHODS

Our first goal was to simulate reaction 1 in the gas phase using *ab initio* quantum mechanical techniques. We divided the reaction pathway into two parts. The first focused on OH^- attack on the carbonyl carbon of formamide and subsequent formation of a stable tetrahedral complex (Figures 1-4). We broke this part of the reaction into four distinct steps, each characterized by the distance from the hydroxide oxygen to the carbonyl carbon; the distances being 6.0, 3.08, 2.08 and 1.48 Å. The second part of the pathway represented water catalyzed breakdown of the tetrahedral complex (Figures 5-7). Similarly, three of these "snapshots" are denoted by the distance between the formamide nitrogen and a hydrogen of the incoming water molecule at 1.75, 1.23 and 1.15 Å. The final geometry (Figure 8) corresponds to the products: formate and ammonia separated by 6.0 Å. All of the geometries were refined using a gradient optimization routine at the *ab initio* level (4–31G[23] basis set). Each degree of freedom was allowed to vary, with the sole constraint of restrained C---O (part 1) or N---H (part 2) distances. Subsequently, we carried out single point SCF 4–31+G[24], 6–31G*[25] and 6–31G*/MP2[26] calculations on these eight optimized geometries to assess the effects of basis set dependence and correlation energy upon the reaction profile.

To incorporate the solvent into the calculation, we placed our quantum mechanically optimized structures within a cube of 216 water molecules. This cube was a single snapshot from a Monte Carlo simulation of pure water[8]. The starting geometries were determined by inserting the solute into the mass weighted center of the solvent box. Any water molecule within 1.55 Å of the solute was discarded. For each of the eight structures, two or three solvent molecules were removed to accommodate the solute. The electric field generated by the solute at each water position was calculated and the water molecules reoriented, about their oxygens, by pointing the hydrogens along the direction of the largest electric field component vector. Using our molecular mechanics program AMBER[27], we energy refined these modeled solute-solvent "snapshot" structures. The empirical potential energy function used appears in equation 5.

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\vartheta (\vartheta - \vartheta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\varphi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] + \sum_{H-bonds} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right] \quad (5)$$

The water potentials, A_{ij}, B_{ij}, and q_i, were taken directly from the TIPS3P studies of Jorgensen[7]. Stretching and bending force constants, K_r and K_ϑ, for the water molecules were derived from a best fit of the calculated vibrational frequencies to the experimentally determined ones[28]. Values of q_i for the solute were determined by a fit of quantum mechanically generated electrostatic potential points to a point charge model[29]. A constant dielectric constant (ε = 1) was used. Consistent with the TIPS3P potential for H_2O --- H_2O interactions, no explicit hydrogen bonding function was evaluated. The cartesian coordinates of the solute were constrained, using a harmonic potential, with a weight of 2000 kcal/mol Å^2, while the solute intramolecular

force constants were set equal to zero. These two steps were taken to assure that the internal geometries of the solute stayed fixed to the optimized *ab initio* structures.

Finding a true local minimum in this solute + 216 water molecules system would only be guaranteed after achieving a large number of energy evaluations. Such a minimum would, in any case, correspond to a 0K structure. For these reasons, we chose instead to consider convergence after a fixed number of energy evaluations, to achieve not necessarily a local energy minimum, but rather a reasonable low energy structure for each solute geometry. We eventually decided upon 1000 energy evaluations as a convergence criterion since the rms gradient ≈ 0.2 kcal/Å, with the energy changing only 0.1 kcal per function evaluation. More importantly, the ΔE (difference in solvation energy) *between* any two solvated structures was essentially the same after 600 and 1000 energy evaluations.

The most difficult aspect of this approach was to develop a method for accurately extracting solvation energies from the molecular mechanics refined structures. It has been shown in Monte Carlo simulations of ions in water[30] that there are two predominant energy contributions to the solvation energy of anions or small molecules, solute-solvent ($E_{solute-solvent}$) interactions and the *change* in solvent-solvent interactions upon introduction of the solute ($\Delta E_{solvent-solvent}$).

The solute-solvent interaction energy can be calculated directly from the molecular mechanical interaction energy of solute with all of the solvent molecules (most of this energy comes from waters within 4 Å of the solute). To enable us to evaluate the change in solvent-solvent energy upon

introduction of solute, $\Delta E_{solvent-solvent}$, we performed a simulation on pure water. The difficulty was in how to quantitatively extract the solvent-solvent energy from the molecular mechanics refined structures and to avoid the artifacts caused by edge effects from waters far from the solute. Given the recent results of Chandrasekhar *et al*[30], which showed that the solvent perturbation in ionic solvation is dominated by the first coordination shell, we focused on those waters in the first solvation shell, i.e. closer than the first minimum in the radial distribution function. In all cases the first minimum was well defined and there were five or six water molecules closer to the solute than this minimum. We evaluated the water-water energies for these first coordination waters and compared them with the corresponding water-water energies from the molecular mechanics optimized structure of pure water, calculated to be -24.2 kcal/mol per water molecule. $\Delta E_{solvent-solvent}$ is the difference between these.

In all of our calculations, we assumed that the total energy can be represented as a sum of three terms:

$$E_T = E_{internal\ solute} + E_{solute-solvent} + \Delta E_{solvent-solvent} \qquad (6)$$

$E_{internal\ solute}$ represents the intrinsic energy of various solute structures taken directly from the gas phase quantum mechanical calculations. The later two terms, $E_{solute-solvent}$ and $\Delta E_{solvent-solvent}$, can come from the molecular mechanical energies after molecular mechanics optimization on the solute in the box of water molecules. The total energy for this model $E_T(MM)$ is given by:

$$E_T(MM) = E_{internal\ solute}(QM) + E_{solute-solvent}(MM) + \Delta E_{solvent-solvent}(MM) \qquad (7)$$

However, an alternative approach is to use this geometry and to evaluate the sum of $E_{internal\ solute}$ + $E_{solute-solvent}$ directly using quantum mechanical methods.

The approach which we used for including the electrostatic environment into the quantum mechanical calculation was to represent the solvent molecules in terms of point charges, q_j, situated at their atomic centers. These point charges enter into the quantum mechanical calculation through the one electron Hamiltonian as:

$$h_i = p_i^2 + \sum_{Atom} \frac{Z_A}{r_{Ai}} + \sum_j \frac{q_j}{r_{ji}} \tag{8}$$

where j represents the atoms of the solvent system. In this manner the solute can be studied in the *ab initio* framework and this energy is $E_{internal\ solute}$ (QM + Electrostatic)

In addition to the electrostatic interaction, this model also includes polarization of the solvent on the solute. To calculate the remaining polarization effects (the solute on the solvent and the solvent-solvent polarization) we used the classical method. If each atom in the solvent system is assumed to have an atomic polarizability, then the induced polarization is:

$$\mu_j = \alpha_j E_j \tag{9}$$

where the electric field E_j on atom j is given by:

$$E_j = \sum_k^{solute} \frac{q_k r_{jk}}{r_{jk}^3} + \sum_i^{solvent} \frac{q_i r_{ij}}{r_{ij}^3} + \sum_{i \neq j} - \nabla \frac{\mu_i r_{ij}}{r_{ij}^3} \tag{10}$$

Equation 10 is solved iteratively to give the induced polarization μ_j. Only the nonbonded interactions were evaluated for the solute-solvent and the solvent-solvent. The induction energy is then given by:

$$E_{induction} = -\frac{1}{2}\sum_{k}^{solute}\sum_{j}\frac{q_k \mu_j r_{kj}}{r_{kj}^3} - \frac{1}{2}\sum_{ij(i\neq j)}^{solvent}\frac{q_i \mu_j r_{ij}}{r_{ij}^3} \qquad (11)$$

The point charges for the solute were evaluated by our approach for fitting electrostatic potential points to a point charge model[29]. The exchange interaction energy, due to solvent-solute, was calculated empirically using a 6-12 pair potential:

$$E_{nonbonded} = \sum_{i<j}\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \qquad (12)$$

Thus, we can estimate the sum $E_{internal\ solute} + E_{solute-solvent}$ by the three terms $E_{internal\ solute}$(QM + Electrostatic) + $E_{induction}$ + $E_{nonbonded}$, even though $E_{induction}$ contains both solute-solvent and solvent-solvent polarization. However, it does not contain the term that is most difficult to determine, the differences in water-water interactions upon perturbation by the solute, $\Delta E_{solvent-solvent}$. These water-water interactions have been calculated with the molecular mechanical approach described above. This leads to an alternate formulation of the energy system E_T(QM):

$$E_T(QM) = E_{internal\ solute}(QM + Electrostatic) + E_{induction} + E_{nonbonded} + \Delta E_{solvent-solvent}(MM) \qquad (13)$$

All of the simulations were performed on the U.C.S.F. Structural Biology VAX-11/780 and the structures displayed on the U.C.S.F. Computer Graphics Lab Evans and Sutherland Picture System.

RESULTS

(A) Formation of the Tetrahedral Complex

Our first focus will be on the quantum mechanical and molecular mechanical results for the steps leading up to tetrahedral complex formation. We have modeled the four initial "snapshot" structures to be representative of reasonable steps along the pathway of OH^- attack. The starting distance of 6.0 Å between hydroxide ion and formamide was selected since it is long enough for the reactants to be considered as essentially separated species, yet small enough for each molecule to be completely solvated within a single cube of 216 Monte Carlo water molecules. The structure of the 1.48 Å complex was determined by complete relaxation of *all* parameters during the *ab initio* optimization, with the 2.08 and 3.08 Å structures being logically selected intermediates and optimized with respect to the fixed O2-C distance.

We optimized formamide and hydroxide at the 4-31G level and then carried out a single point *ab initio* calculation, using these internal geometries, for the reactants separated by 6.0 Å. After optimization of the 3.08 Å structure, the hydroxide ion was found to have migrated over towards the nitrogen end of formamide and to have abstracted one of the amide protons, forming $CHONH^- + H_2O$. We found this complex as the lowest energy structure on the gas phase potential surface; -29 kcal/mol relative to the tetrahedral complex. The structure corresponded to a hydrogen bond between the water and formamide anion. This is qualitatively consistent with the expectation[31] that $CHONH^- + H_2O$ is more stable than $H_2NCHO + OH^-$ in the gas phase. Hence, we did not completely gradient optimize the 3.08 Å structure (with only the C-O distance constraint) but stopped the optimization after ten

cycles, in order to assure that the OH^- was in an intermediate position for attack on the carbonyl carbon. The 2.08 and 1.48 Å geometries each took ≈ 35-40 cycles of optimization before the largest component of the gradient was less than 0.003 a.u.. To better assess the basis set dependence on the energetics of the four structures, we subsequently performed single point *ab initio* quantum mechanical calculations on the 4-31G optimized geometries using a 6-31G* basis set. Moller-Plesset perturbation theory at the second order level (MP2) was also used to estimate the correlation energy and the effect which this may have on stabilizing intermediates or transition states along the pathway. The gas phase quantum mechanical results for hydroxide attack appear in Table 1.

As can be seen from Table 1, both basis sets suggested that the energy of the system for OH^- attack on NH_2CHO monotonically decreased throughout. The ΔE for this half of the reaction was -38.9 kcal/mol (4-31G) and the 4-31G energies were similar to the 6-31G* over this part of the reaction. When MP2 correlation energy is included, the 1.48 and 2.08 structures are stabilized by ≈ 6.5 kcal/mol relative to the 3.08 and 6.00 geometries. Although in solution the tetrahedral complex might be expected to correspond to a high energy transition state, in the gas phase this structure is energetically much more stable than the reactants. Upon approach of the hydroxide ion, the C-N bond begins to lengthen from 1.34 to 1.47 Å and the hybridization of both the carbon and the nitrogen atoms changes from sp^2 to one having more sp^3 character (Figures 1-4). The amide protons can be seen to flip up, out of the plane; consistent with the tendency for the nitrogen lone pair to be antiperiplanar to the C-O bond[32].

The four optimized structures were placed in a water bath, restrained to their starting geometries and the water structure energy refined with 1000 energy evaluations of conjugate gradient minimization. After the molecular mechanics refinement, each of these systems was found to have a relatively strong hydrogen bonding network for the solvent structure and a well-defined first coordination shell (figures 1-4).

We wished to assess whether the solvation properties of the various structures made physical sense. We thus analyzed the number and geometry of the water molecules forming hydrogen bonds with the solute atoms (Table 2). Some points are worth noting. First, over the course of the first part of the reaction, the distance of the water hydrogens interacting with the nitrogen decreases as the hydroxide approaches. This is consistent with the fact that in the 6.0 Å structure, the nitrogen is an amide which is a rather poor hydrogen bond acceptor. Upon attack by OH^-, the nitrogen begins to take on more amine character and becomes a much better hydrogen bond acceptor. Second, in the 6.0 Å structure, there are six water molecules forming hydrogen bonds, with distances ranging from 1.62-1.72 Å, about the hydroxide (O2). As the reaction approaches the tetrahedral complex, with the subsequent transfer of charge from hydroxide to formamide, the total number of hydrogen bonds about the oxygen (O2) decreases to two, with only one of these a strong, near linear H-bond. Third, the number and (inferentially) strength of hydrogen bonding to the formamide oxygen increases during the course of the reaction, as O1 takes on more negative ion character. However, the hydroxide hydrogen has little tendency to hydrogen bond because it bears too little positive charge. Finally, the amide hydrogens both form reasonable H-bonds in the 6.00 structure but not for the tetrahedral complex, consistent

with amides being better hydrogen bond donors than amines.

Table 3 summarizes the values for the solvation energy calculated at the molecular mechanical level. It is clear that the solvent dramatically changes the energy profile of the first part of the reaction relative to the gas phase values. The solute-solvent energy is strongest for the separated reactants, much more so than either of the other three structures. Perturbations of the solvent in the first shell were also found to be greatest for the reactants, where the highly negative hydroxide ion strongly interacts with six water molecules. Here the $\Delta E_{solvent-solvent}$ is 71.7 kcal/mol relative to pure water. The other three structures have $\Delta E_{solvent-solvent}$ ranging from 40.5-49.6 kcal/mol. The net effect is a 42 kcal/mol of stabilization of separated reactants over the tetrahedral complex due to solvation.

We also used quantum mechanical methods to estimate $E_{internal\ solute}$ + $E_{solute-solvent}$ by carrying out *ab initio* calculations as described in METHODS. In Table 3 we compare this sum to the corresponding values from the combined quantum mechanical and molecular mechanical approach. As one can see, the two sets of energies give a very similar reaction profile.

(B) **Water Catalyzed Breakdown of the Tetrahedral Complex**

The second part of the reaction is water catalyzed breakdown of the tetrahedral complex to formate and ammonia. There are two ways that TC breakdown can occur; 1) direct donation of the internal proton (H4) to the nitrogen with concurrent C-N bond breakage, 2) water mediated proton donation to the nitrogen. Williams *et al*[11] have shown that, in the gas phase, a single ancillary water molecule lowered the barrier by 41 kcal/mol for the

hydrolysis of formaldehyde. For this reason, we expect that water mediated proton transfer will be more favorable than direct donation. Thus, we have added a single water molecule into the quantum mechanical tetrahedral complex and have optimized this structure at the 4-31G level followed by single point *ab initio* calculations at the 4-31+G, 6-31G* and 6-31G*/MP2 levels. This structure is shown in Figure 5 and is labeled 1.75 Å to correspond to the distance between the water hydrogen (H6) and nitrogen. The water proton was then forced to move to 1.23 Å and 1.15 Å from the nitrogen. At each point the *ab initio* energies were optimized using gradient methods.

Figures 5-7 illustrate the geometry of these gas phase gradient optimized structures. Figure 5 shows the water molecule forming a hydrogen bond, 1.75 Å, with the nitrogen of the tetrahedral complex. As this hydrogen is being donated to the recipient nitrogen, Figures 5-7 show a concurrent breaking of the C-N bond, with the distance going from 1.51 to 1.77 Å through these three steps. Interestingly, the hydroxyl hydrogen (H4) is found to remain bound to the oxygen throughout these three steps (0.99 Å when N---H = 1.15 Å). As the reaction proceeds, H4 was found to swing around and form an H-bond with the eventual recipient water oxygen, O3 (H-bond = 1.77 Å in Figure 7). We carried out further gradient optimization, and found this consistent with abstraction of H4 by the water *after* the C-N distance had reached 2.0 Å. By this point, the quantum mechanical energy was much lower than the 1.15 Å structure, so the transition state for this step seems to occur near the 1.23 Å structure. These results imply that, in the gas phase, tetrahedral complex breakdown is a stepwise process; beginning first with H_2O proton donation to the nitrogen and followed by proton transfer from the tetrahedral complex to the recipient water oxygen. Also in Figure 5, one

sees that the N lone pair has inverted from the structure of TC and are now no loinger antiperiplanar (app) to the C-O bond. This lone pair inversion prior to the H4 --> N transfer was also found in the calculations by Alagona *et al*[9]. They found that H4 transfer to N without water catalysis involved ≈ 35 kcal/mol barrier. Without N inversion, the barrier would certainly have been larger both in our calculations and those of Alagona *et al*[9].

The quantum mechanical energies are summarized in Table 4. Focusing only on the first three steps, the barrier for proton donation (over the distance 1.75 - 1.15 Å) is 14.2 kcal/mol (4-31G), 19.8 kcal/mol (6-31G*) and 12.2 kcal/mol (6-31G*/MP2).

The molecular mechanics hydrogen bonding geometries and refined energies are summarized in Tables 5 and 6, respectively. In the 1.75 Å structure the catalytic water oxygen (O3) is forming two hydrogen bonds (1.80 and 1.86 Å). As the reaction proceeds and the proton (H6) is donated to the nitrogen, this same oxygen forms three strong hydrogen bonds (1.61, 1.72 and 1.75 Å) with the solvent. It is not until the C-N bond is completely broken, and the products are released, that the nitrogen is "freed" for forming hydrogen bonds. This is illustrated in Table 5 where it is shown that no solvent molecules are closely associated with the nitrogen until ammonia is created. Both O1 and O2 are highly solvated over all of the modeled reaction steps, with the number and the quality of the hydrogen bonds increasing to a maximum when the anionic formate is formed. The two formate oxygens are solvated by six water molecules, all with distances less than 1.99 Å. Figure 8 illustrates the solvated products (NH_3 + $HCOO^-$).

In contrast to OH^- attack, the molecular mechanics solute-solvent energies (Table 6) for the second part of the hydrolysis are very similar for each of the three structures. This is consistent with the fact that one has a relatively large, diffuse anion throughout the reaction.

(C) **Entire Reaction Profile**

At this point we assess our methods for calculating the $E_{internal\ solute}$. By carrying out calculations on OH^-, H_2O, $HCOO^-$ and HCOOH with the 4-31G and 6-31G* basis sets, both with and without MP2 correlation correction, we find that all of these levels of theory overestimate the proton affinity of OH^- by ≈ 35 kcal/mol and that of formate by ≈ 15 kcal/mol. Such large differential errors are unsatisfactory, particularly for comparing the energy of small anions like OH^- and large diffuse ions like the tetrahedral intermediate.

Fortunately, Clark *et al*[24] have shown that *ab initio* calculations performed with an augmented 4-31G basis set (4-31+G) gave excellent agreement with experiment for the proton affinities of OH^- and $HCOO^-$; within 5 kcal/mol and 2 kcal/mol respectively. We have thus performed single point *ab initio* calculations on each of the eight gradient optimized structures employing this augmented 4-31+G basis set. Single diffuse *p* and *sp* functions were added to each heavy atom in line with the exponents reported by Clark *et al*[24].

We can also use the 4-31+G basis set to calculate a ΔE for reaction 1 in the gas phase and solution (Table 7). We assume that both the 4-31+G and 6-31G*/MP2 models are better than 4-31G and that each is "correcting" different defects in the more limited 4-31G model. Thus, our "best" quantum mechanical energies are the 4-31+G values plus the energy difference

between 4-31G and 6-31G*/MP2. To these quantum mechanical energies, we have added the solvation energies calculated as described above (using the first shell waters to estimate the change in solvent-solvent energy). As one can see, both the ΔE_T calculated for this reaction in the gas phase and in solution is in encouraging agreement with experiment and support the use of this approach for evaluating the energies along the reaction pathway.

Table 8 summarizes our "best" values for the various energies and energy components. Because the diffuse functions in the 4-31+G basis set give large counterpoise errors[33] for intermediate (partially bonded) structures, we used the 4-31+G energies only to estimate the energies for the reactants (6.0Å) and the TC. We confirmed that there was negligible counterpoise error in the 6.00 Å (4-31+G) calculations by comparing its total energy with the same reactants separated at infinity. The 2.08 and 3.08 Å energies were scaled between the 6.00Å and 1.48 Å (TC) energies in the same proportion as found at the 6-31G*/MP2 level. We then added the difference between 4-31G and 6-31G*/MP2 energies to these 4-31+G values to arrive at our "best" estimate of the quantum mechanical energies for these structures as we had done for the overall reaction ΔE (Table 7). In the second part of the reaction we used the 6-31G*/MP2 energies as our "best" estimate, since these structures are all similar, diffuse, anions and the proton affinity error will not be so large.

As a final step in the analysis, we wished to compare the energy of the TC (the final structure of the first part of the reaction profile) and that of 1.75 Å structure, which has been created by *ab initio* gradient optimization of 1.48 with an additional water hydrogen bonding to the nitrogen (the first struc-

ture of the second part of the reaction). We do this in the following way: to the molecular mechanics solute-solvent energy of 1.75 Å (-143.5 kcal/mol), we add the 6-31G*/MP2 *ab initio* calculated energy lowering due to the water-tetrahedral intermediate interaction (the energy difference between the 1.75 Å structure + an infinitely separated water vs. the 1.48 structure), which is 14.0 kcal/mol. This gives a net water-tetrahedral intermediate solute-solvent interaction energy (where the quantum mechanical water has been included with the classical waters) of -157.5 kcal/mol, essentially identical to the 1.48 solute-solvent energy of -158.0 kcal/mol. However, the five waters in the first coordination shell of the 1.48 structure have an $\Delta E_{solvent-solvent}$ of -80.5 kcal/mol, and the four classical waters and the one quantum mechanical water of the 1.75 Å have an $\Delta E_{solvent-solvent}$ = -89.7 kcal/mol, leading to a net stabilization of 1.75 Å relative to 1.48, of 8.7 kcal/mol. Thus in Figure 9, we use the values in Table 8 plus the estimated energy difference between 1.48 and 1.75 Å, to describe a complete reaction profile for gas phase and aqueous hydrolysis of formamide.

DISCUSSION

We first wish to assess the accuracy of the gas phase and solution phase reaction energies presented in figure 9. There are four experimental points which we can compare with our calculated values. First, the experimental ΔH for $H_2NCHO + OH^- \rightarrow NH_3 + HCOO^-$ in the gas phase can be compared with the results of our quantum mechanical calculations. The $\Delta H_{exptl.}$ = -46.3 kcal/mol (ref. 34), is in very good agreement with ΔE_{QM} = -48.9 kcal/mol.

Secondly, the experimental solution phase enthalpy ΔH = -5.2 kcal/mol (ref. 35) can be compared with the two endpoints in figure 9 and is found to be in

reasonable agreement with the calculated ΔH = -10.2 kcal/mol. Thirdly, kinetic isotope effect experiments on base catalyzed hydrolysis of amides[36] are consistent with the rate of hydrolysis (k_h) being nearly equal to the rate of O_{18} exchange (k_e). This result implies that the energy difference between the first barrier and TC is nearly equal to the energy difference between the second barrier and TC. This is consistent with our calculated energy differences of 13.0 kcal/mol and 13.3 kcal/mol respectively. Finally, it is encouraging that the calculated barrier reported here is quite consistent with the thermodynamic analysis of Guthrie[22] on base catalyzed hydrolysis of amides. Guthrie suggests an effective $\Delta G^{\neq} \approx 22$ kcal/mol for amide hydrolysis, consistent with our calculated $\Delta E^{\neq}$ of 22.0 kcal/mol; with the expectation that $\Delta S \approx 0$ *eu*. being discussed below.

The encouraging agreement of calculated and experimental energies supports our assumed mechanism for the reaction, although it does not prove it. In the (hypothetical) gas phase reaction, OH^- attack proceeds without a barrier but the second step, water catalyzed proton transfer, involves a barrier of 12 kcal/mol. In solution, on the other hand, there is a solvent induced barrier to OH^- attack, due to the more favorable solvation of OH^- than the more diffuse anions on the pathway to tetrahedral intermediate formation. The second step, breakdown of the tetrahedral intermediate, also involves a barrier and solvent H_2O mediated proton transfer from the carbonyl end of the molecule to the nitrogen end. We have simulated this step with a single water to concertedly abstract and transfer the proton, but it is equally likely that an OH^- could abstract the proton and a different H_2O could deliver it to the amine. Even though the anomeric effect is important in determining the conformation of the tetrahedral complex (N lone pair antiperiplanar to C-O2

bond), it does not appear to be important in causing a *net* lowering of the barrier in solution.

Let us now critically assess the features of this approach. The use of *ab initio* theory at the SCF/MP2 level offers a powerful approach to studying the intrinsic energies of chemical reactions, provided of course that an adequate basis set can be used for the problem at hand. In the calculations presented here, the use of diffuse basis functions was crucial in reasonably representing the relative energies of reactants, tetrahedral intermediate and products even in the gas phase. At the highest level of theory, the calculated gas phase ΔE for reaction 1 was -48.9 kcal/mol, in agreement with the experimental enthalpy of -46.3 kcal/mol for this reaction[35].

The second part of our approach involved a molecular mechanical calculation on the solvation of the various reactants, products and intermediate steps along the hydrolysis reaction. We have evaluated $E_{internal\ solute} + E_{solute-solvent}$ in two ways: the first involves adding quantum mechanical energies for $E_{internal\ solute}$ to the solute-solvent interaction energy calculated using molecular mechanical approaches. We have also evaluated both $E_{internal\ solute}$ and $E_{solute-solvent}$ by an alternate method which involves quantum mechanical calculations to include the electrostatic part of the solute-solvent interaction and classical calculations to determine solute-solvent van der Waals and polarization interactions. The fact that these two approaches gave rather similar results is encouraging and supports the use of simple molecular mechanical calculations to evaluate $E_{solute-solvent}$.

To evaluate $\Delta E_{solvent-solvent}$ upon introduction of the solute required a number of subjective decisions on how many waters to include in this calculation and

how to determine the solvent-solvent energies. Our decision to include only those waters closer than the first minimum in the radial distribution to evaluate $\Delta E_{solvent-solvent}$ was due to the difficulty of consistently implementing any other model. Such an approach underestimates the absolute magnitude of $\Delta E_{solvent-solvent}$. This underestimate of solvent-solvent energy changes leads to absolute $\Delta H_{solvation}$ of OH^- (table 7) too exothermic[37] by ≈ 34 kcal/mol. However, the approach seems to give reasonable relative values, such that the net calculated ΔE for reaction 1 in solution (-10.2 kcal/mol) is reasonably close to the experimental value (ΔH = -6 kcal/mol).

One of the reasons that these molecular mechanical (energy refinement) approaches to calculating solvation energies work at all is that the energies are dominated by very strong ionic interactions. Thus, the need for extensive averaging, inherent in much more time consuming Monte Carlo and molecular dynamics approaches, is not so great. It has been our experience from Monte Carlo simulations on dimethyphosphate[38] that $E_{solute-solvent}$ converges relatively rapidly with such an approach, but $\Delta E_{solvent-solvent}$ is much more time-consuming to accurately determine. To summarize, the weakest part of the approach presented here is the method for extracting $\Delta E_{solvent-solvent}$. By carrying out Monte Carlo or molecular dynamics calculations with periodic boundary conditions, one could avoid the problems of "edge effects" and a limited sampling of solvent-solvent energies. Thus, we plan to compare the results of our simpler model with the results of Monte Carlo or molecular dynamics solvation calculations on a number of the eight "snapshot" structures discussed above; the results of these more time consuming calculations will be reported in due course. We stress, however, that our simpler molecular mechanical approach is faster (6 VAX 11-780

hours/point versus ≈ 100 or more for Monte Carlo), gives qualitatively reasonable energies and physically reasonable solute-solvent hydrogen bonding patterns and is likely to be easier to extend to more complex systems than are full Monte Carlo simulations. Alternatively, one could carry out an approach whereby very short molecular dynamics runs are used to "heat up" the system, followed by molecular mechanics energy refinement. In this way, molecular dynamics simulations would help "pop" the system out of local energy minima. The computer time would be greater than straight molecular mechanics but have the advantage of sampling more of the conformational space.

Upon submission of the original version of this manuscript, we were stimulated to further assess this approach for simulating reaction pathways involving ions in solution by the recent paper of Chandrasekhar *et al*[39] who studied the S_N2 displacement reaction:

$$Cl^- + CH_3Cl \longrightarrow ClCH_3 + Cl^- \qquad (14)$$

using a combination of *ab initio* quantum mechanics and Monte Carlo umbrella sampling methods. They calculated a gas phase barrier of 2.4 kcal/mol, a solution phase $\Delta G^{\neq}$ = 26.3 kcal/mol and a solution $\Delta E^{\neq}$ = 28±8 kcal/mol, in good agreement with the experimental $\Delta G^{\neq}$ = 26.6 and $\Delta E^{\neq}$ = 23±3 kcal/mol (ref. 40). We used the geometry, Lennard Jones and electrostatic parameters from the Chandrasekhar *et al* study and our molecular mechanical solvation model on the reactants separated by 6.00 Å and on the pentagonal bypyramidal transition state. Table 9 summarizes the results of our calculations and the calculated $\Delta E^{\neq}$ = 23.4 kcal/mol is in fortuitously good agreement with that of the more accurate calculations and experiment.

Thus, the fact that our approach can calculate reasonable relative energies for two rather different reactions suggests that the approach may work on other ionic chemical reactions. Our approach is less rigorous and accurate than that of Chandrasekhar *et al*, but can be more easily applied to complex systems and involves 1-2 orders of magnitude less computer time.

Our calculations determined the energy (enthalpy) of the reaction, whereas the analysis by Guthrie[22] focuses on the free energy. We expect there to be two major contributions to the entropy differences along the reaction pathway. First, the loss of translational and rotational entropy in forming the tetrahedral intermediate would stabilize both reactants and products relative to all intermediate points. Using molecular mechanics techniques[41], we can estimate that the gas phase $T\Delta S$ at 298K for reactants --> tetrahedral complex is -9.0 kcal/mol. It is more difficult to quantify this $T\Delta S$ contribution in solution, but it will almost certainly be negative and somewhat smaller in magnitude than in the gas phase. The second major contribution comes into play only in the solution reaction and is the solvent electrostriction due to ionic effects. Smaller anions such as OH^- will reduce the entropy of the surrounding H_2O molecules more than large, diffuse anions such as the intermediate structures in the reaction. The change in entropy upon solvation of OH^- is more negative than that of I^- by 26 e.u. at 298K[31], corresponding to a $T\Delta S$ of 7.7 kcal/mol. We see that this effect is of opposite sign as the loss of translational and rotational entropy change and of the same order of magnitude. Thus, it is reasonable to suppose that the calculated energy profile (Table 9) may be a reasonable approximation to the free energy profile for formamide hydrolysis.

A nice feature of the approach presented here is that it is straightforward to extend to complex enzyme-substrate-water systems, such as catalysis by the serine proteases[42]. Along these lines, we are currently carrying out combined quantum/molecular mechanical calculations on peptide hydrolysis in the active site of trypsin[43]. One of the features of the serine proteases that has intrigued enzymologists for some time has been the unusual reactivity of the serine -OH in the enzyme, compared to an alcoholic OH, in hydrolyzing a peptide bond[44]. It is clear now that this reactivity is *not due* to the $-COO^-$ in the active site facilitating proton transfer from Ser 195 to Asp 102 through His 57[45]. We suggest that a substantial proportion of the cause of the unusual reactivity of this serine is that, once it has begun to deliver a proton to His 57 ($\Delta pK_a \approx 7$, $\Delta G \approx 10$ kcal/mol), the groups in the enzyme reduce the solvation of the incipient $R-O^-$ sufficiently to allow attack on the peptide bond without any further barrier (analogous to the gas phase first step presented here), and the barrier to proton transfer back to the substrate amine (formerly amide) NH_2 is also relatively facile (given the relative pK_a of these groups[46]). Thus, the role of desolvation by the enzyme of reactive or incipiently reactive groups should not be overlooked as a mechanism by which enzymes improve their catalytic efficiency over solution reactions. The importance of solvation in this regard has been stressed by Wolfenden[47] and the calculations reported here and related approaches applied to enzyme catalysis as well as biomimetic models will be able to assess the role of solvation/desolvation vs. propinquity in specific cases.

SUMMARY AND CONCLUSIONS

We have presented the development of an *ab initio* quantum mechanical plus molecular mechanical approach to simulating complex reactive processes in

the "gas phase" and in solution and have applied this approach to formamide hydrolysis by hydroxide ion. In the gas phase, the first step, OH^- attack, involves no barrier, and water catalyzed proton transfer to the amine (formerly amide) and accompanying peptide C-N bond cleavage has a ≈ 13 kcal/mol barrier. In solution, on the other hand, the first step involves a barrier of 22 kcal/mol while the barrier in the second step is little effected by solvation. The energetics of the reaction pathway calculated in solution and the gas phase and solution energies (enthalpies) for the overall reaction are quite consistent with available experimental data for these processes.

ACKNOWLEDGEMENTS

We gratefully acknowledge Caterina Ghio who wrote the first version of the program to incorporate a solute "water bath" as an option in AMBER. We thank W. Jorgensen for sending us the Monte Carlo snapshot of 216 TIPS4P H_2O molecules and for sending us reference 39a prior to publication. This study was supported by NIH (GM-29072 to PAK). The U.C.S.F. computer graphics lab Evans and Sutherland Picture System 2 (supported by RR-1081, R. Langridge director and T. Ferrin system manager) was very useful for displaying and analyzing all preliminary and refined structures. We thank J. Kirsch, W. Jencks and J. Caldwell for their helpful discussions.

Table 1
Energies for the Quantum Mechanical Model for Hydroxide Attack[a]

	GasPhase[b]			Solution[c]		
Structure[d]	4-31G	6-31G*	6-31G*//MP2	4-31G	6-31G*	6-31G*//MP2
Reactants	0.0[e]	0.0[f]	0.0[g]	0.0[h]	0.0[i]	0.0[j]
3.08	-20.4	-21.2	-22.2	22.6	22.7	21.7
2.08	-28.5	-27.3	-34.6	38.6	40.9	35.8
1.48(TC)	-38.9	-39.8	-46.6	38.4	40.0	34.5

a All energies are relative to the reactants.

b *Ab initio* quantum mechanical calculations ($E_{internal\ solute}$); kcal/mol.

c *Ab initio* quantum mechanical calculations incorporating the electrostatic, polarization and nonbonded energies of the solvent into the calculation ($E_{internal\ solute} + E_{solute-solvent}$); kcal/mol.

d Notation for the structures appears in fig. 1-4.

e Total quantum mechanical energy is -243.91061 a.u.

f Total quantum mechanical energy is -244.25502 a.u.

g Total quantum mechanical energy is -245.90514 a.u.

h Total quantum mechanical energy is -244.30997 a.u.

i Total quantum mechanical energy is -244.65346 a.u.

j Total quantum mechanical energy is -245.30045 a.u.

Table 2
Solute-Solvent Hydrogen Bonding Geometries
(Distance and Angle)[a]

Atom	Complexes 6.00(React.)	3.08	2.08	1.48(TC)
Formamide				
H1	2.36(136)	1.76(158)	2.29(164)	-
Formamide				
N	-	-	2.21(147)	2.18(143)
Formamide				
H2	1.92(174)	-	2.22(139)	-
Formamide				
O1		1.60(170)	1.61(176)	1.55(174)
-		-	1.72(166)	1.56(166)
-	-	-	2.22(129)	1.57(172)
Hydroxide				
O2	1.62(177)	1.51(179)	1.66(169)	1.77(166)
-	1.62(175)	1.58(175)	1.68(171)	2.17(127)
-	1.65(171)	1.59(171)	1.89(152)	-
-	1.66(179)	1.62(164)	2.17(151)	-
-	1.66(170)	-	-	-
-	1.72(171)	-	-	-
Hydroxide				
H4	2.26(106)	2.22(102)	2.38(105)	-
-	2.37(93)	2.31(95)	-	-

[a] The structures appear in figures 1-4. The hydrogen bond distance (in Å) is measured from water proton/oxygen to solute acceptor/donor atom. The angle is defined by Donor-Hydrogen---Acceptor (in °). All H-bond distances less than 2.4 Å are reported.

Table 3
Energies for the Molecular Mechanics Model for Hydroxide Attack

Struct.[a]	H_2O[b]	$E^{c}_{solute-solvent}$	$\Delta E^{d}_{solvent-solvent}$	$E^{e}_{solvation}$	$\Delta E^{f}_{solvation}$	$\Delta E(MM)^{g}$	$\Delta E(QM)^{h}$
React.	6	-231.5	71.7	-159.8	0.0	0.0	0.0
3.08	5	-185.9	49.6	-136.3	23.5	23.4	21.7
2.08	6	-157.2	40.6	-116.6	43.2	39.7	35.8
1.48(TC)	5	-158.0	40.5	-117.5	42.3	26.9	34.5

a The structures appear in fig. 1-4.

b The number of water molecules in the first shell. The first shell is defined as those waters found with oxygen distances less than the first minimum in the radial distribution function of solute-water (oxygen).

c Molecular mechanics nonbonded (van der Waals + electrostatic) energy due to solute-solvent interactions (kcal/mol).

d Energy difference between the water molecules in the first shell and an equivalent number of "ideal" water molecules. An ideal water here has an interaction energy of -24.2 kcal/mol.

e $E_{solute-solvent} + \Delta E_{solvent-solvent}$ (kcal/mol).

f Same as *e* but relative to the reactants (kcal/mol).

g Gas phase quantum mechanical energy of the optimized structures at the 6-31G*/MP2 level with molecular mechanics solute solvent energy; $E_{internal\ solute}$(QM) + $E_{solute-solvent}$ (MM); kcal/mol.

h 6-31G*/MP2 quantum mechanical calculations incorporating the electrostatic environment; $E_{internal\ solute}$(QM + Electrostatic) + $E_{induction}$ + $E_{nonbonded}$; kcal/mol.

Table 4
Energies for the Quantum Mechanical Model
for Water Catalyzed Hydrolysis[a]

	GasPhase[b]			Solution[c]		
Structure[d]	4-31G	6-31G*	6-31G*/MP2	4-31G	6-31G*	6-31G*/MP2
1.75	0.0[e]	0.0[f]	0.0[g]	0.0[h]	0.0[i]	0.0[j]
1.23	13.0	17.0	9.8	10.9	13.9	6.7
1.15	14.2	19.8	12.2	8.6	13.8	6.1

a All energies are relative to the reactants (kcal/mol).

b *Ab initio* quantum mechanical calculations ($E_{internal\ solute}$); kcal/mol.

c *Ab initio* quantum mechanical calculations incorporating the electrostatic, polarization and nonbonded energies of the solvent into the calculation ($E_{internal\ solute} + E_{solute-solvent}$); kcal/mol.

d Notation for the structures appears in fig. 5-7.

e Total quantum mechanical energy is -319.90559 a.u.

f Total quantum mechanical energy is -320.34904 a.u.

g Total quantum mechanical energy is -321.19680 a.u.

h Total quantum mechanical energy is -320.17997 a.u.

i Total quantum mechanical energy is -320.61432 a.u.

j Total quantum mechanical energy is -321.46208 a.u.

Table 5
Solute-Solvent Hydrogen Bonding Geometries
(Distance and Angle)[a]

Atom	Complexes 1.75	1.23	1.15	Products[b]
Ammonia				
H1	-	2.27(1.46)	2.29(135)	1.90(158)
	-	-	2.32(136)	-
Ammonia				
N	-	-	-	2.14(137)
-	-	-	-	2.15(136)
Ammonia				
H2	-	2.33(149)	-	2.22(134)
Water				
O3	1.80(156)	1.68(175)	1.61(175)	x
	1.86(154)	1.72(162)	1.72(151)	x
	-	-	1.75(166)	x
Water				
H5	-	-	-	x
Water				
H6	-	-	2.26(144)	x
Formate				
O1	1.62(167)	1.56(177)	1.61(167)	1.53(176)
	1.62(167)	1.59(178)	1.64(162)	1.62(178)
	-	1.69(164)	1.67(168)	
Formate				
O2	1.68(171)	1.87(149)	1.80(145)	1.60(172)
	-	-	-	1.76(172)
	-	-	-	1.80(165)
	-	-	-	1.99(159)

a The structures appear in figures 5-8. The hydrogen bond distance (in Å) is measured from water proton/oxygen to solute acceptor/donor atom. The angle (in °) is defined by Donor-Hydrogen---Acceptor (in °). All H-bond distances less than 2.4 Å are reported.

b Products are separated by 6.0Å. See fig. 8. (For only the product structure, the water atoms (O3, H5 and H6) are no longer treated in the quantum mechanical model and, hence, we report no hydrogen bond values for these atoms in the table).

Table 6
Energies for the Molecular Mechanics
Model for Tetrahedral Breakdown

Struct.[a]	H_2O[b]	$E^{c}_{solute-solvent}$	$\Delta E^{d}_{solvent-solvent}$	$E^{e}_{solvation}$	$\Delta E^{f}_{solvation}$	$\Delta E(MM)^{g}$	$\Delta E(QM)^{h}$
1.75	6	-143.5	41.8	-101.7	0.0	0.0	0.0
1.23	6	-156.7	58.5	-98.2	3.5	-3.4	6.7
1.15	6	-155.0	36.9	-118.1	-16.4	0.7	6.1

a Notation for the structures appears in fig. 5-7.

b The number of water molecules in the first shell. The first shell is defined as those waters found with oxygen distances less than the first minimum in the radial distribution function of solute-water (oxygen).

c Molecular mechanics nonbonded (van der Waals + electrostatic) energy due to solute-solvent interactions (kcal/mol).

d Energy difference between the water molecules in the first shell and an equivalent number of "ideal" water molecules. An ideal water here has an interaction energy of -24.2 kcal/mol.

e $E_{solute-solvent} + \Delta E_{solvent-solvent}$ (kcal/mol).

f Same as *e* but relative to the reactants (kcal/mol).

g Gas phase quantum mechanical energy of the optimized structures at the 6-31G*/MP2 level with molecular mechanics solute solvent energy; $E_{internal\ solute}$(QM) + $E_{solute-solvent}$ (MM); kcal/mol.

h 6-31G*/MP2 quantum mechanical calculations incorporating the electrostatic environment; $E_{internal\ solute}$(QM + Electrostatic) + $E_{induction}$ + $E_{nonbonded}$; kcal/mol.

Table 7
Energies for the Separated Reactants and Products

Struct.	6-31G*/MP2[α]	4-31+G[β]	4-31+G+CF[χ]	$E^{\delta}_{solu-solv}$	$E^{\varepsilon}_{solv-solv}$	ΔE_f^{ϕ}
OH^-	-75.51313	-75.28990	-	-173.7	34.0	-
$HCONH_2$	-169.39297	-168.69238	-	-20.3	8.5	-
NH_3	-56.34848	-56.11499	-	-15.2	8.6	-
$HCOO^-$	-188.66725	-187.93266	-	-146.0	39.8	-
React.	0.0	0.0	0.0	0.0	0.0	0.0
Prod.	-68.8	-41.0	-48.9(-46.3)	32.8	5.9	-10.2(-5.2)

α Absolute quantum mechanical energies with a 6-31G*/MP2 basis in a.u.

β Absolute quantum mechanical energies with 4-31+G basis in a.u.

χ Relative 4-31+G energies + correction factor (CF). CF is defined as the energy difference between reactants and products calculated with a 4-31G and 6-31G*/MP2 model. Experimental gas phase value appears in parentheses (ref. 34); kcal/mol.

δ Molecular mechanics $E_{solute-solvent}$; kcal/mol.

ε Molecular mechanics $\Delta E_{solvent-solvent}$; kcal/mol.

φ "Best" estimate of the energy of reaction in aqueous solution. 4-31+G + CF + $E_{solute-solvent}$ + $\Delta E_{solvent-solvent}$ Experimental aqueous phase value appears in parentheses (ref. 35); kcal/mol.

γ Relative energy of reactants and products in kcal/mol.

Table 8
Reaction Pathway Energies in the Gas Phase and in Solution

Struct.[a]	4-31+G[b]	CF[c]	ΔE(g.p.)[d]	ΔE(aq.)[e]
React.	0.0	0.0	0.0	0.0
3.08	-9.9	-1.8	-11.7	11.8
2.08	-15.4	-6.1	-21.5	21.7
1.48	-18.9	-7.4	-26.3	16.0
1.75	-	-	-40.3	8.7
1.23	-	-	-30.5	22.0
1.15	-	-	-28.1	4.4
Prod.	-	-	-48.9	-10.2

a Structures appear in fig. 1-7; product energies taken for the infinitely separated species and not the configuration shown in fig. 8.

b Quantum mechanical energies with 4-31+G basis; kcal/mol.

c Correction factor for correlation energy, taken as the energy difference between each structure calculated with a 4-31G and 6-31G*/MP2 model; kcal/mol.

d Our "best" estimate of the gas phase reaction. For part 1 and the products we used 4-31+G+CF. For the three steps in part 2 we used 6-31G*/MP2; kcal/mol.

e Our "best" estimate of the aqueous phase reaction. To the $E_{internal\ solute}$ ($\Delta E(g.p.)_T$) we add $E_{solute-solvent}$(MM) + $\Delta E_{solvent-solvent}$ (MM); kcal/mol.

Table 9
Solution Phase Energetics for the Reaction:
$Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$

Struct.[a]	ΔE_{QM}[b]	$\Delta E_{solute-solvent}$[c]	$\Delta\Delta E_{solvent-solvent}$[d]	$\Delta E_{Solvation}$[e]	$\Delta E(MM)$[f]
React.	0.0	0.0	0.0	0.0	0.0
BiPyr.	3.6	29.0	-9.2	19.8	23.4

a React. corresponds to the reactants separated at infinity. BiPyr. represents the bipyramidal transition state.

b The difference in quantum mechanical energies for the two structures was taken from ref. 39a.

c Difference in calculated molecular mechanical solute-solvent interaction energy between the reactants and transition state using the same approach as described for formamide/OH^-; kcal/mol.

d Difference between calculated $\Delta E_{solvent-solvent}$ between the reactants and transition state using the same approach as described for formamide/OH^-; kcal/mol.

e $\Delta E_{solute-solvent} + \Delta\Delta E_{solvent-solvent}$; kcal/mol.

f $\Delta E_{QM} + \Delta E_{solute-solvent} + \Delta\Delta E_{solvent-solvent}$; kcal/mol.

Figure 1: Stereo view of reactants at 6.00Å after 1000 energy evaluations of molecular mechanics refinement. All solute structures were optimized at the quantum mechanical level with a 4-31G basis set.

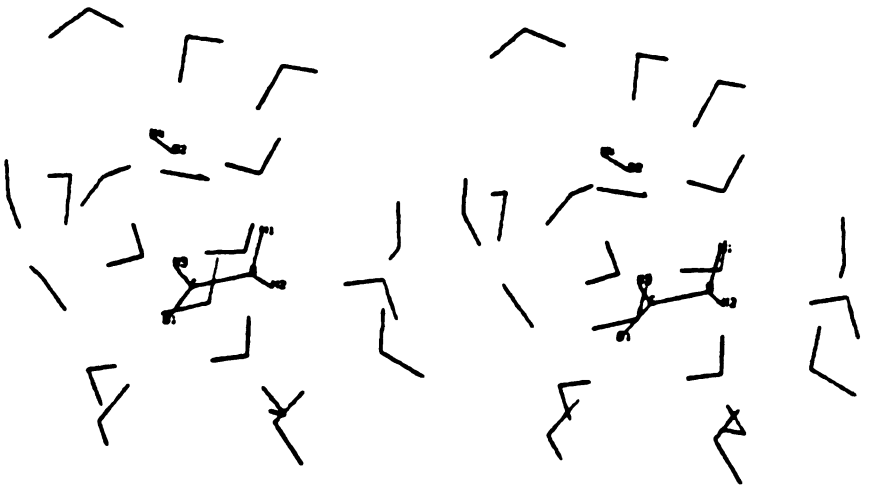


Figure 2: 3.08Å structure, same caption as figure 1.

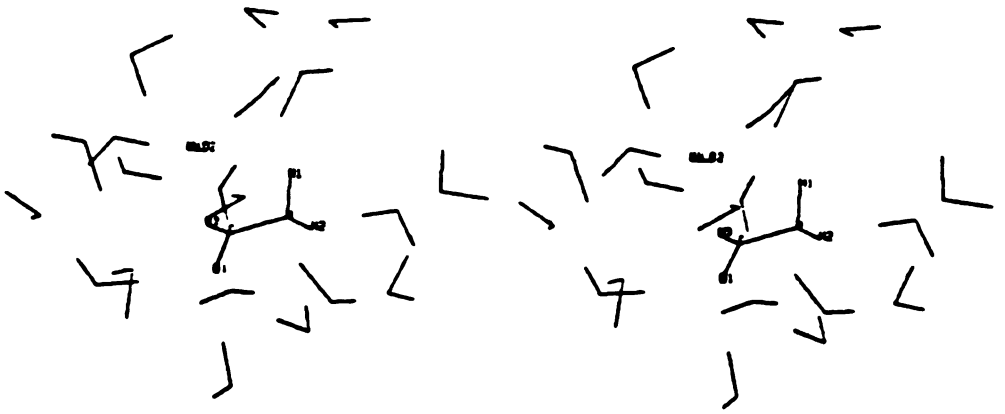


Figure 3: 2.08Å structure, same caption as figure 1.

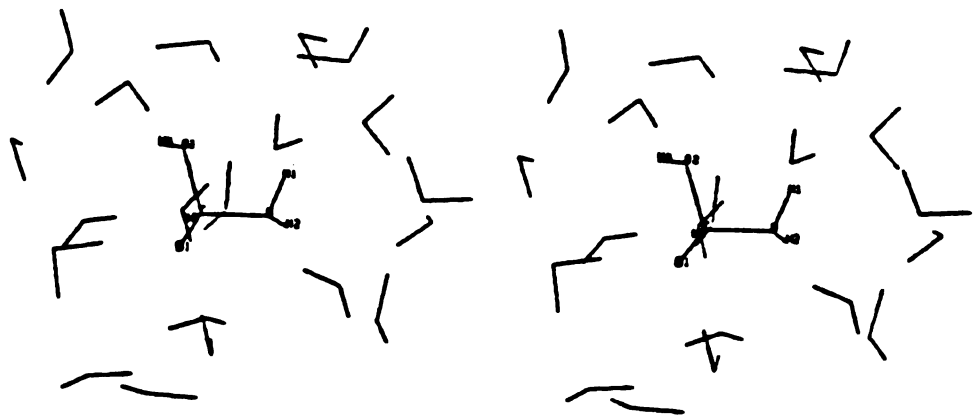


Figure 4: 1.48Å structure (TC), same caption as figure 1.

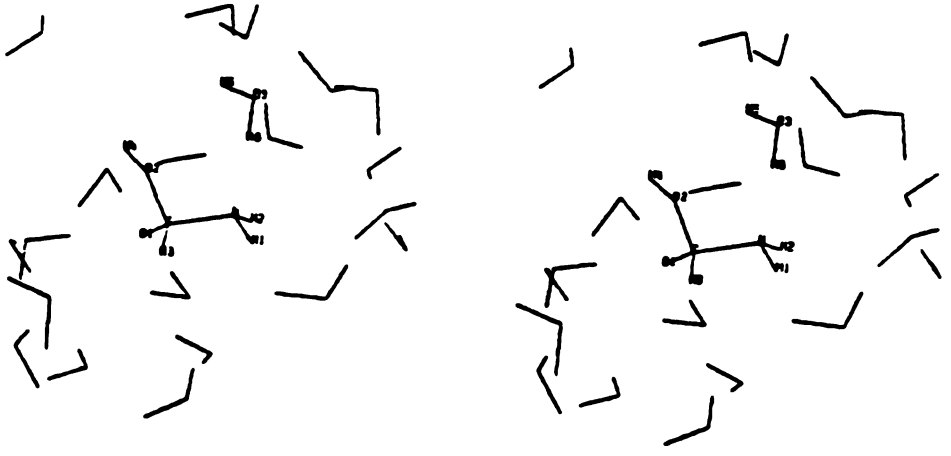


Figure 5: 1.75Å structure, same caption as figure 1.

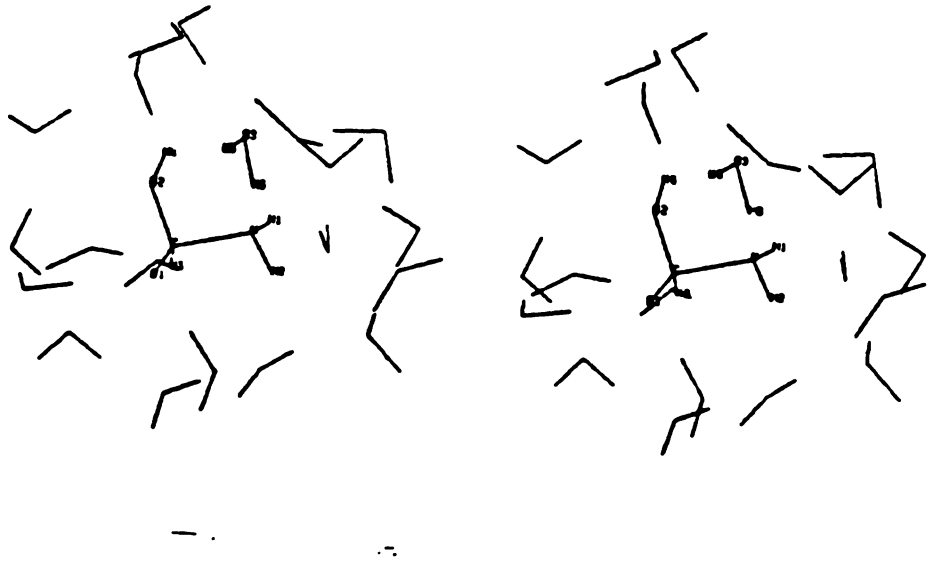


Figure 6: 1.23Å structure, same caption as figure 1.

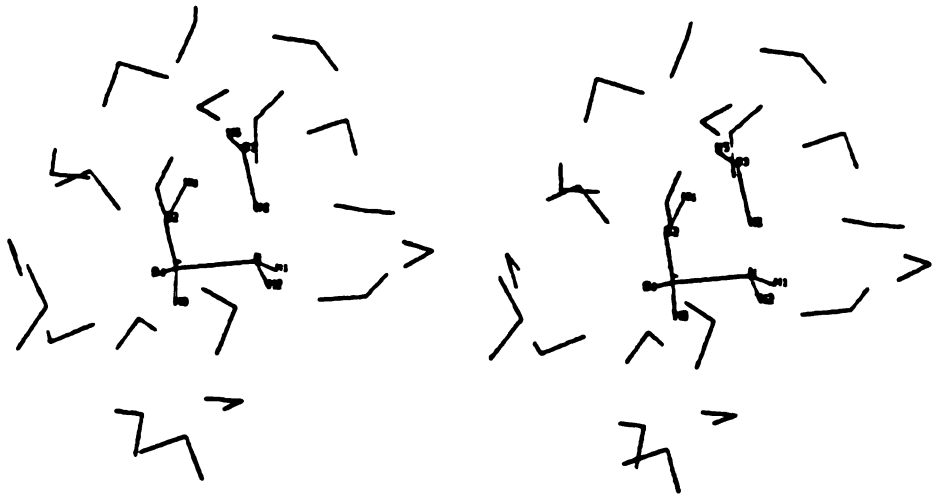


Figure 7: 1.15Å structure, same caption as figure 1.

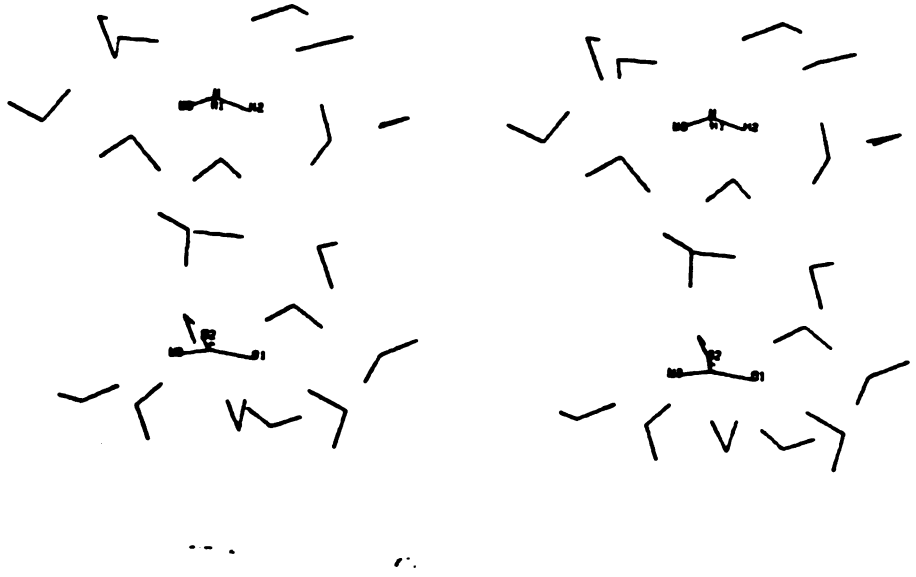


Figure 8: Products at 6.00Å, same caption as figure 1.

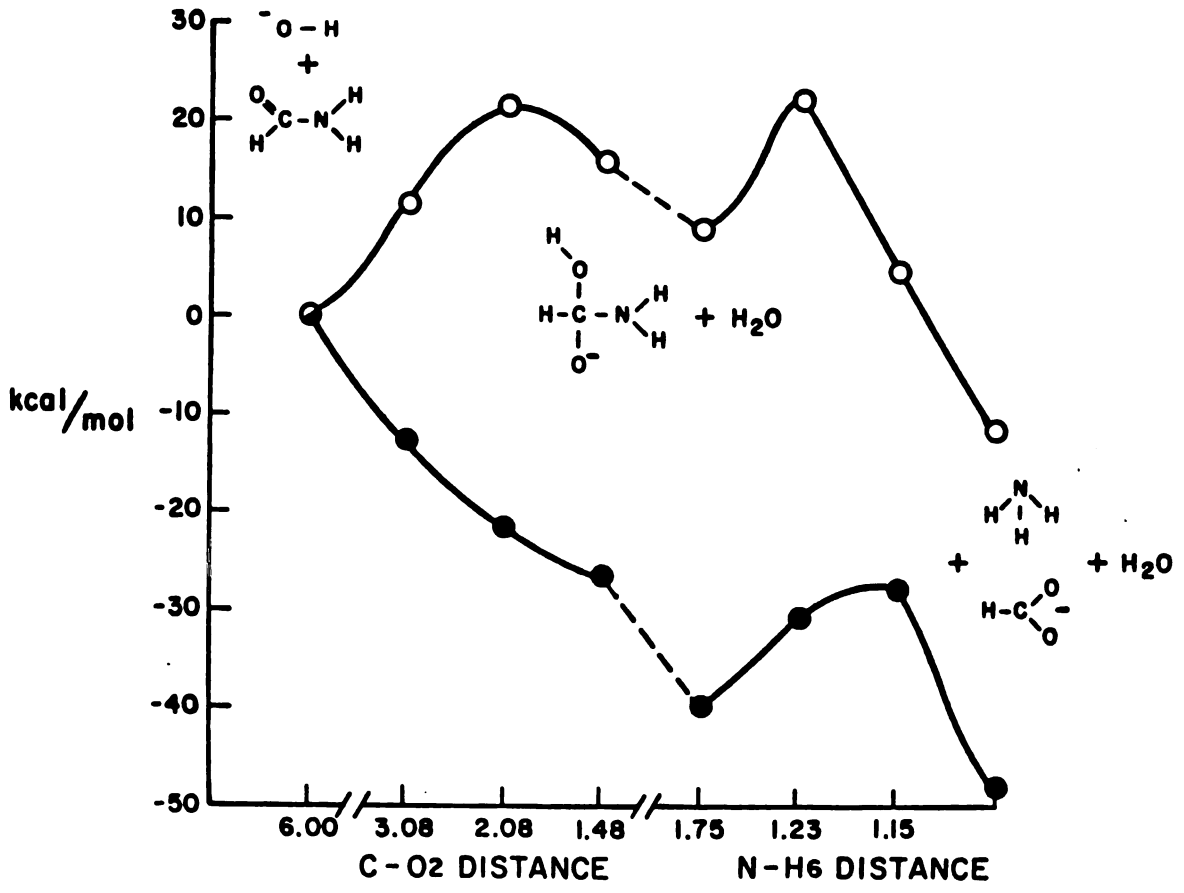


Figure 9: Representation of reaction coordinate (ordinate) for gas phase (solid circles) and aqueous phase (open circles) of the formamide hydrolysis reaction. The relative energy values are taken from table 8. The gas phase and aqueous phase energies are set equal to each other at 6.00 Å in order to better compare the energy profiles. The structures represented on the figure depict reactants, transition state and products respectively. The points on the abscissa represent the structures in figures 1-8. They are equally spaced only for representational purposes.

(1) General reviews: C. Walsh in "Enzymatic Reaction Mechanisms", W.H. Freeman and Co., San Francisco, 1979. A. Fersht in "Enzyme Structure and Mechanism", W.H. Freeman and Co., San Francisco, 1977.

(2) G. Wipff, A. Dearing, P. Weiner, J. Blaney and P. Kollman, J. Amer. Chem. Soc., 1983, **105**, 997.

(3) S. Scheiner and W. Lipscomb, Proc. Nat. Acd. Sci., 1976, **73**, 432.

(4) P. Van Duijnen, B. Thole and W. Hol, Biophy. Chem., 1979, **9**, 273.

(5) A. Warshel and M. Levitt, J. Mol. Bio., 1976, **103**, 227.

(6) J. Binkley, R. Whiteside, R. Krishnan, R. Seeger, D. Defrees, H. Schlegel, S. Topiol, L. Kahn and J. Pople, "Gaussian 80", Quantum Chemistry Program Exchange, (1980).

(7) W. Jorgensen, J. Chandrasekhar and J. Madura, J. Chem. Phys., 1983, **79**, 926.

(8) The Monte Carlo cube of 216 water molecules was kindly provided to us by W. Jorgensen.

(9) G. Alagona, E. Scrocco and J. Tomasi, J. Amer. Chem. Soc., 1975, **97**, 6876.

(10) T. Oie, G. Loew, S. Burt, S. Binkley and R. MacElroy, *ibid*, 1982, **104**, 6169.

(11) I. Williams, D. Spangler, D. Femec, G. Maggiora and R. Schowen, 1983, *ibid*, **105**, 31.

(12) B. Jonsson, G. Karlstrom, H. Wennerstrom, S. Forsen, B. Roos and J. Almlof, *ibid*, 1977, **99**, 4628.

(13) T. Oie, G. Loew, S. Burt, J. Binkley and R. MacElroy, Int. J. Quant. Chem.: Quant. Bio. Symp., 1982, **9**, 223.

(14) P. Claverie, B. Pullman and J. Caillet, J. Theor. Bio., 1966, **12**, 419.

(15) D. Beveridge, M. Kelley and R. Radna, J. Amer. Chem. Soc., 1974, **96**, 3769.

(16) M. Newton, J. Chem. Phys., 1973, **58**, 5833.

(17) R. Bonaccorsi, P. Palla and J. Tomasi, J. Amer. Chem. Soc., 1984, **106**, 1945.

(18) A. Pullman and B. Pullman, Q. Rev. Biophys., 1975, **7**, 505.

(19) A. Warshel, J. Phys. Chem., 1979, **83**, 1640.

(20) D. Beveridge, M. Mezei, P. Mehrotra, F. Marchese, G. Ravi-Shanker, T.

Vasu and S. Swaminathan in "Molecular-Based Study and Prediction of Fluid Properties", J. Haile and G. Mansoori eds., Advances in Chemistry Series, American Chemical Society, 1982.

(21) F. Stillinger and A. Rahman, J. Chem. Phys., 1974, **60**, 1545.

(22) J. Guthrie, J. Amer. Chem. Soc., 1974, **96**, 3608.

(23) W. Hehre, R. Stewart and J. Pople, J. Chem. Phys., 1969, **51**, 2657.

(24) T. Clark, J. Chandrasekhar, G. Spitznagel and P. Schleyer, J. Comp. Chem., 1983, **4**, 294.

(25) P. Hariharan and J. Pople, Theor. Chim. Acta., 1973, **28**, 213.

(26) J. Binkley and J. Pople, Int. J. Quant. Chem., 1975, **9**, 229.

(27) P. Weiner and P. Kollman, J. Comp. Chem., 1981, **2**, 287.

(28) T. Shimanouchi in "Tables of Molecular Vibrational Frequencies", parts 1-3, National Standard Reference Data Series - National Bureau of Standards, Washington, D.C., 1967.

(29) U.C. Singh and P. Kollman, J. Comp. Chem., 1984, **5**, 129.

(30) J. Chandrasekhar, D. Spellmeyer and W. Jorgensen, J. Amer. Chem. Soc., 1984, **106**, 903.

(31) The proton affinity of OH^- is 390 kcal/mol; that of $HCONH^-$ is likely to be similar to that of $HCOO^-$ (342 kcal/mol). This difference in anion stabilities will dominate any differences in bond energies and make $HCONH^- + H_2O$ much more stable than $HCONH_2 + OH^-$. P. Kebarle, Ann. Rev. Phys. Chem., 1977, **27**, 235.

(32) J. Lehn and G. Wipff, J. Amer. Chem. Soc., 1980, **102**, 1347.

(33) S. Boys and F. Bernardi, Mol. Phys., 1970, **19**, 558.

(34) Gas phase ΔH_f; $\Delta H_f(OH^-)$ = -33.7 kcal/mol, $\Delta H_f(HCONH_2)$ = -44.5 kcal/mol, $\Delta H_f(HCOO^-)$ = -114.5, $\Delta H_f(NH_3)$ = -11.0 kcal/mol. NH_3 and OH^- from "Selected Values of Thermodynamic Properties", D. Wagman, U.S. Govt. Printing Off., Washington D.C., 1968. $HCONH_2$ from A. Bander and H. Gundhard, Helv. Chim. Acta., 1958, **41**, 670. $HCOO^-$ from thermodynamic cycle using $\Delta H_f(HCOOH)$, $\Delta H_f(H)$, $\Delta H_f(H^+)$ and $\Delta H_{rxn}(HCOOH \rightarrow HCOO^- + H^+)$; R. Yamdagni and P. Kabarle, J. Amer. Chem. Soc., 1975, **95**, 4050 and Cox and Pilcher in "Thermodynamics of Organic and Organometallic Compounds", Academic Press, N.Y., 1970.

(35) Aqueous phase ΔH_f; $\Delta H_f(OH^-)$ = -55.0 kcal/mol, $\Delta H_f(HCONH_2)$ = -60.7 kcal/mol, $\Delta H_f(HCOO^-)$ = -101.7, $\Delta H_f(NH_3)$ = -19.2 kcal/mol. "Selected Values of Thermodynamic Properties", D. Wagman, U.S. Govt. Printing Off.,

Washington D.C., 1968.

(36) C. Bunton, B. Nayak and C. O'Connor, J. Org. Chem., 1968, **33**, 572.

(37) H.L. Friedman and C.V. Krishnan in "Water: A Comprehensive Treatise", v.2, Franks ed., Plenum Press, 1973.

(38) G. Alagona, C. Ghio and P. Kollman, "Monte Carlo Simulations of the Solvation of the Dimethylphosphate Anion", J. Amer. Chem. Soc., 1984, (submitted).

(39) a) J. Chandrasekhar, S. Smith and W. Jorgensen, "Theoretical Examination of the S_N2 Reaction Involving Chloride Ion and Methyl Chloride in the Gas Phase and Aqueous Solution", J. Amer. Chem. Soc., 1984, (Submitted)
b) J. Chandrasekhar, S. Smith and W. Jorgensen, J. Amer. Chem. Soc., 1984, **106**, 3049.

(40) R. Bathgate and E. Moelwyn-Hyghes, J. Amer. Chem. Soc., 1959, 2642.

(41) A. Hagler, P. Stern, R. Sharon, J. Becker and F. Naider, *ibid*, 1979, **101**, 6842.

(42) See J. Kraut, Ann. Rev. Biochem., 1977, **46**, 331.

(43) S. Weiner and P. Kollman, unpublished.

(44) D. Blow, Acct. Chem. Res., 1976, **9**, 145.

(45) A. Kossiakoff and S. Spencer, Nature, 1980, **288**, 414.

(46) M. Komiyama and M. Bender, Proc. Natl. Acad. Sci., 1979, **76**, 557.

(47) R. Wolfenden, Science, 1983, **222**, 1087.

CHAPTER 4

Why study enzymatic reactions at the theoretical level? First, a thorough theoretical simulation could help shed light on some of the more subtle structural effects occurring in and around the active site during enzymatic catalysis. Computer graphics enables one to easily and readily look at energy refined structures along a modeled reaction pathway and assess from a structural point of view what is occurring at the molecular level. Second, from an energetic standpoint, a molecular mechanics simulation can partition the energy terms so that one can calculate the magnitude of the contribution from various atoms such as solvent or specific peptide side chain residudes. With this in mind, we have simulated the hydrolysis of a model tripeptide by trypsin.

We selected trypsin for several reasons. First, its high degree of structural resolution[1] makes it one of the best protein structures available for studying with theoretical techniques. Second, there is a wealth of experimental data[2,3] available on the catalytic mechanism of this important enzyme. It should be noted that although a given mechanism can never be proved, it can be shown to be consistent with a wide body of experimental data. Specifically, trypsin has a well defined mechanism which is in good accord with many structural and kinetic studies. Thus, it is straightforward for us to model bond breaking/making reactions consistent with this mechanism. In a study on the hydrolysis of over 40 tripeptides by Pozsgay *et al*[3], they found that ACE-PHE-VAL-LYS-NME possessed a higher catalytic efficiency than nearly all of the other model tripeptides (k_{cat} / K_m = 4.5 x 10^4 /M sec). For

this reason, we selected this tripeptide (denoted SUB; where ACE is the acetyl group and NME is an N-methyl group) as our substrate of choice.

Part of the motivation for this study came from earlier work on α-chymotrypsin by Wipff *et al*[4]. Wipff *et al* focused on the increased catalytic efficiency of α-chymotrypsin for L type substrates vs. D. They focused on the Michaelis complex and on a model for the tetrahedral intermediate. The the molecular mechanics simulations were consistent with the much greater observed rate of hydrolysis for the L isomer because the L type substrate is preferentially stabilized (over the D) by specific interactions occurring in the tetrahedral intermediate. There is little energy difference between the L and D Michaelis complexes, also consistent with experimental K_m and K_D values. However, the absolute energy of the tetrahedral intermediate or transition state complex was found to be lower than that of the Michaelis complex, which was physically unrealistic. Thus, there were clearly important terms left out of this molecular mechanical model.

The two most important omissions from the α-chymotrypsin study were a means of evaluating the energetics of bond breaking/making and a framework for including the solvent into the molecular mechanics model. To properly evaluate the bond breaking/making energies, we have employed *ab initio* quantum mechanics. To handle the solvent, we have opted for a model which surrounds the solute (protein active site) in a "sea" of explicit water molecules taken from a Monte Carlo simulation of pure liquid water[5]. Both of these approaches were discussed extensively in Chapter 3.

Up to this point in the thesis, the united atom force field has been completely developed. Insulin and myoglobin were energy refined and the results

were encouraging along the lines of maintaining the important structural features of both proteins. Since the force field reasonably well maintains the structural properties (protein compaction and conservation of internal cavities in myoglobin), we can proceed with some confidence in exploring the functional aspects of enzymes. Our initial study involved a simulation, both in the gas phase and in aqueous solution, of a model system for the active site of a typical serine protease. We focused on the hydrolysis of formamide by hydroxide ion and found the results to be quite consistent with the available experimental data. With this in mind, we now possess the methods for simulating a chemical reaction in an enzyme and outline how we employed these methods in our study of trypsin catalysis.

METHODS

In its most general sense, the proteolysis of peptides by trypsin can be envisioned to occur in six steps: 1) Native free enzyme (NATIVE), 2) Substrate bound to trypsin or Michaelis complex (MICH), 3) Attack by nucleophilic reactive SER 195; forming a tetrahedral intermediate (TET1), 4) Breakdown of the tetrahedral complex by base assisted catalysis of HIS 57; yielding leaving group and acyl-enzyme complex (ACYL), 5) Water facilitated formation of a second tetrahedral complex (TET2) and 6) Final breakdown of TET2; releasing the remaining peptide strand and regenerating the native enzyme (PROD). We have modeled each of these six steps.

Analogous to our earlier *ab initio* quantum mechanical study on formamide hydrolysis by hydroxide ion, we have also added three intermediate points where bonds are being broken and formed. We selected the first two of these between MICH and TET1 and denoted them 3.08 and 2.08 to characterize the

distance between attacking SER O^- and recipient carbonyl carbon. The third structure corresponds to an intermediate point between TET1 and ACYL and is designated 1.75 to again represent the distance between SER OG and SUB C. The ACYL includes an explicit model built water molecule in a position ready for attack on the carbonyl carbon of the acyl enzyme. A complete listing of the initial model built structure and a brief summarized description of the complexes appears in table 1.

Each model built structure consists of the entire trypsin molecule; taken from the x-ray crystal structure by Chambers and Stroud[1] (resolution 1.5Å). Each complex also contains 200 explicit water molecules. Of these 200, 100 of the oxygen positions were taken from the work of Chambers and Stroud and the neutron diffraction study by Kossiakoff[6]. These water molecules were selected if the water oxygen was less than 3.4 Å from any two trypsin protein atoms. We selected this criterion since over 700 water oxygen positions have been located and we were interested only in those closest water molecules which interact with the protein. The hydrogen atomic positions were then placed manually, maximizing hydrogen bonding potential with both protein and neighboring water structure.

To model build SUB into the active site of trypsin we used structural information from the x-ray crystal structure of trypsin + BPTI. The key protein/ligand hydrogen bonds in trypsin + BPTI and trypsin + SUB are presented in table 2. All model building was carried out on an Evans and Sutherland PS II in the Computer Graphics Laboratory at U.C.S.F.

The next step was to take the trypsin + SUB + 100 "tightest" interacting water molecules and to submerse them into a "giant" cube of water. This

"giant" cube was generated by taking a single snapshot from a Monte Carlo simulation performed on a cube of 216 water molecules[5]. This cube was then translated in the ±x, ±y and ±z directions to form the "giant" cube (27 cubes of 216 water molecules). Once immersed, we used software within AMBER to select those water molecules within 15 Å of the active site and which were not within 2.35 Å of any Kossiakoff/Stroud (KOS/STROUD) trypsin or water atom. The effect of this is to obtain a continuous water structure about the active site. At this point, the model consisted of native trypsin + SUB + 100 KOS/STROUD H_2O + 100 AMBER H_2O.

At this point, there were several close water-protein and water-water interaction distances. To "clean-up" the structure we used the BELLY option in AMBER and allowed the entire water structure to relax about the enzyme for 100 cycles of steepest descent. In the BELLY option, part of the system is held fixed (the BELLY; trypsin in this case) and the rest is allowed to relax (200 water molecules). The full gradient and force evaluation is calculated for all water-water and water-protein interactions but **not** for protein-protein (in general BELLY-BELLY). This has the advantage of relieving close water contacts due to inaccuracies in positioning the KOS/STROUD water hydrogen atoms while not allowing any movement in the trypsin structure. This model of trypsin + SUB + 100 KOS/STROUD H_2O (relaxed) + 100 AMBER H_2O (relaxed) is denoted as MODEL1. It was used as the starting structure of NATIVE and taken as the basis for the other nine complexes.

As the reactive SER 195 O^- attacks the carbonyl carbon center, the peptide bond to be cleaved lengthens and the peptide linkage carbon and nitrogen atoms begin to take on more sp^3 character. To best incorporate this into the

molecular mechanics model, we have used geometries from snapshots taken from our earlier *ab initio* quantum mechanical calculations on formamide hydrolysis by hydroxide ion. The formamide molecule is analogous to the peptide linkage to be cleaved while the OH^- is representative of the attacking SER 195 O^-. To keep the 3.08, 2.08, TET1, 1.15 and TET2 intermediate geometries about those atoms undergoing rehybrdization and bond breaking/making, we have used large bond and angle force constants (K_r = 5000 kcal/Å^2 and K_θ = 1000 kcal/rad^2) and appropriate equilibrium bond and angle values about these groups in the molecular mechanics model. The effect of this is to allow the protein atoms and water molecules to relax about the restrained intermediate steps. Molecular mechanics is an energy minimization technique; as such it cannot refine around energies which are not at their minima. It is for this reason that one needs to play "tricks" with the force field in order to focus on the intermediate steps.

At this point a specific example will suffice to illustrate exactly how an intermediate step was model built. In our previous study on formamide hydrolysis by hydroxide ion, one of the intermediate structures corresponded to the quantum mechanically geometry optimized tetrahedral intermediate (FORM-TET). TET1 is the analogous trypsin complex. All of the internal geometries from FORM-TET were taken as starting r_0 and Θ_0 for this force field (e.g. $r_0(O1-C)_{FORM-TET} = r_0(OG-C)_{TET1}$ = 1.48Å, $\Theta_0(O1-C-O2)_{FORM-TET} = \Theta_0(OG-C-O)_{TET1}$ = 110°, etc.). Heavy force constants, as mentioned above, were used to conserve the internal geometries of the structure. The point charges for the atom were taken from the AMBER data base for standard amino acid residues and from a fit of the quantum mechanical electrostatic potential to a point charge model for the intermediate structures.

The next step was to allow each of our model built structures to energy refine. The most computationally intensive part of the molecular mechanics simulation involves evaluation of the nonbonded interactions. To reduce the time which is spent performing these calculations, we have employed a distance based and an explicit "residue by residue" based cutoff for selecting the nonbonded interaction pairs. Any residue or water molecule which had at least one atom within 8 Å of the reactive SER 195 was included in the ACTIVE SITE model of the protein, the rest of the atoms were treated as the BELLY (see above). In this fashion, all BELLY-BELLY nonbonded pair interactions need not be calculated (but the BELLY-ACTIVE SITE nonbonded atom pair interactions are evaluated). For the ACTIVE SITE model, we then used a residue based distance cutoff of 8 Å (e.g. any residue (or water molecule) which had at least one atom within 8 Å of **any** ACTIVE-SITE atom was taken as a nonbonded pair interaction). This corresponded to approximately 165,000 nonbonded pair interactions, which was a manageable number to obtain our goal of being able to perform the entire set of quantum mechanical and molecular mechanical simulations within 10 cpu hours on a CRAY-XMP. The molecular mechanics model built structures were then allowed to energy refine until the root mean squared (rms) gradient was less than 0.1 kcal/Å.

The next step was to evaluate the ***ab initio*** quantum mechanical energy of the important active site atoms. For the steps leading up to the acylenzyme, this model consisted of imidazole (or imidazolium), methanol (or methoxide ion) and formamide; with the key heavy atom positions coming from the energy refined coordinates of HIS 57, SER 195 and the SUB backbone. For the ACYL and TET2 geometries, our quantum mechanical model consisted of imidazole (or imidazolium), water (or attacking hydroxide) and methyl for-

mate. All hydrogen atoms not included in the molecular mechanics model were placed on with standard bond lengths and bond angles or taken from analogy with the formamide/OH^- study. Results are reported in table 3 for both SCF and MP2 energies.

Ideally, we would have liked to use a modified 4-31+G basis set (4-31G + diffuse *s* and *sp* functions), similar to the one used in our earlier study and taken from Clark *et al*[7]. However, the number of basis functions inherent to 4-31+G would have been too computationally intensive. The appeal of the 4-31+G basis set is in its ability to reproduce proton affinities of anions; which standard basis sets *without* diffuse functions are not able to do. For these reasons, we selected a modified Huzinaga[8] basis set (MINI2) and parameterized our own *s* and *sp* diffuse functions to best fit the proton affinities of hydroxide and methoxide ion (MINI+D). We put diffuse functions on the oxygen of methanol (methoxide) and on the carbon, nitrogen and oxygen of formamide models. We calculated a proton affinity of -393 kcal/mol and -403 kcal/mol for hydroxide and methoxide respectively, these numbers are found to be in reasonable agreement with the experimental values of 390 kcal/mol[9] and 384 kcal/mol[9].

We have evaluated the proton affinities of methoxide and imidazole to compare our MINI2+D with 4-31G basis sets. Using MINI2+D, we calculated 229 kcal/mol and 403 kcal/mol for imidazole and methoxide respectively; for 4-31G the results are 248 kcal/mol and 412 kcal/mol. Using 233 kcal/mol as the experimental value for the proton affinity of imidazole[10], the experimental energy for proton transfer (PT) at infinite separation is 384 - 233 = 151 kcal/mol. The calculated PT energy at infinite separation for MINI2+D and

4-31G are 174 kcal/mol and 164 kcal/mol respectively. This leads to a correction factors of 23 kcal/mol and 13 kcal/mol for MINI2+D and 4-31G respectively.

To test the basis set dependence of employing these correction factors we carried out quantum mechanical calculations on imidazole --- $HOCH_3$ and imidazolium --- CH_3O^- We took the optimized geometries from KM and KMION for the methanol/imidazole vs. methoxide/imidazolium interaction. The SCF ΔE between these structures was 74 kcal/mol (MINI+D) and 52 kcal/mol (4-31G). When we correct for the proton affinity errors, we calculated a MINI2+D(SCF) + CF = 51 kcal/mol vs. 4-31G(SCF) + CF = 39 kcal/mol; which are in better agreement with each other than the uncorrected values. It is sensible that the more flexible 4-31G basis set is able to more effectively stabilize the ion pair structure.

The molecular mechanics refinement and ***ab initio*** quantum mechanical calculations were carried out at Cray Research Inc. in Mendota Heights, Minnesota. All of the structures were model built and analyzed visually on the Evans and Sutherland PS2 in the Computer Graphics Lab at U.C.S.F.

RESULTS

The first focus will be on the results of the molecular mechancics energy refined structures. As stated above, each model was allowed to minimize until the root mean squared gradient was less than 0.1 kcal/mol. For the intermediate steps in which bonds were being broken/formed (e.g. 303 and 208), the systems generally refined to this rms within ≈ 5500 energy evaluations; for the more stable complexes (e.g. KM and KMION) only about 3500 molecular mechanics energy evaluations were needed. These structures

appear in figures 1-8. The discussion will concentrate on the unique structural and energetic effects occurring in the active site region as trypsin hydrolyzes its substrate.

The first important active site interaction occurs for the stabilization of HIS 57 by the ASP 102. In all eight of the structures, a good hydrogen bond is found between the donor HND of HIS 57 and one of the acceptor oxygen atoms of the ASP COO^-. It is interesting to note that the strongest hydrogen bond is formed for both of the tetrahedral complexes (TET1 and TET2); with a hydrogen bond distance of 1.68 Å and 1.64 Å respectively. This is physically reasonable since in both of the TET structures, the OG of SER 195 is bonded to the carbonyl carbon of the scissle peptide bond and this oxygen is found to be furthest away from the other HIS 57 proton (HNE). There is not much "pull" on HNE from OG and thus HIS 57 is allowed to maximize its hydrogen bonding potential with ASP 102.

Another important structural feature is the interaction of the carbonyl oxygen in the oxyanion hole. The oxyanion hole is defined by the two backbone hydrogens of GLY 193 and SER 195. Again, the overall "best" hydrogen bonding structure is found for both of the TET structures, with TET1 forming two hydrogen bonds of 1.74 Å and 1.76 Å with the backbone hydrogens. This aids us in understanding how the enzyme is able to stablilize intermediates or transition states along the reaction pathway; part of the energy lost in the bond breaking/making steps is regained by stabilization, via hydrogen bonds, of the intermediate complexes. Energetically it was found that the increased stabilization of the carbonyl group with the external environment, over the KM structure, was $\Delta E_{stabilization}$ = -10.8 kcal/mol for TET1.

The final important active site interaction involves the hydrogen bonding of the LYS side chain in a specificity binding pocket. Trypsin is a specific protease, hydrolyzing peptide bonds adjacent to a LYS or ARG residue. At the bottom of the pocket is the side chain of ASP 189 which interacts with the cationic ammonium and guanidinium chromophores of LYS and ARG respectively. Although this interaction is clipped away in figures 1-8, it is well maintained throughout the simulation.

The next focus is a discussion of the *ab initio* quantum mechanical results for each of the modeled structures. The full results are presented in table 3. As can be seen from table 3, all of the complexes in which the proton has been donated to the histidine are found much higher in energy than the neutral pair. It is interesting to note that the 175 structure is 22.1 kcal/mol higher in energy than KM at the SCF level but at the MP2 level it becomes 2.8 kcal/mol more stable.

In table 3 we present a correction factor (CF) based upon calibrating the calculated proton affinities of the monomeric units with experimental results. We first corrected for the difference in the calculated vs. experimental proton affinity of imidazole; 229 kcal/mol and 233 kcal/mol respectively. This stabilizes the KMION, 308, 208, ACYL and 175 realative to KM and TET2 relative to ACYL by the difference in proton affinities of 4 kcal/mol. In KMION the SER OG^- is a very localized anion. However, as it attacks the carbonyl carbon, the complex becomes much more diffuse. It is for this reason that we have selected methoxide and formate to represent localized and diffuse anions respectively. The proton affinity error for methoxide was calculated above to be -19 kcal/mol. We also calculated a proton affinity for formate of

336 kcal/mol, in good agreement with the experimental value of 339 kcal/mol[11]. We now have two endpoints to use as calibration points in going from KMION to TET1 and we have scaled the intermediate points linearly. The full results for the correction factor appear in table 3 and these values are added to the MP2 energies (ΔE(MP2) + CF) give our best estimate of the energetics associated with bond breaking/making.

To assess the overall stabilization effects of the protein and the solvent on the substrate, we have partitioned the molecular mechanics energy into several different groups. QM refers to those protein atoms which are included in the quantum mechanical simulation (e.g. CB and OG of SER 195) and SUB denotes the remaining substrate atoms. We have combined all of these atoms into one group and label this as QM/SUB. The first interaction involves QM/SUB interacting with all of the non quantum mechanical protein atoms found in the BELLY (QM/SUB + PROT). The second interaction involves QM/SUB interacting with every water molecule also found in the BELLY (QM/SUB + PROT). We have evaluated the internal strain energy of the substrate and denote this energy as SUB + SUB. The full results are presented in table 4.

The first point to make here is the dramatic protein stabilization of TET1 relative to KM; calculated to be 58 kcal/mol. The trend is also found to be consistent in going from 308 to 208 to TET1; the energies relative to KM decrease from -19 kcal/mol to -31 kcal/mol to -58 kcal/mol respectively. Similarly, TET2 is stabilized by trypsin by 15 kcal/mol relative to ACYL. However, the most encouraging result from this entire study is the effect which solvent has upon the intermediate structures along the reaction pathway. As

can be clearly seen from table 4, the water structure has a large stabilizing effect on all ion pair structures relative to the neutral pair interactions. KMION, 308, 208 and TET1 are all stabilized by between 30-40 kcal/mol relative to KM. Analogously, TET2 was calculated to be 24 kcal/mol lower in energy than ACYL. Finally, the internal SUB + SUB interaction energy was found to be relatively equal in energy for all structures. In other words, there does not appear to be much internal strain energy imparted upon the substrate as the reaction proceeds.

Finally, the entire reaction pathway is evaluated and presented in table 6. All of the important environmental and internal energies were included: protein-substrate, solvent-substrate and substrate-substrate molecular mechanics interaction energies and bond breaking/making quantum mechanical energies. First the reaction is found to go over a small barrier of 6 kcal/mol. In our previous study of formamide hydrolysis by hydroxide ion, the initial barrier of 22 kcal/mol was found to be solvent mediated and associated with the desolvation of the hydroxide anion. In the KMION structure (figure 2), there are two water molecules interacting with the OG of SER 195, while the remaining complexes have a single water hydrogen bonded to the attacking oxygen. It is interesting to speculate whether this first small barrier might also be solvent mediated. The rest of the steps leading up to TET1 formation and eventual breakage of the scissle peptide bond in 175 are found to be facile and proceed downhill. Although we have not included any contributions to the changing entropy of the system, it behooves us to make a first order approximation of this important term. The relative entropic contribution due to solvation effects will be negligible in this specific case. The important entropic contribution to the free energy is realized when a trans-

lational and rotational degree of freedom is "frozen" out upon formation of the tetrahedral complex. We have estimated this value (chapter 3, p.) for the gas phase reaction at 298K, $T\Delta S \approx 9.0$ kcal/mol. This would give an overall free energy of -5.1 kcal/mol for formation of TET1 relative to KM.

DISCUSSION AND CONCLUSION

We have presented an approach for simulating an enzymatic reaction; incorporating the important internal and environmental energies as best that we can at the time. However, this study is only a first order approach and it is important for us to outline its limitations and to discuss where improvement on this model is needed.

First, our quantum mechanical model is crude. As computer power continues to increase, we can think of expanding the active site model, of using better basis sets and of incorporating as much of the external protein and solvent environment into the quantum mechanics as is possible. Expanding on this last point, in our earlier study on formamide hydrolysis by hydroxide, we incorporated the point charges of the solvent into the one electron hamilton to give a more complete model at the quantum mechanical level. Similarly, this can be done with both protein and solvent in the trypsin study.

We have not had time to address in detail here the effect which long range electrostatic effects have upon the active site. We have used a dielectric constant $\varepsilon = 1$, rather than our standard $\varepsilon = R_{ij}$, since we have added 200 water molecules to the system. Our cutoff for evaluating nonbonded interactions was 8 Å and it is clear that with $\varepsilon = 1$, these effects are still contributing to the energy past this cutoff. The overall charge of trypsin is +7, so it is important to establish how these longer range electrostatics are affecting

the active site region.

Finally, we have only briefly discussed the contribution of entropy to the total energy profile but it is clear that this term is important and cannot be tossed aside. We can make approximations at this time since we are presenting only a first order approach, but it is clear that a firm grasp of the entropic effects is needed before we can accurately evaluate the reaction profile.

However, the important point to make here is to stress the dramatic effect which both the protein and the solvent have on the reaction profile. As we discussed above, we calculated stabilization of TET1 relative to KM of 58 kcal/mol by protein and 33 kcal/mol by the solvent. Clearly, it is necessary to include the solvent in any simulation in which intermediary structures are being compared along a reaction coordinate.

Table 1
Model Built Trypsin Structures

Structure	Description
NATIVE	Crystal structure of Chambers and Stroud.
SUB	Model tripeptide ACE-PHE-VAL-LYS-NME.
KM	Michaelis complex of NATIVE + SUB.
KMION	Same as KM but with proton transfered from SER 195 to HIS 57, forming this ion pair interaction.
308	Same as KMION but with the SER 195 OG^- restrained at a distance of 3.08 Å from the carbonyl carbon of the scissle peptide bond on SUB.
208	Same as 308 but with the OG-C distance restrained to 2.08 Å
TET1	Tetrahedral intermediate; OG-C distance restrained at 1.48 Å
175	Proton transfered to the leaving group methyamine with a N-C distance of 1.75 Å
ACYL	Acyl enzyme complex formed with remaining SUB.
TET2	Second tetrahedral intermediate on the way to deacylation; OG-C distance restrained at 1.48 Å
PRODUCTS	Cleaved SUB with a LYS-COOH end group.

Table 2
Hydrogen Bonding Considerations in
Model Building SUB into NATIVE

SUB	LYS NZ	LYS NZ	LYS O	LYS O	LYS HN	VAL O
NATIVE	ASP189 OD1	ASP189 OD2	GLY193 HN	SER195 HN	SER214 O	GLY216 HN
Crystal	3.66	3.37	1.67	1.88	2.96	1.92
Model	2.73	3.08	1.87	1.92	3.47	1.90

Table 3
Quantum Mechanical Model

Struct[a].	ΔE(SCF)[b]	ΔE(MP2)[c]	CF[d]	ΔE(MP2)+CF[e]
KM	0.0	0.0	0.0	0.0
KMION	72.9	62.9	-23.0	39.9
308	85.8	76.7	-16.0	60.7
208	85.8	71.1	-9.0	62.1
TET1	86.7	79.9	-1.0	78.9
175	22.1	-2.8	0.0	-2.8
ACYL	0.0	0.0	0.0	0.0
TET2	53.6	54.6	-1.0	53.6

a For description of the structures see table 1.

b SCF energy with MINI2+D basis set in kcal/mol. Absolute energy of KM is -505.6571 au.

c MP2 energy with MINI2+D basis set in kcal/mol. Absolute energy of KM is -506.2489 au.

d CF is the quantum mechanical correction factor, as defined in the text, to adjust the calculated proton affinities of methoxide, formate and imidazole to the experimental values; in kcal/mol.

e Total relative energy for MP2 + CF in kcal/mol.

f SCF energy for ACYL is -525.3965 au and MP2 energy is -525.9976 au.

Table 4
Molecular Mechanical Model

Struct[a].	QM/SUB + PROT[b]	QM/SUB + WATER[c]	SUB + SUB[d]	ΔE^{e}_{tot}
KM	0	0	0	0
KMION	7	-43	3	-33
308	-19	-41	-2	-62
208	-31	-33	-2	-66
TET1	-58	-33	-2	-93
175	-10	-9	6	-13
ACYL	0	0	0	0
TET2	-15	-24	1	-38

a See table 1 for description of the structures.

b QM is defined as those protein atoms which are included in the quantum mechanical model. SUB is defined in table 1. The energy reported is the molecular mechanics interaction energy (electorstatic + van der Waals + hydrogen bond) of the QM and SUB atoms with all of the protein atoms defined in the BELLY (see text). Energy is in kcal/mol relative to the KM structure.

c Interaction energy of QM and SUB with all of the water molecules defined in the BELLY. Energy is in kcal/mol relative to the KM structure.

d The intrinsic internal interaction energy of SUB with itself. Energy is in kcal/mol relative to the KM structure.

e $\Delta E_{tot} = \Delta E_{QM/SUB + PROT} + \Delta E_{QM/SUB + WATER} + \Delta E_{SUB + SUB}$. Energy is in kcal/mol relative to the KM structure.

Table 5
Complete Reaction Profile

Struct[a].	ΔE^{b}_{MM}	ΔE^{c}_{QM}	ΔE^{d}_{tot}
KM	0	0	0
KMION	-33	39.9	6.1
308	-62	60.7	-1.3
208	-66	62.1	-3.9
TET1	-93	78.9	-14.1
175	-13	-2.8	-15.8
ACYL	0	0	0
TET2	-38	53.6	19.6

a See table 1 for a description of the structures.

b Molecular mechanical ΔE_{tot} taken from table 3.

c Quantum mechanical ΔE_{tot} taken from table 4.

d $\Delta E_{tot} = \Delta E_{MM} + \Delta E_{QM}$

Figure 1: Active site energy refined model for KM. The important active site residues are labeled.

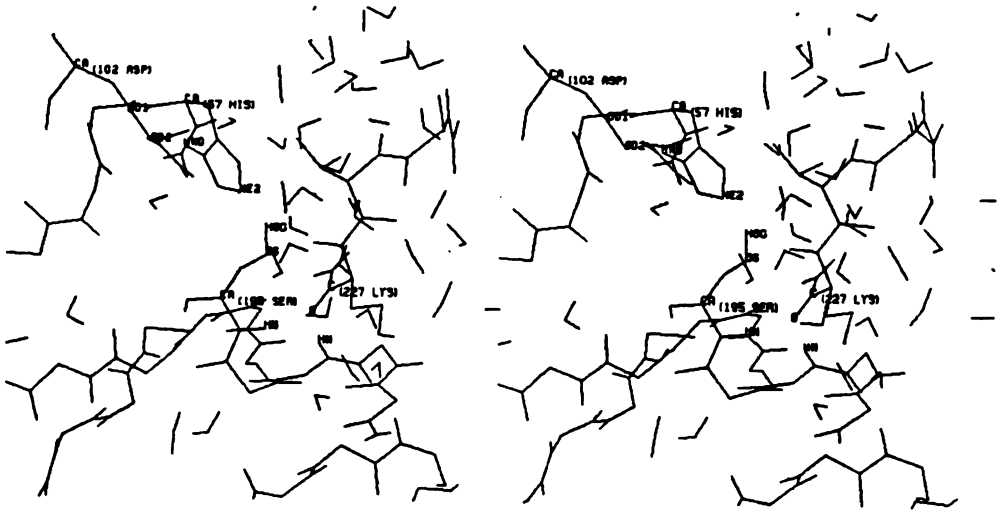


Figure 2: Active site energy refined model for KMION. The important active site residues are labeled.

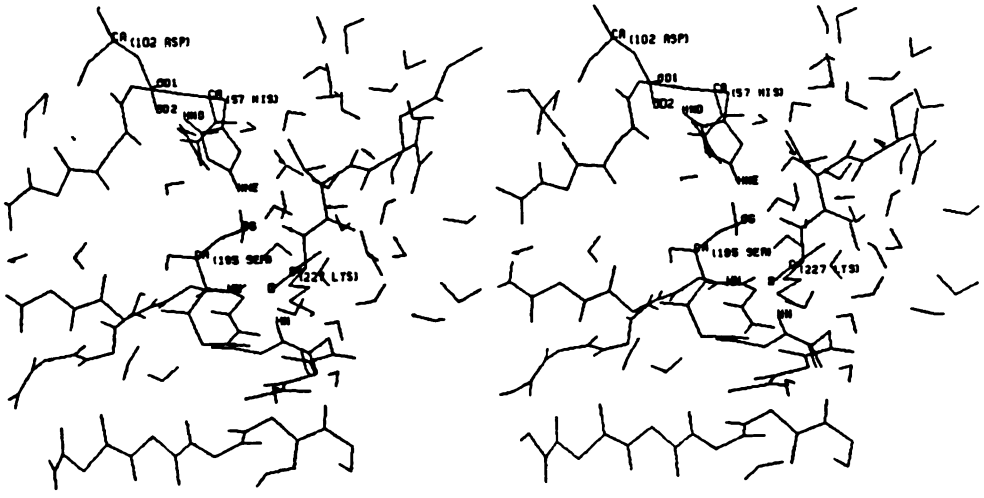


Figure 3: Active site energy refined model for 308. The important active site residues are labeled.

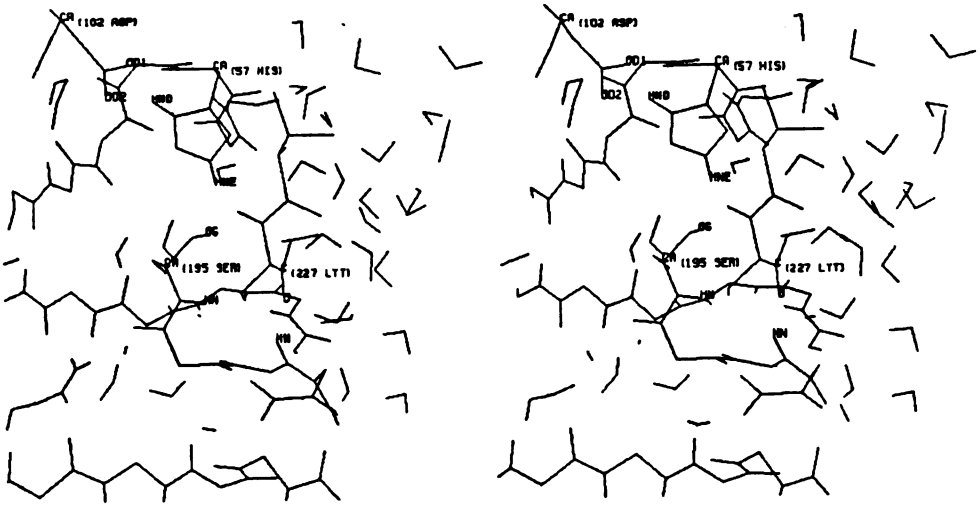


Figure 4: Active site energy refined model for 208. The important active site residues are labeled.

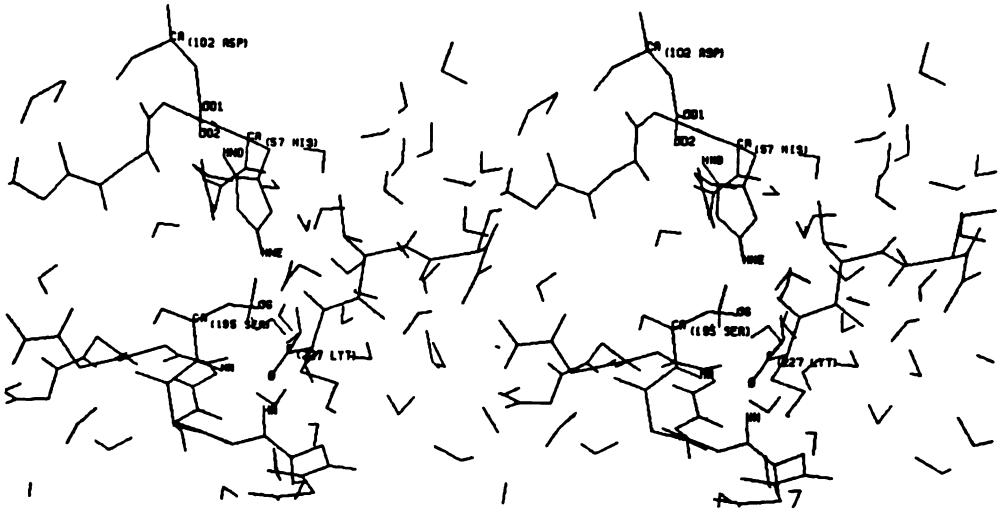


Figure 5: Active site energy refined model for TET1. The important active site residues are labeled.

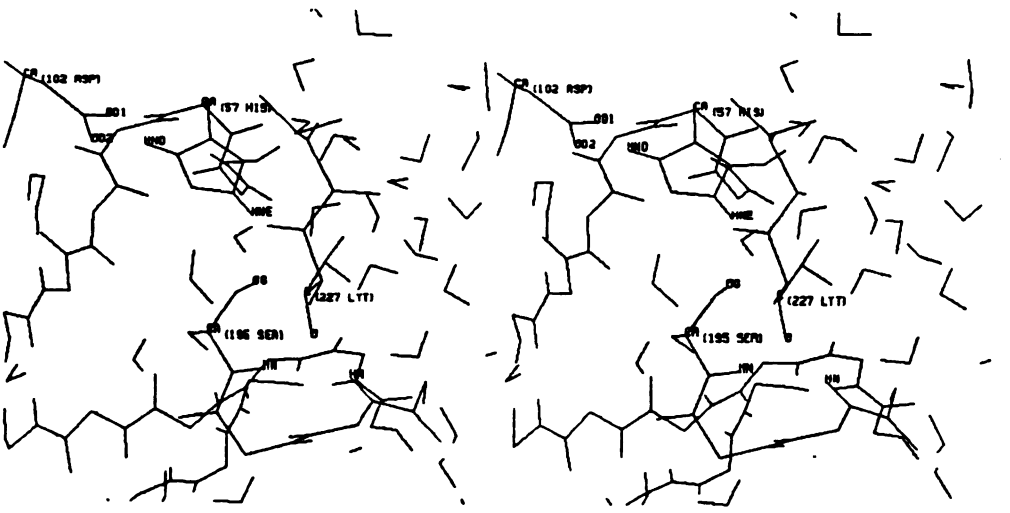


Figure 6: Active site energy refined model for 175. The important active site residues are labeled.

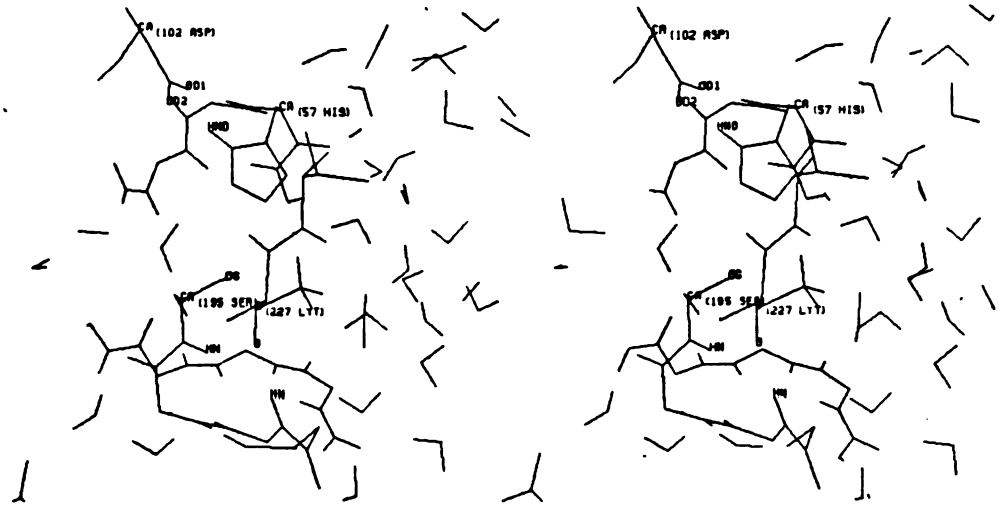


Figure 7: Active site energy refined model for ACYL. The important active site residues are labeled.

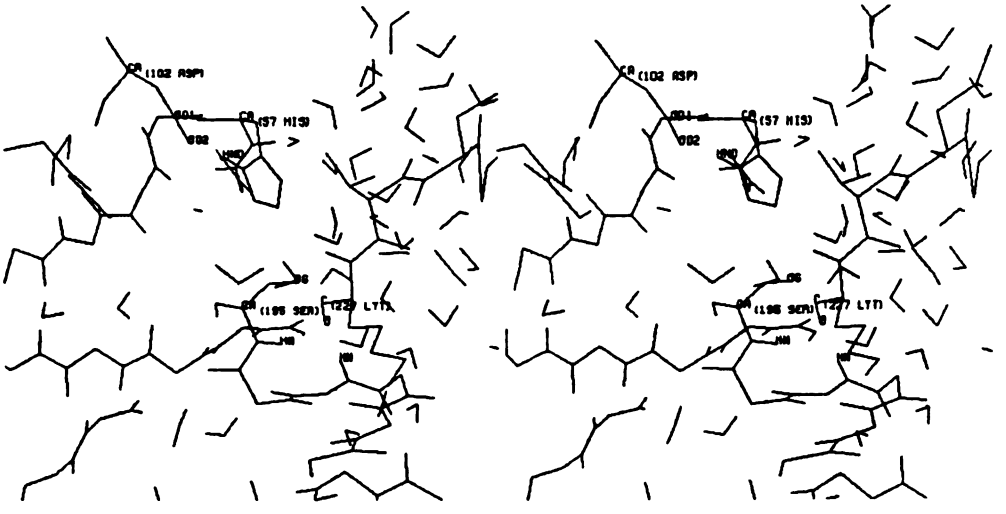
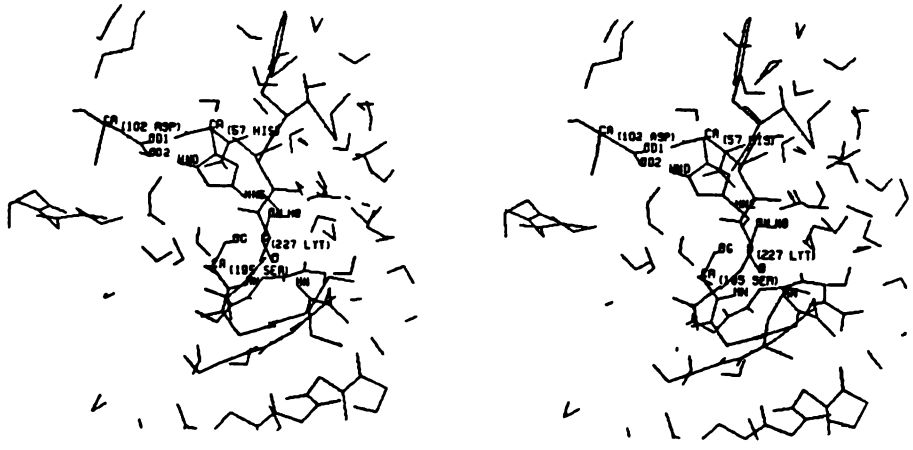


Figure 8: Active site energy refined model for TET2. The important active site residues are labeled.



(1) J. Chambers and R. Stroud, Acta Cryst., Sect. B, 1977, **33**, 1824.

(2) A. Fersht in "Enzyme Structure and Mechanism", W.H. Freeman and Co., San Francisco, 1977.

(3) M. Pozsgay, G. Szabo, S. Bajusz, R. Simonsson, R. Gaspar and P. Elodi, Eur. J. Biochem., 1981, **115**, 497.

(4) G. Wipff, A. Dearing, P. Weiner, J. Blaney and P. Kollman, J. Amer. Chem. Soc., 1983, **105**, 997.

(5) Provided to us by W. Jorgensen.

(6) Personal communication with A. Kosiakoff.

(7) T. Clark, J. Chandrasekhar, G. Sptiznagel and P. Schleyer, J. Comp. Chem., 1983, **4**, 294.

(8) H. Tatewaki and S. Huzinaga, J. Comp. Chem., 1980, **1**, 205.

(9) S. Ikuta, J. Comp. Chem., 1984, **5**, 374.

(10) Private communication from R. McIver (U.C. Irvine) that the proton affinity of histidine is 28.5 kcal/mol higher than NH_3. We use of value of 205 kcal/mol for HN_3 and assume that the proton affinity of histidine is the same as imidazole.

(11) R. Yamdagni and P. Kebarle, J. Amer. Chem. Soc., 1973, **95**, 4050.

CONCLUSION

This thesis represents the culmination of my graduate study in theoretical biophysical chemistry. In a sense, I liken the nature of my research to the building of a house. The force field is analogous to the carpentry tools and materials, while the methodology developed for simulating chemical reactions (in the gas phase, in aqueous solution and in an enzymatic environment) can be likened to the housing foundation. Just as a house with a faulty foundation will surely crumble; an enzymatic simulation carried out with a non-representative methodology will surely lead to spurious results. Similarly, inaccurate or nontransferable parameters will more likely than not lead to nonrealistic conformations or erratic energies upon energy refinement of biological macromolecules. Along these lines, the discussion in this section will be limited to a more qualitative overview of the thesis.

Force fields are empirical in nature and are derived from experimental and, in some cases, *ab initio* quantum mechanical data. Ideally, one would want to parameterize his force field completely from the system which is desired (e.g. an entire protein). Unfortunately, there is just not enough data to accomplish this. Thus, we are left to select small model systems which we feel best represent the repeating structural units of nucleic acids and proteins. For example, for the repeating nature of the polypeptide backbone, we selected N-methylacetamide (NMA). NMA contains the important amide linkage (H-N-C-O) as well as two methyl groups. Formamide also contains the amide linkage but the two methyl groups are replaced by hydrogens, making it not representative enough. A larger peptide would perhaps be a better

choice, however the experimental data becomes more limited as peptides get bigger. So, the model systems upon which the force field will eventually be parameterized invariably become compromises.

Once the model molecule is chosen, it is important to find the most accurate structural and energetic information which is available. In the case of the peptide backbone, we opted for structural data averaged from the crystal structures of many proteins rather than solely from NMA. In this manner we obtained data from the whole, so it was not necessary to take it from the parts. However, we did use infrared calculated vibrational frequencies from NMA to "fine tune" our bond stretching and angle bending force constants about the peptide backbone. The salient point here is that it is incumbent upon the researcher to clearly state exactly where the parameter set came from and what experimental data was used in the process. Then, when the force field is applied, one can have confidence in its transferability and be aware of its limitations.

Just as the force field represents the materials needed to build the house, the empirical potential energy function, $E(rv)$, is analogous to the machinery or tools needed for construction. First, I should point out that there is no unique or absolute $E(rv)$, they are afterall *empirical* in nature. Second, one should always be aware of the limitations inherent within the potential energy function. For example, for the bond stretching term, AMBER uses a simple harmonic $K_r(r - r_{eq})^2$ term. Although a Morse type potential is physically more realistic, the shape near the bottom of the harmonic potential well nicely mimics that of the of the Morse potential. However, as bonds are being broken/formed or a given bond is stretched far beyond its equilibrium

value, one needs to realize that a simple harmonic function is no longer appropriate. In the case of simple polypeptides or nucleic acids, our first order potential function appears to perform well from both a structural and energetic standpoint.

As a final word on parameter derivation, force fields are constantly evolving species and need to be continually updated and revised. As more accurate structural information on small peptides and proteins becomes available through neutron diffraction or x-ray crystallographic studies or new experimental data becomes available on the relative energies of small dipeptide conformations, for example, then it is important to adjust the corresponding parameters accordingly. More important is the ability to be able to derive additional parameters for molecules which are not standard amino nor nucleic acids (e.g. a modified residue in a protein or an inhibitor for an enzyme). This should be relatively straight forward as chapter 1 outlines in detail the steps to be taken from initial charge generation through a scaling algorithm for bond and torsion force constants. In this manner the force field can be modified or appended, with relative ease, to suit the goals of the research group.

Now that a reasonable force field has been developed, our focus shifts to the foundation or methodology for simulating chemical reactions. The first question to be addressed concerns what I would call mechanical methodologies. By this I am refering to those variables, within the molecular mechanics machinery, which can be controlled and varied at the start of a simulation (e.g. a nonbonded cutoff distance, dielectric constant, scale factor for 1-4 nonbonded interactions, root mean squared gradient convergence criterion,

etc.). For example, the most computationally intensive part of the algorithm takes place in the part of the code which evaluates the nonbonded pair interactions. To reduce the total number of these paired interactions ($\approx N(N+1)/2$; where N is the total number of atoms in the system), AMBER has a nonbonded pair cutoff. For a given atom i, its nonbonded interaction with atom j will only be calculated if the distance between i and j is less than the stated cutoff. Surely evaluating all nonbonded pair interactions is not necessary as the interactions are relatively short ranged and far interactions will contribute almost zero to both the gradient and energy terms. However, what is an appropriate nonbonded cutoff to use? For proteins, we have found that at least a 9AG atom based cutoff is needed to assure smooth convergence during the energy refinement and to enable the system to achieve low rms gradients (less than 0.1 kcal/AG). A 12AG or even 15AG cutoff is even better but for an average sized protein (trypsin), this corresponds to far too many nonbonded interactions to make the simulation feasible on a VAX 11/780 type computer. Invariably a compromise is made between an ideal value and one which enables the system to become computationally manageable. The salient point here being that one needs to be aware of these user controlled variables and, more importantly, needs to perform controls to assure that the system is being treated in an accurate manner.

The next methodology is associated with the system itself. The ultimate goal of this thesis was to develop a framework for simulating enzymatic reactions. Prior to developing this framework, one needs to take a step back and ask, "What are the important environmental factors and have I included them in the molecular mechanics model?" The first factor which this thesis dealt with was to include the energetics associated with bond breaking/making. At

the present time, the best theoretical tool for evaluating bond breaking/making energies is *ab initio* quantum mechanics. The methodology itself is quite simple to carry out and a simple example taken from our formamide hydrolysis by hydroxide ion study will show this. Consistent with a mode of attack whereby the hydroxide oxygen attacks the carbonyl carbon of formamide, one need only pick intermediate points along the reaction pathway (the pathway in this case being defined by the distance between the O and C), fix that distance and let the system undergo quantum mechanical z-matrix gradient optimization of the remaining degrees of freedom. Intermediate "snapshot" geometries are realized and the corresponding energies can be compared with the unrestrained reactants to arrive at a gas phase energy profile. These geometries can then be placed in the molecular mechanical model and the protein and/or water structure allowed to refine about the solute. Nevertheless, we have shown that it is necessary to include this energy associated with bond breaking/making in the total energy in order to realize an accurate energy profile.

It should also be pointed out that there are limitations in the quantum mechanical model and these need to be dealt with accordingly. Most importantly, there is a very definite basis set dependence associated with each set of calculations. Our trypsin pathway involves transfer of protons back and forth between protein, substrate and water. Because of this, we have focused on the proton affinities of the separate monomeric units in the active site model: immidazole (immidazolium), methoxide (methanol) and formate (formic acid). To help correct some of the problem we have added diffuse *p* and *sp* basis functions to our basis set. Still, we were not able to reproduce the experimental values exactly for the monomeric units; so we included an

empirical correction factor to each intermediate structure. Although our model is not perfect, we have attempted to correct for its limitations as best as one can do at the current time.

The other important environmental factor which we incorporated in the molecular mechanics model was the inclusion of solvent. We have opted for an explicit water model and have shown that for small anions we can well reproduce the experimental solvation energies. The most important point to make here is the dramatic effect which solvent has on the energy profile for a chemical reaction. In the formamide study, the gas phase reaction was found to go completely downhill for the first part of the reaction. However, in solvent there was found to be a barrier of ≈ 22 kcal/mol associated with the desolvation of the hydroxide anion. This result is consistent with experiment and stresses the importance which solvent has in even a simple small molecule hydrolysis. Similarly, in the trypsin study, a simple model which neglected solvent in the simulation would never have been able to realize the stabilizing effect which the water structure has on the intermediate ion pairs vs. the Michaelis complex. Clearly we have shown the importance which such environmental factors have on the energy profile of chemical reactions occurring in aqueous phase.

Just as force fields are constantly evolving species which can always use improvement, so is the methodological part of the simulation. This model was an improvement over the earlier work done in our group by Wipff *et al*, but similarly, improvements need to be made to this approach. The most obvious limitation which comes to mind is an evaluation of the important entropic contributions; both configurational entropy of the solvent and con-

formational entropy. Currently, we are only able to evaluate enthalpies; ideally we would like to evaluate free energies.

Finally, the thesis has followed a logical progression from the initial development of the force field, to the simulation of a small molecule reaction in the gas phase and in aqueous solution and to a simulation of trypsin hydrolysis. The effects of solvent were shown to be dramatic and its inclusion is certainly needed to achieve proper energtics of chemical and enzymatic reactions.

FOR REFERENCE

NOT TO BE TAKEN FROM THE ROOM

BRODART CAT. NO. 23 012 PRINTED IN U.S.A.