Title
Molecular dynamics and free energy perturbation calculations on fluorocarbons and aspartyl protease inhibitors

Author
Gough, Craig Alan

Publication Date
1992

Peer reviewed|Thesis/dissertation

MOLECULAR DYNAMICS AND FREE ENERGY PERTURBATION CALCULATIONS ON FLUOROCARBONS AND ASPARTYL PROTEASE INHIBITORS

by

Craig Alan Gough

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

[redacted]

Date **University Librarian**

Degree Conferred: 9/6/92

Chapter Three of this Dissertation will be published in a future issue of the Journal of Computational Chemistry, as "On the Derivation of Fluorine and Hydrogen Atom Parameters Using Liquid Simulations", by Craig A. Gough, Stephen, E. DeBolt, and Peter A. Kollman.

ACKNOWLEDGEMENTS

The text of Chapter Three of this dissertation will be published as it appears here, as an article in a future issue of the Journal of Computational Chemistry, with coauthors Stephen E. DeBolt and Peter A. Kollman. The latter coauthor directed and supervised the research which forms the basis for this dissertation.

The publisher of the Journal of Computational Chemistry, John Wiley and Sons, Inc., has granted permission to reproduce this article as Chapter Three of this dissertation, and to have this article microfilmed by University Microfilms as part of this dissertation. I thank John Wiley and Sons for granting this permission.

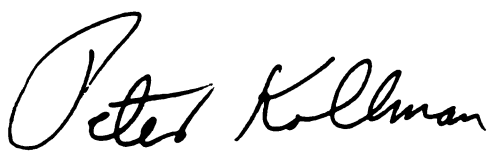
I wish to thank those who provided stimulating scientific advice during the period of my graduate research, including George Seibel, David Pearlman, and Stephen DeBolt. I also wish to thank George Seibel for invaluable assistance in helping my learn the idiosyncrasies of the computer program AMBER.

I am pleased to acknowledge research support from the NIH (GM-39552, G. Kenyon, P.I.). The facilities of the UCSF computer graphics laboratory, supported by NIH RR-1081, R. Langridge, P.I., are gratefully acknowledged, as is the supercomputer time from the San Diego Supercomputer Center, supported by the NSF.

STATEMENT BY THE CHAIRMAN OF THE DISSERTATION COMMITTEE REGARDING CO-AUTHORSHIP

This dissertation contains one chapter, Chapter Three, that has been accepted for publication in the Journal of Computational Chemistry, and which has Stephen DeBolt as a co-author. The work in this chapter consists primarily of the work of Craig Gough, but Stephen DeBolt is listed as a co-author because computer programs written by him were used to generate periodic boxes of tetrafluoromethane and trifluoromethane for molecular dynamics simulations, which were all performed by Craig Gough. The testing and optimization of fluorine van der Waals parameters by means of these simulations was carried out by Craig himself.

Although Chapter Three can stand by itself as a published paper, it is an integral part of this dissertation, as it involves the determination of parameters needed for the calculations on fluorocarbons carried out in Chapters Four and Five. Thus, Chapter Three and the rest of this dissertation together comprise a coherent whole.



Peter A. Kollman
Professor of Chemistry and Pharmaceutical Chemistry;
Chairman of the Dissertation Committee

ABSTRACT

MOLECULAR DYNAMICS AND FREE ENERGY PERTURBATION CALCULATIONS ON FLUOROCARBONS AND ASPARTYL PROTEASE INHIBITORS

by Craig Alan Gough

Adviser and Chaiman of Dissertation Committee:
Peter A. Kollman

The difference in the free energies of hydration and binding of two pencillopepsin inhibitors was calculated using the free energy perturbation (FEP) technique, using three different models for protonation state and intra-inhibitor interactions. The calculated free energy of hydration of the two inhibitors varied from 2.4 to 4.4 kcal/mole, and the calculated free energies of binding varied from 2.0 to 5.7 kcal/mole. The calculated difference between the free energies of binding and of hydration varied from 2.3 to 12.7 kcal/mole. These results agree qualitatively with the experimental value of 1.5 kcal/mole.

Molecular dynamics (MD) simulations of pure tetrafluoromethane and trifluoromethane were performed to determine the van der Waals parameters R* and ε for fluorine and for the hydrogen of trifluoromethane. The best values of R* and ε for fluorine were determined to be 1.75 Å and 0.061 kcal/mole. For the hydrogen, the optimal R* and ε were determined to be 1.21 Å and 0.015 kcal/mole.

The relative free energies of aqueous solvation of several fluorinated derivatives of methane were calculated using the FEP method. The calculations duplicated the experimental free energies relatively well, but the calculation of the bond-potential of mean force (bond-PMF) contribution was necessary in order to get the most satisfactory agreement with experiment. In addition, results of an ethanol-to-ethane perturbation in

aqueous solution show that the bond-PMF contribution is important even for FEP calculations not involving large changes in size if the length of a bond is changed during the perturbation.

MD simulations were run to determine the structure of the waters solvating fluoromethane, trifluoromethane, and tetrafluoromethane. The calculated radial distribution functions and water orientations suggest that, on average, there is one water-fluorine hydrogen bond in the case of fluoromethane. In contrast, there is no evidence of water-fluorine hydrogen-bonding in the cases of trifluoromethane or tetrafluoromethane. These results suggest that the greater aqueous solubility of fluoromethane relative to trifluoromethane is largely due to the poorer quality of the water-fluorine electrostatic interactions in trifluoromethane.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

(a)Thermodynamic cycle for the calculation of the difference in free energy of interaction between two inhibitors.
(b)Thermodyanmic cycle for the calculation of the free energy of hydration relative to the gas phase.

Figure 2.6 - Page 92-93

Key hydrogen bonds between inhibitor peptide and penicillopepsin.
(a) Difluorostatone residue.
(b) Remainder of inhibitor peptide.

Figure 2.7 - Page 94-97

Stereoviews of the environment around the hydroxyls of the statone/statine residue and Asp 33 and 213 of penicillopepsin, crystal structure and calculations using Model 1.

(a) The crystal structure of the difluorostatone/penicillopepsin complex. (Hydrogen atoms were placed randomly with reasonable constraints on bond lengths and angles, and should not be considered respresentative of the proper hydrogen bonding).

(b) A snapshot of the same structure after the addition of waters, energy minimization, and 19 picoseconds of equilibration by molecular dynamics.

(c) A snapshot of the structure resulting from mutation of the molecular dynamics-equilibrated structure to the statine (Lambda=0) form (atom OXT is now a hydrogen).

(d) A snapshot of the structure resulting from mutation of structure (c) back to the statone (Lambda=1) form.

Figure 2.8 - Page 98-100

Stereoviews of the environment around the hydroxyls of the statone/statine residue and Asp 33 and 213 of penicillopepsin, Model 2.

(a) A snapshot of the structure after equilibration by molecular dynamics.

(b) A snapshot of the structure resulting from mutation of the molecular dynamics-equilibrated structure to the statine (Lambda=0) form (atom

OXT is now a hydrogen).

(c) A snapshot of the structure resulting from mutation of structure (c) back to the statone (Lambda=1) form.

Figure 2.9 - Page 101-103

Stereoviews of the environment around the hydroxyls of the statone/statine residue and Asp 33 and 213 of penicillopepsin, Model 3.

(a) A snapshot of the structure after equilibration by molecular dynamics.

(b) A snapshot of the structure resulting from mutation of the molecular dynamics-equilibrated structure to the statine (Lambda=0) form (atom OXT is now a hydrogen).

(c) A snapshot of the structure resulting from mutation of structure (c) back to the statone (Lambda=1) form.

Figure 3.1 - Page 129

Thermodynamic cycle used for the calculation of the enthalpy of vaporization of CHF_3 at 173 K.

Figure 4.1 - Page 157

Thermodynamic cycle used for the calculation of difference in solvation free energies between two species A and B.

Figure 5.1 - Page 182

Radial distribution function for the carbon of fluoromethane and the oxygen of water.

Figure 5.2 - Page 183

Radial distribution function for the fluorine of fluoromethane and the hydrogen of water.

Figure 5.3 - Page 184

Definition of the F-H-O angle θ, as used in succeeding figures.

Figure 5.4 - Page 185

Distribution of cos θ, the F-H-O angle defined in Figure 3, for waters having a hydrogen within 2.5 Å of the fluorine of fluoromethane.

Figure 5.5 - Page 186

(a) and (c) Water orientations around the fluorine of fluoromethane that are consistent with the F-H-O angle distribution of Figure 4, but not with the C-H-O angle distribution of Figure 7.

(b) A water orientation around the fluorine of fluoromethane that is consistent with the F-H-O angle distribution of Figure 4 and with the C-H-O angle distribution of Figure 7.

(d) A water orientation that is not consistent with the F-H-O angle distribution of Figure 4 or the or the C-H-O angle distribution of Figure 7.

Figure 5.6 - Page 187

Definition of the C-H-O angle θ, as used in Figure 7.

Figure 5.7 - Page 188

Distribution of cos θ, the C-H-O angle defined in Figure 6, for waters having a hydrogen within 2.5 Å of the fluorine of fluoromethane.

Figure 5.8 - Page 189

Distribution of cos θ, as defined in Figure 6, for waters having a hydrogen at a distance between 2.5 and 5.8 Å from the fluorine of fluoromethane.

Figure 5.9 - Page 190

(a) A water orientation that is consistent with the cos θ distribution of Figure 8. (b) A water orientation that is not consistent with the cos θ distribution of Figure 8.

Figure 5.10 - Page 191

Radial distribution function for fluorine of trifluoromethane and the hydrogen of water.

Figure 5.11 - Page 192

Distribution of cos θ, as defined in Figure 3, for waters having a hydrogen within 2.5 Å of a fluorine of trifluoromethane.

Figure 5.12 - Page 193

Distribution of cos θ, as defined in Figure 3, for waters having a hydrogen at a distance between 2.5 and 4.8 Å from a fluorine of trifluoromethane.

Figure 5.13 - Page 194

Illustration of the orientation involved in a possible trifurcated hydrogen bond between water and trifluoromethane.

Figure 5.14 - Page 195

Definition of the angle θ, the angle between the O-H vector of water and the C-H vector of trifluoromethane.

Figure 5.15 - Page 196

Distribution of cos θ, the angle between the O-H vector of water and the C-H vector of trifluoromethane defined in Figure 14, for waters having a hydrogen within 4.0 Å of all three fluorines of trifluoromethane.

Figure 5.16 - Page 197

CHAPTER ONE

INTRODUCTION

MOLECULAR MECHANICS, MOLECULAR DYNAMICS, FREE ENERGY PERTURBATION CALCULATIONS, AND THE HYDROPHOBICITY OF FLUOROCARBONS

CHAPTER 1

1.1--MOLECULAR MECHANICS AND MOLECULAR DYNAMICS

This dissertation is based upon molecular mechanics and molecular dynamics calculations. The former refers to the use of minimization of an analytical expression for intra- and inter-molecular potential energy, while the latter refers to the use of this potential to calculate a classical trajectory over time. In both of these techniques, the potential energy is expressed in terms of internuclear positions, as suggested by the Born-Oppenheimer approximation. This states that the motions of the positions of nuclei relative to each other are very much slower than the velocities of electrons in molecules, and the electron distributions thus redistribute themselves essentially instantaneously on the time scale of nuclear motions. It is therefore valid to express in terms of internuclear distances only, a potential which is determined in large part by the electron density distribution.

Most of the potentials used in molecular mechanics and molecular dynamics calculations are two-body additive potentials; i.e., the intermolecular potential energy is expressed as a sum of terms, each of which represents the interaction of a pair of nuclei, and is dependent upon the distance between these two nuclei only. There have been some potentials involving nonadditive interactions[1,2].

1.1.1--Molecular Mechanics

In molecular mechanics, the potential energy of the system of interest is minimized by an algorithm that seeks a conformation for which the gradient of the potential

energy with respect to the Cartesian coordinates of the system is zero. The system thus moves "downhill" through Cartesian coordinate space to the nearest energy minimum, at which point the algorithm halts. Molecular mechanics by itself is therefore not very useful for finding the global minimum of a system, or for conformational searching. It is, however, of great utility for refining the structure of a system for which an approximate conformation has already been determined, and for comparing the energies of similar conformations of systems which differ slightly. In this vein, molecular mechanics has been used for the refinement of crystal structures of proteins and of model-built structures of macromolcules. Pioneers in this field were Levitt and Lifson[3,4], and Scheraga[5]. Molecular mechanics is now part of the standard repertoire of techniques used in the refinement of X-ray structures of macromolecules[6,7]. Energy minimization has also proved useful in the study of small organic molecules[8].

1.1.2--Molecular Dynamics

The first molecular dynamics (MD) calculations were performed by Alder and Wainwright[9,10] on a system of hard spheres. The first MD calculations using a more realistic physical model was that of Rahman[11], who simulated liquid argon using a Lennard-Jones potential. Since then, MD has been used to simulate systems with interaction potentials of increasing complexity, such as water[12–14], hydrocarbons[15–18], polar molecular liquids[19], and polymers[20,21].

The first MD simulations of a protein were performed by McCammon, Gelin, and Karplus[22], who simulated the dynamics of folded proteins using a more detailed potential. Since then, many simulations of proteins and nucleic acids have been performed; an excellent survey of the macromolecular simulations performed up until the mid

1980's is the book by McCammon and Harvey[23].

Because of the long times involved in protein folding from the denatured state, molecular dynamics has not yet proven capable of simulating this process. However, it has been used extensively to study short-time processes that relate to the questions of protein structure. For example, molecular dynamics has been used to demonstrate that units of protein secondary structure such as α-helices can move as rigid bodies[24,25]. In other simulations relating to protein structure, MD has been used to study protein fragments that retain their native structure when separated from the rest of the protein[26–28], and to study the role of water in the bending of α-helices[29,30].

Molecular dynamics calculations on liquids have been used to provide structural information that has not yet been obtained by experimental means, such as the structure of liquid water solvating various hydrophobic solutes[31,32] and polar solutes in ground and excited states[33], and the solvent structure involved in a general "solvophobic" effect[34]. MD has been used to study the structure of water at interfaces[35].

Molecular dynamics also has recently found extensive application in the refinement of macromolecular crystal structures and in the combination of two-dimensional Nuclear Overhauser Effect data and restrained molecular dynamics for the determination of macromolecular structure[36–41]. In these calculations, distance restraints obtained from experiment are included in the interaction potential; in particular, these restraints can be included in a time-averaged manner, such that they are satisfied on average but not necessarily instantaneously. These MD calculations involving restraints may be the most useful and practicable calculations that can be done with MD using the currently available computer resources, as they restrict the amount of conformational space that must be spanned to a reasonable fraction of that available without

restraints.

1.1.3--Key Algorithms of Molecular Dynamics

The key algorithm involved in any molecular dynamics simulation is the predictor-corrector algorithm, which allows a calculation of atomic positions and velocities at a future time, given the positions, velocities, and accelerations (as obtained from the gradients of the potential energy) at an earlier time. A predictor-corrector algorithm often used in its original form or in one of several variations is the Verlet algorithm[42].

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \delta t^2\mathbf{a}(t)$$

Given the positions at time t, this expression provides the positions at time $t+\delta t$, where δt is a small finite difference.

This expression is derived from a Taylor series expansion about $\mathbf{r}(t)$, by adding together the Taylor expansions for $\mathbf{r}(t)$ at times $t+\delta t$ and $t-\delta t$ and truncating after the second order term:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t) + (1/2)\delta t^2\mathbf{a}(t) + \cdots$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \delta t\mathbf{v}(t) + (1/2)\delta t^2\mathbf{a}(t) + \cdots$$

The acceleration $\mathbf{a}$ may be obtained from the potential energy V(r):

$$\mathbf{a} = -\nabla V(\mathbf{r})$$

The velocities may be obtained from the positions at times $t+\delta t$ and $t-\delta t$, in order to calculate the kinetic energy of the system:

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)}{2\delta t}$$

The molecular dynamics calculations described in this dissertation were performed using the leap-frog algorithm[43,44], which is shown below:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t+\frac{1}{2}\delta t)$$

$$\mathbf{v}(t+\frac{1}{2}\delta t) = \mathbf{v}(t-\frac{1}{2}\delta t) + \delta t \mathbf{a}(t)$$

The positions at time (t+ δt) are thus predicted from the current positions and accelerations, and from the mid-step velocities $\mathbf{v}(t-\frac{1}{2}\delta t)$. For the purpose of knowing the energy at time t, the current velocities may also be calculated:

$$\mathbf{v}(t) = \frac{1}{2}(\mathbf{v}(t+\frac{1}{2}\delta t) + \mathbf{v}(t-\frac{1}{2}\delta t)$$

A bit of algebraic manipulation will show that this algortihm is equilivalent to the Verlet algorithm. However, the leap-frog algorithm has an advantage in that the velocities enter into the calculation of the trajectory, and it is therefore possible to control the energy and temperature of the system by scaling the velocities.

The crucial limitation on the length of computer simulations that can be performed is the time step δt, as this must be significantly (approximately about an order of magnitude) smaller than the period of the fastest motion in the system of interest. At each new point in time, the time-consuming calculation of the gradient of the potential energy

must be performed.

In the case of molecular systems, the fastest motions are bond stretching vibrations; however, methods have been developed to allow the use of larger time steps, and thus less computer time per unit of time simulated, even in the presence of chemical bonds. One of these is the SHAKE algorithm[45], which iteratively readjusts coordinates at each time step so that each of the bond lengths is constrained to its equilibrium value. The bond motions thus do not enter into the equations of motion, and the time step that must be used is determined by the other, slower motions of the system.

1.1.4--Constant Temperature Molecular Dynamics

As mentioned above, it is sometimes desirable to scale the temperature of the system. In particular, it is desirable to be able to simulate a constant temperature (constant-NVT or constant-NPT) ensemble, in which the system of interest is considered in contact with a "heat bath" that maintains it at constant temperature (but not a constant total energy). Although the initial temperature of a system can be set by assigning velocities to a Maxwell-Boltzmann distribution, this distribution only describes the kinetic energy distribution of an ideal gas, and any system containing interacting particles will deviate from the desired temperature as the effect of the interparticle interactions is exerted. It is thus necessary to provide some means of controlling the kinetic energies so as to equilibrate the system at the desired temperature. It is also necessary to control the temperature to avoid non-conservation of energy due to the truncation of interatomic interactions: only interactions between those atoms within a specified cutoff distance are calculated, and thus energy is gained or lost as atoms

move in and out of each other's cutoff distance.

The algorithm used for this purpose in the calculations comprising this dissertation is that developed by Berendsen *et al.*[46]. This algortihm involves scaling the velocities by a factor χ at each time step, where

$$\chi = [\, 1 + \frac{\delta t}{t_T}(\frac{T_{ref}}{T} - 1)\,]^{1/2}$$

where T is the current temperature as calculated from the velocities, T_{ref} is the desired equilibrium temperature, δt is the time step, and t_T is a relaxation time which is adjusted to an appropriate value.

This scaling factor is designed to force the system to follow the behavior of a Langevin equation representation of temperature coupling to a heat bath. This representation leads to an expression for the time derivative of the temperature of the system:

$$dT/dt = \frac{2}{3Nk}\sum_{i=1}^{3N} v_i F_i + 2\gamma(T_{ref} - T)$$

where N is the number of particles and γ is a friction coefficient representing a frictional interaction with the heat bath.

The first term in (dT/dt) is the change in temperature due to the force exerted on each particle by the gradient of the potential energy; that is, the influence on the temperature that would occur without temperature coupling. The second term is due to coupling due to the heat bath, and is first-order with respect to (T_{ref}–T). It can be shown

that scaling the velocities by χ as defined above gives, if the temperature is expressed in terms of the squared velocities (i.e., $\frac{3}{2}kT = \frac{1}{2}mv^2$):

$$(dT/dt) = \frac{1}{3Nk}\sum_{n=1}^{3N}(2v_iF_i) + \frac{1}{t_T}(T_{ref} - T_{unscaled})$$

where $T_{unscaled}$ is the temperature derived the unscaled velocities, and $1/t_T = 2\gamma$.

In the context of the leap-frog algorithm, $T_{unscaled}$ would be $T(t - \frac{1}{2}\Delta t)$. $v_{scaled}(t + \frac{1}{2}\delta t)$ would be determined by calculating $v(t + \frac{1}{2}\delta t)$ from the leap-frog algorithm, and then scaling by χ.

The aforementioned temperature-coupling scheme has the disadvantage that does not reproduce the time-dependent properties of any known ensemble, making it impossible, for example, to determine heat capacities from the energy fluctuations. There exists another temperature coupllng scheme, developed by Andersen[47] which correctly simulates the canonical (NVT) ensemble. For example, a simulation of a two-atom harmonic oscillator molecule shows the probability distribution of the deviation of the interatom distance expected for a harmonic oscillator in the canonical ensemble; namely, a Gaussian distribution centered on the equilibrium bond length. This algorithm causes the system of interest to undergo periodic stochastic collisions which transfer kinetic energy to and from the system.

1.1.5--Periodic Boundary Conditions and Constant Pressure Molecular Dynamics

In the case of systems such as liquids and solutes dissolved in liquids, it is necessary to perform the simulation with a finite number of particles, yet avoid edge effects. For example, if a system of waters were simulated as a droplet containing a finite number of waters, the waters on the surface of the droplet would have orientations uncharacteristic of bulk water. For the microscopic droplet size which must be used in a practical calculation, this would have undesirable consequences in terms of the properties of the water, and in terms of interactions with a solute.

One way of dealing with this problem is to used a droplet of freely moving waters surrounded by a region of waters in a fixed orientation that is characteristic of a snapshot of a realistic bulk water structure; this approach is used to solvate an inhibitor in Chapter 2 of this dissertation.

A more widely used approach is the periodic boundary conditions method[48]. This involves performing the simulation with a finite number of molecules, but in such a way that this group of molecules in effect is surrounded by an infinite number of images of itself. That is, this group of molecules has an image of itself as each of its six nearest neighbors, and each image has an image of itself as each of *its* size nearest neighbors, etc. Each periodic image consists of molecules confined to some geometrical shape; the simples image possible is a cube, but shapes as complex as a truncated octahedron have been used. In the case of a cube, the system is defined by characteristic box dimensions l_x, l_y, l_z.

A periodic simulation can be run in two ways: as a constant volume simulation, in which the box size remains constant; or as a constant pressure simulation, in which the box size changes to maintain a constant pressure. The pressure can be calculated from[46]:

$$P = \frac{2}{3V}(E_k - \Xi)$$

where Ξ is the internal virial for pair-additive potentials, as defined below:

$$\Xi = -\frac{1}{2}\sum_{i<j} \mathbf{r}_{ij} \cdot \mathbf{F}_{ij},$$

$$\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$$

and $\mathbf{F}_{ij}$ is the force on particle i due to particle j.

The algorithm used to maintain constant pressure in the simulations described in this dissertation is that of Berendsen *et al.*[46], which couples the system to a constant pressure bath, in a manner similar to that used by the same authors to maintain constant temperature in a molecular dynamics simulation (as described in Section 1.1.4 of this chapter). This algorithm controls the pressure by scaling the volume of the system. For an isotropic system in a cubic box, the coordinates and box length l are scaled, each time step, from x and l to μx and μl, where μ is given by:

$$\mu = [1 - \frac{\beta \delta t}{t_P}(P_{ref} - P)]^{1/3}$$

where β is the isothermal compressibility, δt is the time step, t_P is a relaxation time, P_{ref} is the reference pressure at which the system is to be maintained, and P is the current pressure calculated from the virial. Because the volume V of the system depends upon the cube of the coordinates and the box length, this scaling adds a "bath" term to the change of volume with time:

$$(dV/dt)_{bath} = -\frac{\beta(P_{ref} - P)}{t_P}$$

Recalling that

$$\frac{dP}{dt} = -\frac{1}{\beta V}\frac{dV}{dt},$$

we have:

$$(dP/dt)_{bath} = \frac{P_{ref} - P}{t_P}$$

as the added "bath" term controlling the evolution of the pressure as a function of time.

1.2--THE CALCULATION OF FREE ENERGIES USING MOLECULAR DYNAMICS SIMULATIONS

One of the most useful applications of molecular dynamics has been in the calculation of ensemble averages which can be used to obtain the difference in free energy between two species. Free energy calculations by computer simulation in general seek to find the free energy difference between two species. There are three main methods in use for this purpose, all of which use the coupling factor approach developed by Kirkwood[49]. This approach is based on the introduction of a parameter λ into the Hamiltonian of the system, in the fashion originally proposed by Kirkwood; given two states A and B, one state would be represented by a Hamiltonian in which λ equals 0, and the other would be represented by a Hamiltonian in which λ equals 1. Intermediate, nonphysical states are represented by a lambda between 0 and 1. This parameter is similar

to a reaction coordinate, in that it determines the degree to which the Hamiltonian of the system resembles the Hamiltonians of the two systems A and B.

There are two methods commonly in use for representing a state intermediate between the λ=0 and the λ=1 states. In one method, that used in the AMBER free energy module[50,51], and used in this dissertation, a given parameter in the Hamiltonian is composed of a linear combination of the parameters assigned to the two extreme states. For example, the atomic partial charge on a given atom, for a given value of λ, would be equal to

$$q_\lambda = \lambda q_{\lambda=1} + (1-\lambda) q_{\lambda=0}$$

One drawback of this method of coupling the two extreme states is that changes in geometry of a system that are not reflected in the Hamiltonian, such as changes in an equilibrium bond length that is fixed by the SHAKE slgorithm, do not contribute to the calculated free energy difference, and that the contribution to the free energy due to geometry changes must be included explicitly[52]. This will be discussed in more detail in Chapter Four, in the context of the relative free energies of hydration of several fluorinated derivatives of methane.

Another means of coupling the Hamiltonians of the two states is to make the Hamiltonian of an intermediate state a linear combination of the two endpoint Hamiltonians; that is,

$$H_\lambda = \lambda H_{\lambda=1} + (1-\lambda) H_{\lambda=0}$$

This approach is used in the CHARMM molecular dynamics package[53,54].

This approach to coupling does include free energy changes due to changes in geometry, since some both states exist simultaneously during the calculation, and are mixed according to the expression above.

One method of calculating the difference in free energy between two species is thermodynamic integration. In thermodynamic integration, the difference in free energy between two species represented by the extreme values of λ=0 and λ=1 is equal to:

$$\Delta G = \int_0^1 \frac{\delta G}{\delta \lambda} d\lambda$$

$$= \int_0^1 \frac{\delta}{\delta \lambda} (-kT \ln \int exp\,[-H(\lambda)/kT]\,)\, d\lambda$$

$$= \int_0^1 (-kT)^{-1} (-kT \frac{\int H(\lambda) exp\,[-H(\lambda)/kT]}{\int \exp[-H(\lambda)/kT}) d\lambda$$

$$= \int_0^1 <\frac{\delta H}{\delta \lambda}>_\lambda d\lambda$$

In practice, the integral is integrated numerically over a range of ensemble averages $<\delta H/\delta\lambda>_\lambda$. Each ensemble average is calculated as a time average over the duration of an MD trajectory at that value of λ.

A second method is the perturbation theory method, using an expression originally derived by Zwanzig[55]. Consider the difference in Helmholtz free energies between two species A and B with Hamiltonians H_A and H_B, expressed in terms of their partition functions:

$$G_B - G_A = \Delta G = -kT \ln \frac{\int exp\,[-H_B/kT]}{\int \exp[-H_A/kT]}$$

where the integral is performed over all of phase space.

Multiplying the integrand in the numerator by $1 = exp\,[H_A/kT]exp\,[-H_A/kT]$ gives:

$$\Delta G = -kT \ln \frac{\int exp\,[-(H_B - H_A)/kT]\; exp\,[-H_A/kT]}{\int exp\,[-H_A/kT]}$$

$$= -kT \ln \langle exp\,[-\Delta H/kT] \rangle_A$$

where $\langle\rangle_A$ indicates an ensemble average over the phase space charcteristic of state A. Again, this is calculated as a time average over the duration of an MD trajectory.

This method works well only if the states A and B are not very different; if this is not true, the trajectory calculated with a Hamiltonian characteristic of state A will not sample, within a reasonable amount of simulation time, all of the regions for which $exp\,(-\Delta H/kT)$ makes a significant contribution to the ensemble average $\langle exp\,[-\Delta H/kT] \rangle_A$. To remedy this problem, a ΔG calculation is usually split into several calculations between intermediate states, each of whose Hamiltonian is represented by an intermediate value of λ.

A third method of calculating ΔG is slow growth, using an expression derived from the perturbation theory expression by Taylor expansion about ΔH of the exponential and the logarithm in the perturbation expression, and truncating where appropriate for small values of ΔH:

$$\Delta G = -kT \ln \langle exp\,[-\Delta H/kT] \rangle_A$$

$$= -kT \ln \langle 1 - \Delta H/kT + ... \rangle_A$$

$$= -kT \ ln(1 - <\Delta H/kT>_A + ...)$$

$$= -kT < - \Delta H/kT>_A + \cdots$$

$$= \Delta H$$

If the simulation is run such that λ changes at every time step; i.e., the change in λ per time step is equal to 1 divided by the total number of time steps in the simulation, then ΔH may be small enough to allow the truncation of the Taylor series in the derivation above.

The same expression can be derived from the expression used in thermodynamic integration if one makes the assumption that $<\delta H/\delta\lambda>_\lambda = \delta H/\delta\lambda$, and

$$\Delta G = \int_0^1 \Delta H d\lambda$$

There are assumption in this derivation that is not valid in general, but may be valid in certain special cases. The assumptions that $<\Delta H> = \Delta H$, and that $<\delta H/\delta\lambda> = \delta H/\delta\lambda$ are not true in general, even if the system is in equilibrium, since equilibrium consists of an ensemble of states, some of which may have different energies, rather than a single state. If the system has significant probability of occurring in several states, this assumption is assuredly invalid. An additional problem is a "lag" in the Hamiltonian, in which the trajectory never quite keeps up to the change in the Hamiltonian as λ changes[56]. However, slow growth does seem to work well, and give better results than the perturbation method, in cases involving large van der Waals changes between the two species of interest[50]. The reasons for this, and a theoretical elucidation of when and where slow growth should be expected to work, need further exploration.

There are two major problems involved in calculating accurate free energies. One, of course, is that the energies calculated in the Hamiltonian depend upon the molecular mechanical parameters used. The other is that a time average is used in place of an ensemble average; any simulation performed in a finite time will only approximate the actual ensemble average. For systems with only a few degrees of freedom, the available computer resources can calculate a very good approximation to the actual ensemble average. But for larger systems this is not possible. What is even more problematic is that it is not possible to know when the calculation has converged, and no upper limit can be placed on the error bounds. The hysteresis between a "forward" and a "backward" free energy calculation provides only a *lower* limit on the error.

The calculation of free energies by means of computer simulation became very prevalent in the mid 1980's, with the advent of sufficient computer power. Among the first was Jorgensen[57], who calculated the free energy difference between methanol and ethane in aqueous solution by using a Monte Carlo method[58] to evaluate the ensemble average in the perturbation theory expression.

Since then, the use of molecular dynamics calculations for the calculation of ΔG's has seen extensive application in areas relevant to the interaction of enzyme inhibitors. Such calculations include the free energy of the binding of NADP to dihydrofolate reductase[59], binding of peptide inhibitors to aspartyl proteases[60], and the free energies involved in ion binding and transport by valinomycin[61].

These techniques have been ideally suited for theoretical studies of the effects of site-directed mutagenesis of proteins. The effects of amino acid substitutions upon binding and catalysis have been studies in the cases of subtilisin[62] and ribonuclease *T1*[63].

1.3--HYDROPHOBICITY

Hydrophobicity has been and is still a somewhat controversial subject, both in terms of its definition and its molecular underpinnings. This has been discussed at length by Dill[64] in a recent review of the forces involved in protein folding, to which I am indebted for much of the information in this section. The term hydrophobicity usually refers to the low solubility of certain compounds, such as benzene and other hydrocarbons, in water. The dissolution of hydrophobic compounds in water is accompanied by an increase in the heat capacity ΔC_P[65–68]. At room temperature, a negative excess entropy of hydration, ΔS_{hyd} is also observed[65,69,70]

However, at higher temperatures, ΔS_{hyd} becomes less negative, and as the temperature increases actually becomes positive; at these temperatures the positive ΔG_{hyd} is due to ΔH_{hyd} rather than ΔS_{hyd}[54,65]; at room temperature ΔH_{hyd} is negative or a small positive value. It therefore may be preferable to define hydrophobicity in terms of the ΔC_P, since the ΔC_P is positive regardless of the temperature.

The proposed molecular explanations of the hydrophobic effect seek to explain the negative ΔS_{hyd} at room temperature in microscopic terms. The first attempt at an explanation for the molecular basis for the observed thermodynamics of the hydration of hydrophobic species was advanced by Frank and Evans[66]. They explained the negative entropy of hydration in terms of an increased structuring of the hydrating waters; i.e., the water structure around the solute becomes more tetrahedral or ice-like.

This interpretation has been supported and clarified by computer simulations[31,71–73]. The molecular dynamics simulations of Rossky and Karplus[31] are illustrative of the general concept. These calculations involved studies of the water structure around the methyl group of alanine dipeptide, showed that the waters solvating this

hydrophobic moiety tend to have their charges (lone pairs and hydrogen atoms) pointing away from the methyl group carbon; the C-H-O angle distribution has a maximum corresponding to zero degrees. The other charges of a given water molecule are oriented in a broad distribution of angles corresponding to approximately 95 degrees to 145 degrees, or somewhat tangential to the surface of the methyl group.

These molecular dynamics studies showed that each water solvating the methyl group, however, maintains more or less unchanged the number of hydrogen bonds with other waters that it would have in the bulk solution. The fraction of waters forming between zero and four hydrogen bonds is about the same both in bulk solution and near the methyl group. This is consistent with evidence from nuclear magnetic resonance studies[64,75] that indicate no significant change in the degree of hydrogen bonding in water due to the dissolution of nonpolar solutes. However, the fraction of waters having five or more hydrogen bonds is decreased, and this is compensated for by a slight *increase* in the fraction of waters forming four hydrogen bonds.

It therefore seems that the presence of a hydrophobic species does not disrupt water-water hydrogen bonds, but rather introduces a geometrical constraint upon the waters that must be satisfied if hydrogen bonding similar to that in bulk water is to be maintained. The C-H-O angle distribution discussed above is the ideal one to maximize water-water hydrogen bonding near the surface of a hydrophobic solute. This constraint decreases the amount of conformational space available to each water near the solute, and can explain the negative ΔS_{hyd}: the entropy of the waters decreases because of the reduction in the number of librational states available to the waters.

This microscopic explanation of hydrophobicity can also account for the positive ΔC_P[64]. Consider what occurs as the temperature of a solution of water and a

hydrophobic solute is raised. The water-water hydrogen bonds that are predominant at a lower temperature are to a certain extent bent or broken at a higher temperature, since the Boltzmann probability of energetically unfavorable states becomes higher at a higher temperature; the $T\Delta S$ contribution to the ΔG becomes more important at a higher temperature. The energetically unfavorable configurations contribute to a more positive enthalpy at a higher temperature. $\Delta C_P = (\delta H/\delta T)_P$, and this model describes a mechanism by which a solution of a hydrophobic solute in water exihibits a greater increase of enthalpy with temperature than does a reference state with water and the hydrophobic solute separated.

Regardless of the molecular mechanism accounting for hydrophobicity, the hydrophobicity of nonpolar compounds, as evidenced by the entropy and heat capacity of transfer of nonpolar solutes into water, is proportional to the surface area of the solutes[76–79]. A hydrophobic molecule creates a cavity in water, and hydrophobicity apparently increases with the size of this cavity.

This is consistent with scaled particle theory (SPT)[80], which has been extended to solutes in water by Pierotti[81]. This theory describes the solution process in terms of two steps. First, a cavity large enough to accommodate the solute is created in the solvent. Next, the solute molecule, with a suitable interaction potential, is introduced in this cavity and interacts with the solvent. The SPT calculations of Pierotti[81], involving the dissolution of argon and N_2 in water at 298 K, gave a free energy of cavity formation that was positive and almost entirely determined by a negative entropy change. The large negative entropy of hydration is due to the entropy of cavity formation. A positive ΔC_P was also calculated for this process. SPT says nothing about the molecular mechanism of this, since it involves no assumptions about the structure of the solvent.

The correlation of the hydrophobicity of a nonpolar solute with its water-accessible surface area has been used to quantify the magnitude of the hydrophobic contribution to protein stability[82,83]. This dependence on surface area suggests that the determinant of hydrophobicity is the extent to which waters are deprived of contact with other waters, and is consistent with the molecular mechanism described above.

1.4--THE PROPERTIES OF FLUOROCARBONS AND FLUOROALCOHOLS AND THEIR HYDROPHOBICITY

The phenomenological definition of hydrophobicity in terms of a ΔC_P of dissolution in water, as discussed in the previous section of this chapter, is relevant to the behavior of fluorocarbons and fluoroalcohols. Table 1.1 shows the changes in free energy, entropy, and heat capacity occuring when several fluoroalcohols[84], methane[85], and several fluorinated methane derivatives[86–88] are dissolved in water. These data show a negative ΔS upon dissolution of these species in water. As with hydrocarbons and hydrocarbon alcohols, there is a positive ΔC_P upon dissolution of tetrafluoromethane and the fluoroalcohols in water. The ΔC_P of tetrafluoromethane is almost twice that of methane, and the ΔC_P of a given fluoroalcohol is approximately twice as large as the ΔC_P for the corresponding alcohol with all fluorines replaced by hydrogens.

These data can be interpreted to mean that some fluorocarbons and fluoroalcohols can be considered hydrophobic, and indeed are more hydrophobic than the corresponding hydrocarbon. In particular, tetrafluoromethane is less soluble in water than methane, and the fluoroalcohols containing one or more trifluoromethyl groups are less soluble in water and produce a larger ΔC_P of solution than the corresponding

hydrocarbon alcohol. Although it is not too surprising that tetrafluoromethane behaves hydrophobically, as it has a zero dipole moment, it is interesting that the heat capacity data suggests that the polar trifluoromethyl group can be considered hydrophobic, and more hydrophobic than the relatively nonpolar methyl group.

The solubility behavior of fluorocarbons can be discussed in terms of the simple microscopic models mentioned in the previous section. Replacing one or more hydrogens by fluorines increases the size of the cavity occupied by the solute in solution, due to both the larger van der Waals radius of fluorine, and the increased length of a carbon-fluorine bond versus a carbon-hydrogen bond. This would make the ΔS more negative, all other factors remaining the same. In terms of the molecular model of the hydrophobic effect, more order would be imposed upon the solvating waters due to the greater surface area of the hydrophobic cavity. This explanation is likely the cause of the ΔS for dissolution of tetrafluoromethane, which has no dipole moment and could reasonably be expected to interact with waters much in the manner of a larger version of methane. One could advance the argument that since fluorine is not very polarizable, CF_4 lacks the favorable solute-water dispersion interactions found with methane, but the polarizability data for the CF_4 group as a whole does not support this interpretation. The dispersion interactions between CF_4 and water actually should be more favorable than those between CH_4 and water, as CF_4 has a greater polarizability than methane[89], leaving the "cavity" explanation as the obvious one. The FEP calculations discussed in Chapter 4 support this "cavity" interpretation also, and suggest a net decrease in the free energy of hydration of CF_4, relative CH_4, due to dispersion.

The discussion above, of course, assumes that the orientations and mobility of the solvating waters is similar for both fluorocarbons and hydrocarbons. This is most

likely not the case for some fluorocarbons; for example, fluoromethane shows a smaller ΔS of hydration than methane, even though its cavity size would be larger. This data suggests that local electrostatic interactions between the fluorine and the waters may result in a solvent water shell that is significantly different in orientation from that around methane. There may be some water-fluorine hydrogen bonding, though it would be expected to contribute less to the free energy of solvation than the hydrogen bonding to methanol, the hydroxyl of which can form three hydrogen bonds to waters[57]. The water structure around, say, fluoromethane, may not resemble that around methane, as there may be some directional hydrogen-bonding interactions between the waters and the fluorine. Scaled-particle theory does not deal well with solutes having strong directional interactions with solvent; under these conditions, it is questionable whether the cavity and interaction terms are simply additive[90]. So the simple model of a "cavity" term added to an interaction term may not work well for calculating the hydration thermodynamics for fluoromethane. The model of scaled-particle theory would be expected to work well for both methane and tetrafluoromethane, as neither have a net dipole moment, and in fact, the thermodynamics have hydration of tetrafluoromethane have been calculated quite accurately using scaled particle theory[88]. Trifluoromethane, although polar, may not interact in a hydrogen-bonding manner with water, and thus scaled particle theory may be adequate to explain its solubility in water.

Trifluoromethane, which has a dipole moment similar to fluoromethane, is 1 kcal/mole less soluble in water than fluoromethane. There are at least two possible explanations for this decreased solubility, and both may be true to some extent. One is that the greater solubility of this species versus methane is due to more general dipole-water interactions, rather than hydrogen bonding. The other is that hydrogen bonding is

present, but the non-hydrogen-bonded waters, which adopt a typical "hydrophobic" orientation and ordering around the solute, see a bigger cavity. No data is available for the ΔS of hydration of trifluoromethane. The trifluoromethyl groups of the fluoroalcohols may be considered somewhat analogous, although the situation may be complicated somewhat by the interactions of the hydrogen of trifluoromethane and the hydrogen attached to the carbon bearing the hydroxyl in the fluoroalcohols. These interactions will contribute to the thermodynamic quantities involved in hydration, in addition to the fluorine-water nonbonded interactions.

The "hydrophobic" properties that appear to be exhibited by the trifluoromethyl group could have implications for rational drug design. Since the trifluoromethyl group is polar, it could have a favorable electrostatic interaction with a region of favorable electrostatic potential gradient in an enzyme active site, and with solvating waters. However, its hydrophobicity would contribute a positive term to its free energy of hydration. Since a trifluoromethyl group is more polar than a methyl group, and possibly more hydrophobic also, the result of replacing a methyl group with a trifluoromethyl group could be a net negative change in the free energy of desolvation and binding to an enzyme active site; this would increase the binding affinity.

Consider the following analysis, which is similar to an analysis performed by Fersht of the net effects of substitution of groups on the free energy of desolvation and binding and on protein stability[91,92]. In the absence of any specific, directional electrostatic interactions such as hydrogen bonding between the trifluoromethyl group and the solvating waters (that is to say, if the water structure around the trifluoromethyl group is that typically found around hydrophobic solutes), the only negative changes in the free energy of solvation upon transformation of a methyl group into a trifluoromethyl group

will be approximately that that of charging a dipole in a dielectric medium[93], and a small increase in the dispersion interactions due to the expected greater polarizability of the trifluoromethyl group (trifluoromethane is slightly more polarizable than methane[84]). In the enzyme active site, a suitable electrostatic environment could lead to electrostatic interactions with the trifluoromethyl group that are comparable to those between the trifluoromethyl group and water. Thus, considering only electrostatics, the net free energy of replacing a methyl group with a trifluoromethyl group could be negative, positive, or approximately zero; in any case, the electrostatic contribution could be small.

The other contribution to be considered, however, is the entropic effect of replacing the hydrophobic cavity of a methyl group with the larger cavity of a trifluoromethyl group. If the water structure around a trifluoromethyl group is typically that found around hydrophobic solutes, this replacement will lead to a net entropy decrease; the negative ΔS of hydrating a trifluoromethyl group at 298 K will greater in magnitude than that of hydrating a methyl group. This contribution could swing the balance of contributions to the free energy of desolvation and binding in favor of binding, since the increased size of the trifluoromethyl group may have a negligible effect on the binding free energy.

Compare this with the free energy of replacing a methyl group with a hydrophilic polar moiety, such as a hydroxyl group. Here, a nonpolar methyl group is replaced with moiety capable of hydrogen bonding. The electrostatic environment of the enzyme may be very favorable to the binding of such a group; but in the absence of ionic residues in the active site, this effect could be compensated for by the loss of hydroxyl-water hydrogen bonds. Thus the net change in the free energy of desolvation and

binding would be very small.

In view of the time-consuming nature of synthesizing an enzyme ihibitor of the desired composition, computer simulation methods provide a means of testing the feasibility of new ideas for inhibitor enhancement such as the one described above. The major thrust of this dissertation is the exploration of the properties of fluorocarbons in aqueous solution, and the testing of the FEP method as applied to fluorocarbons in aqueous solution. Chapter Two of this work describes free energy calculations of the relative binding free energy of two fluorinated aspartyl protease inhibitors to the aspartyl protease penicillopepsin; because of the size of the system, only a qualitatively correct prediction of this free energy was possible. The results of these calculations led to a desire to improve our ability to simulate the interactions of fluorocarbons with water. The remaining chapters involve simulations of fluorocarbons, either pure or in aqueous solution, with the aim of improving our ability to use molecular dynamics and FEP calculations accurately in drug design. In Chapter Three, van der Waals parameters for fluorine, for the purpose of simulation of fluorocarbons, are derived. Chapter Four involves the determination of the relative free energies of hydration of the fluorinated derivatives of methane using the FEP method. Chapter Five explores the reasons for the different solubilities of fluoromethane, trifluoromethane, and tetrafluoromethane by analyzing the water structure around these species by means of molecular dynamics simulations, and by performing FEP calculations. This chapter also aims to investigate the reasons for the "hydrophobicity" of the fluorocarbons, and in particular, to see if a trifluoromethyl group really is hydrated in a manner typical of hydrophobic moieties such as a methyl group.

REFERENCES FOR CHAPTER ONE

(1) Caldwell, J.; Dang, L.X.; Kollman, P.A. *J. Am. Chem. Soc.* 1991, **112**, 9144.

(2) Ahlstrom, P.; Wallqvist, A; Engstrom, S; Jonsson, B. *Mol. Phys.* 1989, **68**, 563.

(3) Levitt, M.; Lifson, S. *J. Mol. Biol.* 1969, **46**, 269.

(4) Levitt, M. *J. Mol. Biol.* 1974, **82**, 393.

(5) Scheraga, H.A. *Adv. Phys. Org. Chem.* 1968, **6**, 103.

(6) Jack, A.; Levitt, M. *Acta Crystallographica* 1978, **A34**, 931.

(7) Konnert, J.H.; Hendrickson, W.A. *Acta Crystallographica* 1980, **A36**, 344.

(8) Molecular Mechanics, U. Burkert; N.L. Allinger. American Chemical Society, Washington. 1982.

(9) Alder, B.J.; Wainwright, T.E. *J. Chem. Phys* 1957, **27**, 1208.

(10) Alder, B.J.; Wainwright, T.E. *J. Chem. Phys* 1959, **31**, 459.

(11) Rahman, A. *Phys. Rev.* 1964, **136A**, 405.

(12) Rahman, A.; Stillinger, F.H. *J. Chem. Phys* 1971, **55**, 3336.

(13) Stillinger, F.H. *Science* 1980, **209**, 451.

(14) Stillinger, F.H. *Adv. Chem. Phys.* 1975, **31**, 1.

(15) Ryckaert, J.P.; Bellemanns, A. *Chem. Phys. Lett.* 1975, **30**, 123.

(16) Ryckaert, J.P.; Bellemanns, A. *Chem. Soc. Farday. Discuss.* 1978, **66**, 95.

(17) Marechal, G.; Ryckaert, J.P. *Chem. Phys. Lett.* 1983, **101**, 548.

(18) Harris, J.G.; Stillinger, F.H. *J. Chem. Phys.* 1991, **95**, 5953.

(19) Yashonath, S.; Rao, C.N.R. *Chemical Physics* 1991, **155**, 351.

(20) Kremer, K.; Grest, G.S. *J. Chem. Phys.* 1990, **92**, 5057.

(21) Blonski, S.; Brostow, W. *J. Chem. Phys.* 1991, **95**, 2890.

(22) McCammon, J.A.; Gelin, B.R.; Karplus. M. *Nature* 1977, **267**, 585.

(23) Dynamics of Proteins and Nucleic Acids. J.A. McCammon; S.C. Harvey. Cambridge University Press, Cambridge, U.K. 1987.

(24) Levy, R.M.; Sheridan, R.P.; Keepers, J.W.; Dubey, G.S.; Swaminathan, S.; Karplus, M. *Biophys. J.* 1985, **48**, 509.

(25) Aqvist, J.; van Gunsteren, W.F.; Leijonmarck, M.; Tapia, O. *J. Mol. Biol.* 1985, **183**, 461.

(26) Tiradorives, J.; Jorgensen, W.L. *Biochemistry* 1991, **30**, 3864.

(27) Klaus, W.; Dieckmann, T.; Wray, V.; Schomburg, D.; Wingender, E.; Mayer, H. *Biochemistry* 1991, **30**, 6936.

(28) Aqvist, J.; Tapia, D. *Biopolymers* 1990, **30**, 205.

(29) Di Capua, F.M.; Swaminathan, S.; BBeridge, D.L. *J. Am. Chem. Soc.* 1991, **113**, 6145.

(30) Di Capua, F.M.; Swaminathan, S.; Beveridge, D.L. *J. Am. Chem. Soc.* 1990, **112**, 6768.

(31) Rossky, P.J.; Karplus, M. *J. Am. Chem. Soc.* 1979, **101**, 1913, and references therein.

(32) Zichi, D.A.; Rossky, P.J. *J. Chem. Phys.* 1985, **83**, 797.

(33) DeBolt, S.E.; Kollman, P.A. *J. Am. Chem. Soc.* 1990, **112**, 7515.

(34) Tobia, D.J.; Brooks, C.L. *J. Chem. Phys.* 1990, **92**, 2582.

(35) Ahlstrom, P.; Teleman, O; Jonsson, B. *Chemica Scripta* 1989, **29A**, 97.

(36) Brunger, A.T.; Karplus, M. *Acc. Chem. Res.* 1991, **24**, 54.

(37) Brunger, A.T. *Ann. Rev. Phys. Chem.* 1991, **42**, 197.

(38) Torda, A.E.; van Gunsteren, W.F. *Comp. Phys. Comm* 1991, **62**, 289.

(39) Gros, P.; van Gunsteren, W.F.; Hol, W.G.J. *Science* 1990, **249**, 1149.

(40) Torda, A.E.; Scheek, R.M.; van Gunsteren, W.F. *J. Mol. Biol.* 1990, **214**, 223.

(41) Pearlman, D.A.; Kollman, P.A. *J. Mol. Biol.* 1991, **220**, 457.

(42) Verlet, P. *Phys. Rev.* 1967, **159**, 98.

(43) Hockney, R.W. *Methods comput. Phys.* 1970, **9**, 136.

(44) Computational Physics. D. Potter. Wiley, New York. 1972.

(45) Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. *J. Comput. Phys.* 1977, **23**, 327.

(46) Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; Di Nola, A.; Haak, J.R. *J. Chem. Phys.* 1984, **81**, 3684.

(47) Andersen, H.C. *J. Chem. Phys.* 1980, **72**, 2384.

(48) Computer Simulation of Liquids. M.P. Allen and D.J. Tildesley. Oxford Science Publications, Oxford, UK, 1987.

(49) Kirkwood, J.G. *J. Chem. Phys.* 1935, **3**, 300.

(50) Singh, U.C.; Brown, F.K.; Bash, P.A.; Kollman, P.A. *J. Am. Chem. Soc.* 1987, *109*, 1607.

(51) Pearlman, D.A.; Case, D.A.; Caldwell, J.C.; Seibel, G.L.; Singh, U.C.; Weiner, P.; Kollman, P.A. (1991), Amber 4.0, University of California, San Francisco.

(52) Pearlman, D.A.; Kollman, P.A. *J. Chem. Phys.* 1991, **94**, 4532.

(53) Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* 1983, **4**, 187.

(54) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. *Science* 1989, **244**, 1069.

(55) Zwanzig, R.W. *J. Chem. Phys.* 1954, **22**, 1420.

(56) Pearlman, D.A.; Kollman, P.A. *J. Chem. Phys.* 1989, **91**, 7831.

(57) Jorgensen, W.L.; Ravimohan, C.. *J. Chem. Phys.* 1985, **83**, 3050.

(58) Monte Carlo Methods in Statistical Physics. K. Binder, ed. Springer-Verlag, Berlin, FRG. 1986.

(59) Cummins, P.L.; Ramnarayan, K.; Singh, U.C.; Gready, J.E. *J. Am. Chem. Soc.* 1991, **113**, 8247.

(60) Rao, B.G.; Singh, U.C. *J. Am. Chem. Soc.* 1991, **113**, 6735.

(61) Eisenman, G.; Aqvist, J.; Alvarez, O. *J. Chem. Soc.-Faraday Trans.* 1991, **87**, 2111.

(62) Mizushima, N.; Spellmeyer, D.; Hirono, S.; Pearlman, D.; Kollman, P.A.; *J. Biol. Chem.* 1991, **266**, 11801.

(63) Hirono, S.; Kollman, P.A. *Protein Engineering* 1991, **3**, 233.

(64) Dill, K.A.; *Biochemistry* 1990, **29**, 7133.

(65) Privalov, P.L.; Gill, S.J. *Adv. Protein Chem.* 1988, **39**, 191.

(66) Frank, H.S.; Evans, M.W. *J. Chem. Phys.* 1945 **13**, 507.

(67) Christian, S.D.; Tucker, E.E. *J. Solution Chem.* 1982, **11**, 749.

(68) Edsall, J.T.;McKenzie, H.A. *Adv. Biophys.* 1983, **16**, 53.

(69) The Hydrophobic Effect. 2nd edition. C. Tanford. Wiley, New York, 1980.

(70) Gill, S.J.; Wadso, I. *Proc. Natl. Acad. Sci. U.S.A.* 1976, **73**, 2595.

(71) Geiger, A.; Rahman, A.; Stillinger, F.H. *J. Chem. Phys.* 1979, **70**, 263.

(72) Pangali, C.; Rao, M.; Berne, B.J. *J. Chem. Phys.* 1982, **71**, 2982.

(73) Ravishanker, G.; Mezei, M.; Beveridge, D.L. *Faraday Symp. Chem. Soc.* 1982, **17**, 79.

(74) Balazs, E.A.; Bothner-By, A.A.; Gergely, J. *J. Mol. Biol.* 1959, **1**, 147.

(75) Bovey, F.A. *Nature* 1961, **192**, 324.

(76) Miller,, K.W.; Hildebrand, J.H. *J. Am. Chem. Soc.* 1968, **90**, 3001.

(77) Gill, S.J.; Dec, S.F.; Olafsson, G.; Wadso, J. *J. Phys. Chem.* 1985, **89**, 3758.

(78) Jorgensen, W.L.; Geo, J.; Ravimohan, C. *J. Phys. Chem.* 1985, **89**, 3470.

(79) Jolicoeur, C., *et al.*. *J. Solution. Chem.* 1986, **15**, 109.

(80) Reiss, H. *Adv. Chem. Phys.* 1965, **9**, 1.

(81) Pierotti, R.S. *J. Phys. Chem.* 1965, **69**, 281.

(82) Sharp, K.A.; Nicholls, A.; Friedman, R.; Honig, B. *Biochemistry* 1991, **30**, 9686.

(83) **Sharp, K.A.**; Nicholls, A.; Fine, R.F.; Honig, B. *Science* 1991, **252**, 106.

(84) **Rochester, C.H.**; Symonds, J.R. *J. Chem. Soc., Faraday Transactions* 1973, **1**, 1577.

(85) Water: a Comprehensive Treatise, Felix Franks, Editor. Plenum Press, New York, 1973.

(86) **Swain, C.G.**; Thornton, E.R. *J. Amer. Chem. Soc.* 1962, **84**, 822.

(87) **Hine, J.**; Mookerjee, P.K.; *J. Org. Chem.* 1984, **40**, 292.

(88) **Wen, W.-Y.**; Muccitelli, J.A. *J. Solution Chem.* 1979, **8**, 225.

(89) Theory of Molecular Fluids. Oxford University Press, New York. 1984.

(90) **Morel-Desrosiers, N.**; Morel, J.-P. *Can. J. Chem.* 1981, **59**, 1.

(91) Enzyme Structure and Function, Enzyme Structure and Mechanism, W.H. Freeman, New York. 1985.

(92) **Fersht, A.** *Trends in Biochemical Sciences* 1987, **12**, 301.

(93) **Beveridge, D.L.**; Schnuelle, G.W. *J. Phys. Chem.* 1975, **79**, 2562.

Table 1.1

Thermodynamic Properties for Hydration of Some Fluorocarbons and Fluoroalcohols[a,b]			
Solute	ΔG^c	ΔS^d	ΔC_p^e
CH_4	6.3[f]	–31[f]	55[f]
CH_3F	4.1[g]	–27[g]	-
CHF_3	5.1[h]	-	-
CF_4	7.4[i]	–37.5[i]	98[i]
CH_3CH_2OH	–0.7[j]	–40[j]	36[j]
$CH_3CH(OH)CH_3$	–0.5[j]	–45[j]	49[j]
CF_3CH_2OH	–0.03[j]	–40[j]	76[j]
$CF_3CH(OH)CF_3$	0.5[j]	–47[j]	99[j]

[a] At 25 degrees C.

[b] All thermodynamic quantities are given for the following choice of standard states: 1 atmosphere pressure of the solute in the gas phase and a hypothetical infinitely dilute solution in which the mole fraction of the solvent is 1.

[c] In $kcal\ mol^{-1}$.

[d] In $cal\ mol^{-1}\ K^{-1}$.

[e] In $cal\ mol^{-1}\ K^{-1}$.

[f] From reference 85.

[g] From reference 86.

[h] From reference 87.

[i] From reference 88.

[j] From reference 84.

CHAPTER TWO

FREE ENERGY PERTURBATION CALCULATIONS ON DIFLUOROSTATINE AND DIFLUOROSTATONE INHIBIORS OF PENICILLOPEPSIN

CHAPTER TWO

INTRODUCTION

The aspartyl proteases comprise a class of enzymes which cleave peptide bonds, in particular those between large hydrophobic residues[1]. This class includes rhizopus pepsin from *Rhizopus chinensis*[2,3]; renin, which converts angiotensinogen to the vasoconstrictor angiotensin II and thus has clinical relevance[4,5]; penicillopepsin, found in the mold *Penicilium janthinellum*[6,7,8]; and *Endothia parasitica pepsin*[9]. The aspartyl proteases for which crystal structures have been obtained are penicillopepsin[8], *Endothia parasitica pepsin*[9], and *Rhizopus chinensis* pepsin[4]), and the proteases of the Rous Sarcoma Virus[10] and Human Immunodeficiency Virus[11,12].

The active sites of these enzymes are located at the junction of domains, each of which contributes a catalytically important aspartyl residue (residues 33 and 213 in penicillopepsin) (see reference 11 for a discussion of the catalytic importance of these residues), which are believed to share a proton, and to act to hydrolyze substrate peptides by a general acid-general base mechanism[12].

Aspartyl protease inhibitors have long been a source of interest because of the desire to design renin inhibitors for the treatment of hypertension[4,5]. Recently, additional interest in aspartyl protease inhibitors has been generated by the discovery that the the Human Immunodeficiency Virus (HIV), the causative agent of AIDS, may use an aspartyl protease[12,13,14]. The genomes of many retroviruses, including HIV, code for a protein with sequence homology to aspartyl proteases[12,13,14] which functions in

the processing of viral translational products[15,16] and which is inhibited by pepstatin, an aspartyl protease inhibitor[17]. Each retroviral protein has two highly conserved regions, with homology to two corresponding highly conserved regions in each of the two domains of aspartyl proteases. One of the conserved regions contains a strictly conserved aspartyl-threonine-glycine triad, which has led some workers to postulate models of the three-dimensional structure of retroviral protease structure in which the catalytically active structure is a dimer which resembles the bilobal structure of the aspartyl proteases[13]. These predictions have been confirmed[11,12].

The importance and clinical relevance of aspartyl proteases, especially vis-a-vis the search for AIDS treatments, has led us to look at the study of these proteins and their inhibitors as a new application of the free energy perturbation (FEP) technique[18,19] as an aid to drug design. The FEP method has been used to predict and rationalize the changes in the energetics of substrate binding and transition state formation due to site-directed mutagenesis in enzymes[20,21] and to duplicate the different binding affinities of protein inhibitors [22,23]. We believe that one of the key future uses of the FEP method is in the "fine-tuning" at the atomic level of drug structures derived from a lead compound, which may be a naturally occurring compound, or a compound derived *ab initio* by other computational methods. That is, the change in relative binding affinity caused by minor changes in a structure can be predicted by the FEP method more readily than this change could be checked experimentally, allowing the screening of large numbers of compounds without unnecessary time being spent in the synthesis of all of these compounds. With this in mind, we present in this paper the results of free energy calculations of relative binding affinities of two inhibitors of the aspartyl protease penicillopepsin.

Gelb *et al.* have synthesized peptides containing the amino acids difluorostatine and difluorostatone (Figures 2.1a and 2.1b), which is a derivative of statine, an amino acid found in the naturally occurring peptide inhibitor pepstatin[24,25]. They have shown inhibition of pepsin[26] and human renin[27] by these compounds. This inhibitor presumably works as a transition-state inhibitor or a collected-substrate inhibitor[12], since the presence of the two fluorines leads to hydration of the ketone to form a *gem*-diol (Figure 2.1c). Peptides containing either statine or difluorostatine have been found to be less tightly bound to human renin than peptides containing difluorostatone. Sielecki and James [28] have obtained crystal structures of difluorostatine- and difluorostatone-containing peptides bound to penicillopepsin. The purpose of the FEP study described in this paper is to predict the relative binding affinities of these two inhibitors starting with these crystal structures.

METHODS

The two inhibitors were synthesized by Gelb *et al.* at the University of Washington[27]. Coordinates for the crystal structures of the two inhibitors bound to penicillopepsin were obtained in the James laboratory at the University of Alberta[28].

The energy minimizations and molecular dynamics trajectories were calculated with the program AMBER, which uses a potential energy function of the following form:

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

All of the parameters needed are found in the standard AMBER force field[29,30], except for parameters involving fluorine, and the point charges of difluorostatine, difluorostatone, and the isovaleryl group.

The point charges of the novel residues used were calculated using the electrostatic potential fitting routine developed by Singh & Kollman[31,32], in which point charges on the atoms are adjusted to give an electrostatic potential that fits that calculated quantum mechanically. The charges for the difluorostatine and difluorostatone residues are a mixture of the charges calculated using two different basis sets: the key atoms involved in determining the electron density distribution at the immediate site of the perturbation were assigned charges calculated with the 6-31G* basis set, while the remaining atoms in the residue were given charges calculated with the STO-3G basis set. Because of the computational costs associated with calculations using the larger 6-31G* basis set, it was not possible to calculate 6-31G* charges for the entire difluorostatone/statine residue; rather, charges were calculated with the structure shown in **Figure 2.2a**, which contains the key atoms involved in the perturbation calculation. The STO-3G charges were calculated with a structure containing the entire residue, and then the charges on the key atoms were replaced with the 6-31G* charges calculated using the smaller structure. Figures 2.2b and 2.2c show the structures used in calculating the charges of the difluorostatine and difluorostatone residues, and the isovaleryl group attached to the N-terminal end of the inhibitor peptide. The small excess charge

remaining after this charge substitution was made was placed on the alpha carbon of the difluorostatone residue, in order to make the residue neutral; this charge was only ~0.04 of an electron. Figures 2.2d-f show the charges after this manipulation; the atoms marked with an asterisk have 6-31G* charges, while the remainder are STO-3G.

The standard AMBER parameter files do not contain parameters for fluorine, and it was therefore necessary to derive them or obtain them them from other sources. The force constants and equilibrium bond distances and angles for bond stretching and bending potential energy terms were taken from the parameter list for MM2[33] and used without modification, after being converted into the slightly different energy expressions used in AMBER; these parameters are listed in Table 2.1. The torsional parameters for dihedral angles involving a fluorine atom were taken to be identical to the parameters used for dihedral angles in which a hydroxyl oxygen is in the same position as the fluorine atom. For example, the dihedral angle in Figure 2.3a would be assigned the same parameters as the dihedral angle in Figure 2.3b. This is obviously only an approximation, but makes intuitive sense on the grounds that fluorine and oxygen are both fairly electronegative atoms, and the dihedral energy dependence due to electron delocalization effects depends primarily on the relative electronegativity of the atoms involved[34]. The torsional parameters used for dihedrals involving fluorine are listed in Table 2.2.

The van der Waals 6-12 parameters for fluorine were derived as follows: a series of single point *ab initio* calculations[31] were carried out on a methane-fluoromethane complex, in which a hydrogen of methane and the fluorine of fluoromethane were placed adjacent to each other at different F-H distance, with the C-F and C-H bonds parallel (Figure 2.4a). The 6-31G* basis set was used. The hydrogen-fluorine distance

differed by 0.1 Å in each single-point calculation.

The quantum-mechanically calculated energy of this complex was at a minimum for the calculation in which the hydrogen-fluorine distance was 2.7 Å. A series of molecular mechanical energy minimizations was performed on the same complex, in which the R* value assigned to fluorine was varied, at intervals of 0.01 Å, until the hydrogen-fluorine distance in the minimized structure was 2.7 Å; this occurred when R* of fluorine was taken to be 1.55 Å. ε for fluorine was taken to be 0.080 kcal/mole, based on the value of 0.078 used in MM2[33]. The charges of methane and difluoromethane were calculated quantum mechanically using the method described previously in this paper. The 6-12 parameters used for fluorine are given in Table 2.3.

It is not obvious how to perform a perturbation between two different potential energy terms having different functional forms, such as between a 6-12 and 10-12 potential, when the perturbation Hamiltonian is non-linear in λ. Therefore the current free energy perturbation module of AMBER (3.0 or 3A) does not allow the use of the 10-12 hydrogen-bond energy term, used by other modules of AMBER, in the calculation of the perturbation energy of the perturbed atoms.

Since the key interaction being observed in this free energy calculation is that between a hydroxyl group and the carboxyl of an aspartate, it was necessary to derive 6-12 van der Waals parameters for hydroxyl hydrogens that would reproduce the geometry of a hydroxyl-carboxyl hydrogen bond. This was accomplished by performing energy minimization on a methanol-acetate anion complex (Figure 2.4b) using the standard AMBER 10-12 parameters for this type of hydrogen bond, and then repeating the minimization using a 6-12 potential function for the hydroxyl-carboxyl interaction. The R* for the hydroxyl hydrogen was varied until the O-O distance for the 6-12 minimized

structure was 2.692 Å, which was close to the desired distance of 2.690 Å obtained in the 10-12 minimized structure. The ε value for the hydroxyl hydrogen was taken to be 0.020, the same value as used for the standard AMBER 6-12 potential for other hydrogen types. The interaction energy using these 6-12 parameters, relative to the energy of isolated methanol and acetate anion, was calculated to be -12.92 kcal/mole, as compared to -13.69 kcal/mole using the standard AMBER 10-12 parameters. The energies and geometries obtained in the above calculations are summarized in Table 2.4.

The difference in the free energy of binding between the two inhibitors is difficult to calculate directly, since it is computationally prohibitive to simulate the desolvation of an inhibitor and its subsequent binding to a protein. Instead, this difference is calculated by taking advantage of the thermodynamic cycle shown in Figure 2.5. ΔG_{sol} and ΔG_{bind} are easily calculable by leaving the inhibitor in the environment in solution or bound to the protein and perturbing the parameters.

The crystal structure of penicillopepsin with no inhibitor bound shows two of the oxygens of the two key aspartyl residues positioned in close enough proximity to be hydrogen-bonded to each other[8]; this suggests that the two aspartates share a proton in the absence of bound inhibitor. However, we have performed the FEP calculation both with two anionic aspartates, (henceforth referred to as Models 1 and 2, which both have anionic aspartates, but differ in the model used for the bound inhibitor peptide, as described later in this paper), and with Asp 213 protonated (referred to as Model 3), to compare the free energies and structures calculated with the two models. The assumption that both aspartates are in the anionic form has also been made by some, who have modeled statine-based inhibitors in the active sites of aspartyl proteases[4]; however, it should be noted that other authors have assumed that the aspartates retain the proton

even when a statine-based inhibitor is bound[12]. We hoped, at the outset of these calculations, that the results might suggest whether or not one aspartate is protonated in the presence of a statine-based inhibitor.

The free energy calculations were carried out using the "windowing" method described previously[18,19]. The two states between which we wish to calculate a free energy are defined as the λ=1 and λ=0 states; in the case of the present calculation, these are the difluorostatone- and difluorostatine-containing inhibitor, respectively. Each of these states has appropriate molecular mechanical parameters, such as atomic partial charges and Van der Waals radii, assigned to it. The free energy calculation involves performing molecular dynamics on the system and changing λ in a series of steps of size $\delta\lambda$, from one extreme value to the other.

At each intermediate value of λ, or "window", λ_i, the dynamics trajectory evolves under the influence of a Hamiltonian that contains a λ-dependent combination of the molecular mechanical parameters of the two extreme states; for a given parameter P, having value P^A and P^B corresponding to the two extreme states A and B:

$$P_{\lambda_i} = \lambda P^A + \left[1 - \lambda\right] P^B$$

The free energy for each intermediate window is calculated using the equation:

$$\Delta G = -RT \ln \langle exp\left[-(H_{\lambda_i + \delta\lambda} - H_{\lambda_i})/RT\right] \rangle_{\lambda_i}$$

where $\langle\rangle_{\lambda_i}$ indicates an equilibrium average at $\lambda = \lambda_i$. At each window, both a "forward" ΔG and a "backward" ΔG can be calculated, corresponding to a negative or

positive $\delta\lambda$ in the equation given above. A comparison of the sums of these forward and backward calculations gives a measure of the uncertainty inherent in the calculation.

ΔG_{sol} was calculated twice, once by perturbing from the λ=1 state (difluorostatone) to the λ=0 state (difluorostatine); and once by perturbing from the λ=0 state back to the λ=1 state, starting with the coordinates from the previous perturbation. It was thus possibly to have an additional check on the uncertainty involved in the calculation by using the hysteresis between these two calculations as a standard deviation. ΔG_{bind} was also calculated twice, using two separate trajectories, but performing both calculations in the same direction, going from λ=1 to λ=0. The second calculation involved starting from a structure that had been equilibrated an additional 5 ps. The uncertainty in $\Delta\Delta G$ was calculated by considering the uncertainties in ΔG_{sol} and ΔG_{bind} as standard deviations, and combining them with the standard method for combining standard deviations of two data sets; i.e.:

$$\sigma_{\Delta\Delta G} = (\sigma^2_{sol} + \sigma^2_{bind})^{1/2}$$

In our initial calculations, the only interactions calculated in the ensemble average of the exponential described above were those between the inhibitor peptide and the protein. In other words, the entire inhibitor peptide was defined as a "perturbed group" and the free energy calculated involved only interactions between this group and the protein and waters; no intra-group interactions were calculated, in order to avoid calculating small differences between the large energies due to intra-peptide interactions. However, as described more fully in the Results section of this paper, this led to problems in maintaining the inhibitor-protein hydrogen bonds, since, as mentioned above,

the free energy perturbation module of AMBER does not allow the use of 10-12 energy terms, but rather replaces the 10-12 parameters normally assigned between hydrogen-bonded atoms with the 6-12 parameters normally assigned between non-hydrogen-bonded atoms. We thus decided to perform a second set of calculations by using as a "perturbed group" only the following atoms of the difluorostatone residue: the α carbon and its associated hydrogen, the two hydroxyl groups and the carbon to which they are attached, and the two fluorines and the carbon to which they are attached. This was the smallest group of atoms including the mutated hydroxyl for which the total charge remained essentially constant when perturbing from difluorostatone (hydrated ketone) to difluorostatine. Only a small amount of residual charge (~0.03 of an electronic charge) had to be added to the α carbon in order to make the total charge of this "perturbed group" remain the same during the perturbation. Figures 2.2g and 2.2h show the smaller perturbed group and the charges in the difluorostatone and difluorostatine states.

The model in which we used the entire peptide as the perturbed group is referred to in this paper as Model 1; the model using the smaller perturbed group is referred to as Model 2.

As mentioned previously in this paper, we used three models of the protein with the inhibitor peptide bound: Model 3 used one protonated aspartate, while Models 1 and 2 both used anionic aspartates; however, they differed in that Model 1 used the entire bound peptide as the perturbed group, while Model 2 used the smaller perturbed group. Model 3 also used the smaller perturbed group.

Figure 2.2i shows the charges used for the protonated aspartate.

The molecular dynamics trajectory from which the free energy was calculated was run using a time step of one femtosecond. For the initial ΔG_{sol} and ΔG_{bind} calculations using Model 1, each window, as described above, consisted of a 400 fs equilibration period; and a 400 fs data collection period, in which the actual calculation of the ensemble average, used in determining ΔG, took place. A complete free energy calculation, perturbing from statone to statine in one direction only, consisted of twenty-one of these windows (for both the protein and the solution calculation), and thus involved 16 picoseconds of dynamics. In subsequent calculations, using Models 2 and 3, we desired to use longer simulations for a greater degree of convergence of the free energies. These subsequent FEP calculations, performed with the smaller "perturbed group", were expected to maintain a structure close to that of the crystal structure even over a longer simulation, so these calculations used windows consisting of 800 fs of equilibration and 800 fs of data collection. These calculations also used more windows than the initial calcalation: since it was found that there were larger changes in the free energy in the windows where λ was in the range 1.0 to 0.8 (i.e., close to the statone state), twice as many windows were used in this range of λ. That is , a $\delta\lambda$ of 0.025 was used in this range, and a $\delta\lambda$ of 0.05 for $\lambda < 0.8$, for a total of twenty-five windows and a total simulation time of 40 ps.

Both the protein-inhibitor complex and the solvated complex were equilibrated by molecular dynamics before the perturbation dynamics trajectory was begun; a time step of one femtosecond was used for this equilibration also. The solvated inhibitor was equilibrated for 10 ps using the STO-3G charges, in the case of the FEP calculation using these charges; for the FEP calculation involving the 6-31G* charges, the STO-3G-equilibrated structure was re-equilibrated for an additional 3 ps with the new

charges before the FEP calculation was carried out. Similarly, for the protein calculation, the protein-inhibitor complex was equilibrated for 16 ps with the STO-3G charges; for the 6-31G* calculation, an additional 3 ps of equilibration with the new charges was carried out.

To carry out the calculation on the protein-inhibitor complex, it was necessary to solvate the active site of the protein and its immediate vicinity. This was done by placing a spherical shell of water, of radius 18 Å, at the active site of the crystal structure. This shell was centered at the oxygen atom of the hydroxyl group being perturbed. The water coordinates were obtained from a snapshot of a Monte Carlo simulation of water. The TIP3P[35] water potential was used. This collection of water molecules was placed in the desired location, and any water molecule with its oxygen within 2.4 Å, or a hydrogen within 1.4 Å, of a protein or inhibitor atom, was removed from the structure.

The waters in the resultant solvated protein-inhibitor complex (and the crystal waters, which were left in place) were then subjected to energy minimization, while keeping the structure of the protein and inhibitor fixed. This was followed by minimization of all residues (including waters) having an atom within 15 Å of the key hydroxyl oxygen, for Model 1, or within 5 Å of any atom of the inhibitor for Models 2 and 3 (the smaller group of mobile residues was used in subsequent calculations to prevent large deviations from the crystal structure during the simulation). The other residues were not allowed to move. The entire structure was then equilibrated with MD as described above; during the equilibration and the consequent FEP calculation, only the residues within 15 Å of the perturbed hydroxyl (for Model 1), or within 5 Å of any atom of the inhibitor (for Models 2 and 3), were allowed to move. In the FEP calculation itself, only the residues within 15 Å of the hydroxyl (Model 1), or within 5 Å of any inhibitor

atom (Models 2 and 3), were allowed to move.

The calculation involving the solvated inhibitor was set up in a similar fashion, starting with a Monte Carlo snapshot of TIP3P waters. The inhibitor peptide structure from the x-ray coordinates was surrounded by a shell of waters such that the waters extended out to 18 Å away from any atom of the inhibitor. The waters alone were subjected to energy minimization first, followed by minimization of the inhibitor and all waters having an atom within 15 Å of any inhibitor atom (for both inhibitor models). MD equilibration of the inhibitor and all waters within 15 Å of any inhibitor atom (for both inhibitor models) was then carried out; this left an outside shell of "frozen" waters that were not allowed to move. In the FEP calculation, also, only the waters within 15 Å of the inhibitor were allowed to move (for both inhibitor models).

As mentioned above, calculations involving Model 2 involved calculating intra-inhibitor interactions. In this case, to calculate a true $\Delta\Delta G_{sol}$), it was necessary to perform the FEP calculation on the isolated (gas phase) inhibitor also, to obtain ΔG_{gas}. In this case, $\Delta\Delta G_{sol} = \Delta G_{sol} - \Delta G_{gas}$ (See Figure 2.5). In the inhibitor solvation FEP calculation involving Model 1, $\Delta\Delta G_{sol}$ is exactly equal to ΔG_{sol}, since not intra-inhibitor interactions are calculated.

As mentioned above, the inclusion of intra-inhibitor interactions in the FEP calculation introduces the problem of calculating small differences between large intra-inhibitor interaction energies. It would be difficult to calculate the differences due to these differences accurately if the same regions of conformational space are not sampled equally in both the "forward" and "backward" calculations; with the sampling limitations inherent in free energy calculations using the currently available amount of computer power, a large hysteresis in ΔG_{sol} is to be expected, and was found, if intra-

inhibitor interaction energies are calculated for a freely moving inhibitor. For this reason, constraints were applied for the gas phase and solution calculations involving calculation of intra-inhibitor interactions. That is, all ΔG_{sol} calculations involving Models 2 and 3 were performed with constraints on all of the inhibitor dihedral angles in which all four atoms were non-hydrogen atoms.

Both the solvation calculation and the protein calculation were carried out using a cutoff distance of 8 Å for the non-bonded interactions.

RESULTS

A. FREE ENERGIES

Table 2.5 gives the differences in binding (ΔG_{bind}) and solvation (ΔG_{sol}) free energies of the two statine derivatives, for each of the calculations performed. The initial calculation we performed, using Model 1, gave an average ΔG_{bind} of 4.3 kcal/mole after 19 ps of equilibration. After an additional 5 picoseconds of MD equilibration using the same model, a ΔG_{bind} of 5.2 kcal/mole was obtained. The overall average of the two different calculations was thus 4.7 kcal/mole, with an uncertainty of 0.5 kcal/mole. The ΔG_{sol} values obtained using Model 1 are 2.8 for the statone to statine calculation, and -1.9 for the the statine to statone calculation. This gives an average ΔG_{sol} of 2.4 kcal/mole, with an uncertainty of 0.5. A $\Delta\Delta G$ of 2.3 kcal/mole, with an uncertainty of 0.7, was obtained

as the difference between the average values of ΔG_{bind} and ΔG_{sol}.

The calculation using Model 2 gave a ΔG_{bind} of 2.4 kcal/mole in the statone to statine direction, and a ΔG_{bind} of -1.6 kcal/mole in the statine to statone direction, resulting in an average ΔG_{bind} of 2.0±0.4 kcal/mole. The calculation on the solvated inhibitor using Model 2 gave a ΔG_{sol} of -7.3 kcal/mole in the statone to statine direction, and 6.7 kcal/mole in the statine to statone direction. This gives an average ΔG_{sol} of -7.0±0.3 kcal/mole. The resulting $\Delta\Delta G$ is thus 9.0±0.5 kcal/mole.

A 40 ps gas phase FEP calculation on the inhibitor, using Model 2, gave a ΔG_{gas} of -11.6 kcal/mole in the statone to statine direction and 11.2 kcal/mole in the statine to statone direction, giving an average ΔG_{gas} of -11.4±0.3 kcal/mole. The true $\Delta\Delta G_{sol}$ is thus 4.4±0.4 kcal/mole.

A third ΔG_{bind} calculation, which used Model 3, gave a ΔG_{bind} of 5.6 kcal/mole in the statone to statine direction and -5.8 kcal/mole in the statine to statone direction, giving an average ΔG_{bind} of 5.7±0.1 kcal/mole. Combined with the ΔG_{sol} described above, this gives a $\Delta\Delta G$ of 12.7±0.3 kcal/mole.

Gelb and coworkers report the experimental difference in $\Delta\Delta G$ as 1.5 kcal/mole.

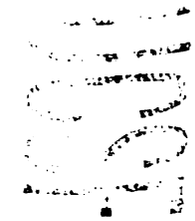
B. RMS DEVIATIONS

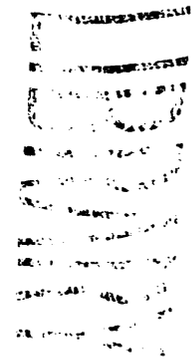
The rms deviations of the calculated structures from the crystal structure of difluorostatone bound to penicillopepsin are given in Table 2.6. In all cases, Models 2 and 3 resulted in structures more closely resembling the crystal structure than did Model 3.

C. HYDROGEN BOND ANALYSIS

Figure 2.6 is a schematic showing the key hydrogen bonds between the inhibitor and the protein; this can be compared with the key hydrogen bonds found in crystal structure of hydrogen bonds found in crystal structures of statine-based inhibitors bound to other aspartyl proteases (see, for example, Reference 4).

Table 2.7 lists, for the difluorostatone-penicillopepsin X-ray structure and for the structures resulting from the calculations using Model 1, whether or not a hydrogen bond exists (+) or does not exist (-) between the atoms involved in these key hydrogen bonds, using a 3.6 Angstrom cutoff criterion for the inter-heavy atom distance. It is obvious that a number of the inhibitor-protein hydrogen bonds present in the X-ray structure have been broken during equilibration and perturbation calculations; many of these are replaced by water-protein hydrogen bonds.





Of particular interest are the hydrogen bonds in the region of the two active site aspartates (Asp 33 and 213). In particular, three hydrogen bonds from the non-perturbed hydroxyl group to the aspartate oxygens are broken during equilibration and do not reform during the perturbation calculations. Of the two hydrogen bonds from the perturbed hydroxyl group (which becomes a hydrogen in the statine form of the inhibitor) to Asp 33, one breaks during equilibration and does not re-form after both directions of the perturbation (from statone to statine and back again) have been traversed. The distance between the two aspartates found in the crystal structure (2.98 Angstroms) is suggestive of a hydrogen bond, and the relevant O-O distance has been included in the tables of hydrogen bonds, even though it is not certain that one of these oxygens carries a proton. In all of the calculated structures shown in Table 2.7, this O-O distance is too large to fit the criterion for a hydrogen bond (the distance, not shown in Table 2.7, are ~3.8 Angstroms).

Figure 2.7 shows stereoviews of the crystal structure of penicillopepsin with statone bound, and of Model 1 after MD equilibration and each direction of perturbation.

Table 2.8 shows the presence or absence of hydrogen bonds for the protein-inhibitor FEP calculation performed with Model 2. Waters have been omitted for clarity, since they no longer take the place of protein-inhibitor hydrogen bonds. It is immediately apparent that the hydrogen bonds resulting from the FEP simulations bear a much closer relationship to the crystal structure than those resulting from the simulations using the entire peptide as the perturbed group. In general, more protein-inhibitor hydrogen bonds are conserved, rather than being displaced by waters.

The hydrogen bonding structure around the two active site aspartates remains similar to that of the crystal structure throughout the equilibration and perturbation calculations. In particular, the hydrogen bonds between the non-perturbed hydroxyl of the statone residue and Asp 213, and the bonds between the perturbed hydroxyl of the statone and Asp 33, are all present after equilibration and after perturbation in both directions has been completed. However, the O-O distance between the two closest oxygens of Asp 33 and 213 is still greater than than that of the crystal structure (the distances, not shown in Table 2.8, are 3.7-3.8 Angstroms).

Figure 2.8 shows stereoviews of Model 2 after equilibration and after each direction of perturbation.

Table 2.9 shows the same hydrogen bond patterns resulting from the calculation using Model 3. The structures are again closer to the crystal structure than those resulting from the calculation using Model 1, particularly in the region around the two active site aspartates. However, one of the hydrogen bonds between the non-perturbed hydroxyl and Asp 33 is not present after both directions of perturbation have been

completed. The distance between atoms OD2 of Asp 33 and OD1 of Asp 213 is suggestive of a hydrogen bond, as in the crystal structure (the actual distances in the calculated structure are in 2.7-2.9 Angstroms, versus 2.98 in the crystal structure).

Figure 2.9 shows stereoviews of Model 3 after equilibration and after each direction of perturbation.

It is difficult to say which of the two preceding structures (described in Tables 2.8 and 2.9) is "better", since the latter is lacking one statone-aspartate hydrogen bond upon completion of both directions of perturbation, and the other has an O-O distance between Asp 33 and Asp 213 that is almost an Angstrom larger than that of the crystal structure. It is obvious, however, that these structures calculated using Models 2 and 3 more closely resemble the X-ray structure than do the structures calculated using the entire inhibitor as the perturbed group. It is thus easier to have faith in the veracity of the free energies obtained from the calculations using the smaller perturbed group.

DISCUSSION AND CONCLUSIONS

The $\Delta\Delta G$ for the mutation of difluorostatone into difluorostatone is qualitatively correct for each of the three calculations. In each case, an answer in qualitative agreement with the experimental results of Gelb et al.; i.e., the statone form of the inhibitor is the tighter-binding form.

The ΔG_{bind} calculations using Model 2 and 3 lead to the best structure; i.e., that which is closest to the difluorostatone-penicillopepsin crystal structure. The difference between Models 2 and 3 is only in the Asp33-Asp213 distances, with Model 3 beting more realistic.

The ΔG_{bind} value is the most believable, if the hydrogen-bonding structure is to be a criterion; it is difficult to have faith in the free energies obtained if the structure diverges greatly from the crystal structure. However, this ΔG_{bind} results in the worst quantitative $\Delta\Delta G$ in comparison to experiment.

The ΔG_{bind} calculation using Model 2 results in a structure almost as good as the calculation in which Asp 213 is protonated, but differing in that the two closest carboxyl oxygens move a considerable distance apart during molecular dynamics equilibration, and remain apart during the perturbation calculations. Here again, the $\Delta\Delta G$ resulting from the ΔG_{bind} calculation was of the same sign as the experimental $\Delta\Delta G$ but considerably larger in magnitude.

The only $\Delta\Delta G$ value in reasonable agreement with experiment is that obtained from the calculations using Model 1. The structure of the inhibitor-penicillopepsin complex resulting from this calculation is very poor, however; the ΔG_{bind} is therefore suspect.

It is difficult to draw a conclusion about the protonation state of the two active site aspartates from the results described above, since the $\Delta\Delta G$'s closest to experiment were the results of calculations involving two anionic aspartates, while the structure most closely resembling the crystal structure resulted from calculations in which one active site aspartate was protonated. The structural results tend to be more convincing, however, since the $\Delta\Delta G$ depends not only on ΔG_{bind}, but also on ΔG_{sol}. The ΔG_{sol}'s

described in this paper may have errors due to constraints being applied, or to inadequate sampling of conformational space (in the shorter calculation, where no constraints were applied).

The absolute values of the ΔG_{bind}'s is surprisingly small. The free energy of interaction between a hydroxyl group and an ionized carboxyl group was expected to be 10 or more kcal/mole or larger; recall that the energy of interaction of acetate and methanol, after energy minimization, was ~13 kcal/mole (Table 2.4b).

The ΔG_{sol}'s of 2.4 and 4.4 kcal/mole were also smaller in magnitude than expected. The difference in free energy of hydration due to the the disappearance of a hydroxyl group was expected to be about 7 kcal/mole, in agreement with other calculations and experiments[19,37]. These results suggest further calculations on the free energy of hydration of difluoroalcohols, particularly difluoro-*gem*-diols. An interesting question is whether the relative free energy of removing a hydroxyl group from a *gem*-diol is significantly less than removing one from a mono-alcohol, and whether or not this depends upon the alcohol in question being fluorinated. We intend to carry out calculations of the relative free energies of hydration of $CHF_2CH(OH)_2$, $CH_3CH(OH)_2$, CHF_2CH_2OH, CH_3CH_2OH, CH_3CH_3, and CHF_2CH_3. Calculations on small difluorohydrocarbon alcohols will demand less in the way of computational resources, and will thus allow long calculations that better sample the conformational space of the solute than the relatively short calculations we have performed upon the entire tripeptide inhibitor. These calculations will reveal whether the surprisingly small ΔG_{sol}'s are merely an artifact of inadequate sampling, due to the dihedral constraints or the short sampling time. It is conceivable that under these conditions the ΔG_{sol} will be greater in magnitude, suggesting that longer calculations involving the entire inhibitor would also

give larger ΔG_{sol}'s, and thus a $\Delta\Delta G$ more in line with the experimental value.

Another issue to be investigated further is the relative merit of performing calculations of hydration free energy in a water "drop", as was done in the present work, as opposed to using periodic boundary conditions. The "drop" approach was used in order to make the treatment of solvent in the calculation of ΔG_{sol} as similar as possible to the its treatment in the ΔG_{bind} calculation, since FEP calculations rely to some extent upon the cancellation of certain systematic errors[38], and we wanted to make these errors similar in the two calculations. Calculations on small difluorohydrocarbon alcohols will be less demanding in terms of computational resources, and this will allow each perturbation calculation to be performed both with a "drop" of water and with periodic boundary conditions.

The difficulties involved in performing FEP calculations involving a 10-12 hydrogen bond potential will be absent in AMBER Version 4.0, the next version to be released. In Version 4.0, it will be possible to perturb between a 10-12 and an 6-12 potential. This will prevent the sort of structural problems seen in our calculations involving Model 1. It may also give more realistic free energies for calculations involving hydrogen bonds, since hydrogen bonds within the "perturbed group" will be represented by the 10-12 potentials derived in the 1984 Weiner et al. force *field*[31], which has been thoroughly tested.

In summary, we can say that the utility of the FEP method for drug design is apparent, at least for the cases , such as this one, in which there are no large structural changes between two enzyme-protein inhibitor complexes. Although quantitative information on relative free energies of binding are not always obtained, as demonstrated in the results of this paper, the qualitative results can provide a guide toward

modifications that may produce a tighter-binding inhibitor. The use of constraints on the peptide inhibitor, however, is unsatisfying, since the ΔG_{sol} may be critically dependent upon sampling configurations of the inhibitor other than that found in the crystal structure. These calculations thus point out that the biggest pitfall in the predictive use of FEP calculations may be the inability to adequately sample the conformational space of the inhibitor in aqueous solution, rather than inadequate sampling of the inhibitor-protein complex. In this calculation, there is no gross difference in conformation between the two protein-inhibitor complexes of interest; but there may be significant differences between the conformations sampled by the two inhibitors in adeqaute solution. This issue will be addressed by the fluoroalcohol calculations discussed above.

REFERENCES FOR CHAPTER TWO

(1) Zinchenko, A.A.; Rumsh, L.D.; Antonov, V.K. *Bioorg. Chem. (USSR)* 1976, **2**, 803.

(2) Suguna, K.; Padlan, E.A.; Smith, C.W.; Carlson, W.D.; Davies, D.R. *Proc. Nat. Acad. Sci. USA* 1987, **84**, 7009.

(3) Suguna, K.; Bott, R.R.; Padlan, E.A.; Subramanian, E.; Sheriff, S.; Cohen, G.H.; Davies, D.R. *J. Mol. Biol.* 1987, **196**, 877.

(4) Foundling, S.I.; Cooper, J.; Watson, F.E.; Cleasby, A.; Pearl, L.H.; Sibanda, B.L.; Hemmings, A.; Wood, S.P.; Blundell, T.L.; Valler, M.J.; Norey, C.G.; Kay, J.; Boger, J.; Dunn, B.M.; Leckie, B.J.; Jones, D.M.; Atrash, B.; Hallett, A.; Szelke, M. *Nature* 1987, **327**, 349.

(5) Carlson, W.; Karplus, M.; Haber, E. *Hypertension* 1985, **7**, 13.

(6) Sodek, J.; Hofmann, T. *Methods Enzymol.* 1971, **19**, 372.

(7) Hofmann, T.; Shaw, R. *Biochim. Biophy. Acta* 1964, **92**, 543.

(8) James, M.N.G.; Sielecki, A.R. *J. Mol. Biol.* 1983, **163**, 299.

(9) Blundell, T.L.; Sewell, B.T.; McLachlan, A.D. *Biochim. Biophys. Acta* 1979, **580**, 24.

(10) Miller, M.; Jaskolski, M.; Rao, J.K.M.; Leis, J.; Wlodawer, A.; *Nature* 1989, **337**, 576.

(11) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B.K.; Baldwin, E.; Weber, I.T.; Selk, L.M.; Clawson, L.; Schneider, L.; Kent, S.B.H.; *Science*

1989, **245**, 616.

(12) Navia, M.A., Fitzgerald, P.M.D.; McKeever, B.M.; Leu, C-T.; Heimbach, J.C.; Herber, W.K.; Sigal, I.S.; Darke, P.L.; Springer, J.P. *Nature* 1989, **337**, 615.

(13) Enzyme Structure and Mechanism, Alan Fersht. W.H. Freeman and Company, New York, 1985.

(14) Rich, D.H. *J. Med. Chem.* 1985, **28**, 263.

(15) Pearl, L.H.; Taylor, W.R. *Nature* 1987, **329**, 351.

(16) Weber, I.; Muller, M.; Jaskolski, M.; Leis, J.; Skolka, A.; Wlodawer, A.; *Science* 1989, **243**, 928.

(17) Yoshinaka, Y.; Luftig, R.B. *Cell* 1977, **12**, 709.

(18) Van der Helm, K. *Proc. Natn. Acad. Sci. USA* 1977, **74**, 911.

(19) Katoh, I.; Yasunaga, T.; Ikawa, Y.; Yoshinaka, Y. *Nature* 1987, **329**, 654.

(20) Bash, P.A.; Singh, U.C.; Langridge, R.; Kollman, P.A. *Science* 1987, **236**, 564.

(21) Singh, U.C.; Brown, F.K.; Bash, P.A.; Kollman, P.A. *J. Am. Chem. Soc.* 1987, *109*, 1607.

(22) Rao, S.N.; Singh, U.C.; Bash, P.A.; Kollman, P.A. *Nature* 1987, **328**, 551.

(23) Seibel, G.; Kollman, P.A. Work in progress.

(24) Bash, P.A.; Singh, U.C.; Brown, F.K.; Langridge, R.; Kollman, P.A. *Science* 1987, **235**, 574.

(25) Merz, K.; Kollman, P.A.; *J. Am. Chem. Soc.* 1989, **111**, 5649.

(26) Marshall, G.R. *Fed. Proc., Fed. Am. Soc. Exp. Biol* 1976, **35, 2494.**

(27) Marciniszyn, M.P.; Hartsuck, J.A.; Tang, J.J.N. *J. Biol. Chem.* 1976, **251**, 7088.

(28) Gelb, M.H.; Svaren, J.P.; Abeles, R.H. *Biochemistry* 1985, **24, 1813.**

(29) Fearon, K.; Spaltenstein, A.; Hopkins, P.B.; Gelb, M.H. *J. Med. Chem.* 1987, **30**, 1617.

(30) Sielecki, A.R.; James, M.N.G. Personal communication.

(31) Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghia, C.; Alagona, G.; Profeta, S.,Jr.; Weiner, P.; *J. Am. Chem. Soc.* 1984, *106*, 765.

(32) Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A.; *J. Comp. Chem.* 1986, **7, 232.**

(33) Singh, U.C.; Kollman, P.A.; Gaussian 80 UCSF, *QCPE Bull.*, 1982, **2**, 17.

(34) Singh, U.C.; Kollman, P.A. *J. Comp. Chem.* 1984, **5**, 129.

(35) Allinger, N.L.; *J. Am. Chem. Soc.* 1977, **99**, 8127.

(36) Brunck, T.K.; Weinhold, F.;*J. Am. Chem. Soc.*, 1979, **101**, 1700.

(37) Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L.; *J. Chem. Phys.* 1983, **79**, 926.

(38) Gelb, M.H.; personal communication.

(39) Jorgensen, W.L.; Ravimohan, C.; *J. Chem. Phys.* 1985, **83**, 3050.

(40) Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications". W. van Gunsteren and P.K. Weiner, eds. Escom Science Publishers, Netherlands, 1989.

Table 2.1

New Bond Stretching and Bending Parameter for Bonds Involving Fluorine and hydroxyl oxygen

Bond	K_R (in kcal/mol Å^2)	R_o (in Å)
CT - F	367	1.38
Angle	**K_θ (in kcal/mol rad ^{2})**	**θ_o (in degrees)**
F-CT-HC	35.0	109.5
F-CT-CT	46.8	109.5
F-CT-C	46.8	109.2
F-CT-F	77.0	109.1
OH-CT-OH	80.0	109.5

Table 2.2

Torsional Parameters for Dihedrals Involving Fluorine

Dihedral Type	n (fold of dihedral)	$V_N / 2$ (kcal/mole)	Phase γ
F-CT-C-O	3	0.067	180
F-CT-CT-OH	3	0.144	0
F-CT-CT-OH	2	0.500	0
F-CT-CT-HC	3	0.144	0
F-CT-CT-HC	2	0.500	0

Table 2.3

6-12 Parameters Used for Fluorine and Hydroxyl Hydrogen		
	R^{*a} (in Å)	ε^{b} (in kcal/mol Å^2)
F	1.55	0.080
HO	0.58	0.020

(a) van der Waals minimum distance

(b) van der Waals well depth

Table 2.4

Acetate-Methanol Interaction Energies[a]

Parameters Used	Interaction Energy[b] (kcal/mole)	Distance (in Å)
Standard Amber 10-12	-13.69	2.690
6-12	-12.92	2.692

[a] Interaction energy = energy of acetate-methol complex - sum of energies of isolated acetate and methanol.

[b] Energies of isolated molecules: acetate: -15.30 kcal/mole; methanol: 0.57 kcal/mole.

Table 2.5(a)

Gas Phase Free Energies ($\Delta G_{gas}^{a,b}$)	
Direction of Perturbation	$\Delta G_{gas}^{a,b}$
Statone to statine	-11.6
Statine to statone	11.2
Average	-11.4 ±0.3[g]

Table 2.5(b)

Solvation Free Energies ($\Delta G_{sol}^{a,c}$)				
Model	$\Delta G_{sol}^{a,b}$ Statone to statine	$\Delta G_{sol}^{a,c}$ Statine to statone	Average	$\Delta\Delta G_{sol}^{a,d}$
Model 1	2.8	-1.9	2.4 ±0.5[g]	2.4 ±0.5[g]
Model 2	-7.3	6.7	-7.0 ±0.3[g]	4.4 ±0.4[g]

Table 2.5(c)

Binding Free Energies ($\Delta G_{bind}^{a,e}$)			
Model	$\Delta G_{bind}^{a,g}$ Statone to statine	$\Delta G_{bind}^{a,g}$ Statine to statone	Average
Model 1	4.3	-5.2	4.7 ±0.5[g]
Model 2	2.4	-1.6	2.0 ±0.4[g]
Model 3	5.6	-5.8	5.7 ±0.1[g]

Table 2.5(d)

$\Delta\Delta G's^{a,J}$	
Model	$\Delta\Delta G^{a,J}$
Model 1	2.3 ±0.7[g]
Model 2	9.0 ±0.5[g]
Model 3	12.7 ±0.3[g]

Table 2.5

(a) In kcal/mole.

(b) ΔG_{gas} is the difference in free energy between the statine and statone states in the gas phase.

(c) ΔG_{sol} is the difference in free energy between the statine and statone state in aqueous solution.

(d) $\Delta\Delta G_{sol}$ is equal to ΔG_{sol} minus ΔG_{gas}.

(e) ΔG_{bind} is the difference in free energy between the statine and statone states when bound to the active site of penicillopepsin.

(f) $\Delta\Delta G$ is equal to ΔG_{bind} minus ΔG_{sol}.

(g) See the Methods and Results sections for an explanation of how these standard deviations were calculated.

Table 2.6

RMS Deviations (in Angstroms)
between X-ray structure and calculated structures

Structure	Model 1	Model 2	Model 3
After MD Equilibration	0.89	0.49	0.62
After statone to statine perturbation	0.97	0.60	0.59
After both directions of perturbation	-	0.76	0.64

Table 2.6

(a) Only non-hydrogen atoms were considered in these calculations.

(b) The MD-equilibrated strcture was obtained from 19 picoseconds of molecular dynamics equilibration of the energy-minimiized statone X-ray structure. The "Perturbed, λ=0" structure is the result of a complete perturbation from the statone to the statine state.

Table 2.7

Key Hydrogen Bonds, Model 1[a,b]

H-Bond		Statone X-Ray	Md-Equilibrated Structure	Perturbed, λ=0	Perturbed, λ=0, After Additional 5ps Equilibration
0 of 324 IVA	Water[c]	+	+	+	+, +
N of 325 Val	OG1 of THR 217	+	-	-	-
	Water[c]		-	+	-
O of 325 Val	N of Thr 217	+	+	-	-
	Water[c]	+	+	+	+
N of 326 Val	OD1 of ASP 77	+	-	-	-
	OD2 of ASP 77	-	-	-	-
O of 326 Val	N of ASP 77	+	-	+	+
	N of GLY 76	+	+	-	+
OD1 of ASP 77	Waters[c]	+	+	+	-
			+	+	+
			+		
			-		
OD2 of ASP 77	Waters[c]	+	+	+	+
		+	+	+	+
			-	+	+
					+
					+
N of Statone(ine)	O of GLY 215	+	-	-	-
	OG1 of THR 216	+	+	+	+
O of GLY 215	Waters[c]	+	+	+	-
					+
O2 of Statone(ine)	N of GLY 76	+	+	+	+
	Waters		+		+
N2 of Statone(ine)	O of GLY 35	+	-	+	-
	OG of SER 36	-	+	+	-
	Waters				+,
					+,
					+,
					+
Statone(ine) O	OD1 of AP 33	+	-	-	-
	OD2 of ASP 33	+	-	-	-
	OD1 of ASP 213	+	-	-	-
	OD2 of ASP 213	+	+	+	+
	OG1 of THR 216	-	+	+	-

Table 7 Key Hydrogen Bonds, Model 1[a,b] (continued)

H-Bond		Statone X-Ray	Md-Equilibrated Structure	Perturbed, λ=0	Perturbed, λ=0, After Additional 5ps Equilibration
Statone(ine) OXT (H in statine)	OD1 of ASP 33	+	-	-	-
	OD2 of ASP 33	+	+	+	-
	OD1 of ASP 213	-	-	-	-
	OD2 of ASP 213	-	-	+	-
	OG of SER 36	-	+	+	+
OD1 of ASP 33	OG of SER 36	+	+	+	-
	Water			+	+
OD2 of ASP 33	N of GLY 35	+	+	+	+
OD2 of ASP 213	OG1 of THR 216	+	+	+	+
OD2 of ASP 33	OD1 of ASP 213	+	-	-	-

(a) The MD-equilibrated structure was obtained from 19 picoseconds of molecular dynamics equilibration of the energy-minimized statone X-ray structure. The "Perturbed, λ=0" structure is the result of a complete perturbation from the statone to the statine state. The structure labeled "λ=0, after additional 5 ps equilibrattion" is the results of a perturbation from the statone to the statine state after the MD-equilibrated structure was equilibrated for an additional 5 ps.

(b) In this table, atoms "O" and "N" are peptide carboxyl oxygens and peptide amide nitrogens, except for water residues (where "O" is the water oxygen) and the statone residue, where atom "O" is the non-perturbed hydroxyl oxygen. Atom OXT is the oxygen atom of the perturbed hydroxyl group in the statone residue. Atoms N2 and O2 are the nitrogen and oxygen of the peptide bond of the N-methyl group in the statone residue. Atoms "OD1" and "OD2" are the carboxyl oxygens of aspartate residues. Atoms "OG" and "OG1" are the hydroxyl oxygen atoms of serine and threonine, respectively. See Figure 5 for a schematic of some of these bonds.

Table 2.7

(a) The MD-equilibrated structure was obtained from 19 picoseconds of molecular dynamics equilibration of the energy-minimized statone S-ray structure. The "Perturbed, λ=0" strucure is the result of a complete perturbation from the statone to the statine state.

(b) In this table, atoms "O" and "N" are peptide carboxyl oxygens and peptide amide nitrogens, except for water residues (where "O" is the water oxygen) and the statone residue, where atom "O" is the non-perturbed hydroxyl oxygen. Atom OXT is the oxygen atom of the perturbed hydroxyl group in the statone residue. Atoms N2 and O2 are the nitrogen and oxygen of the peptide bond of the N-methyl group in the statone residue. Atoms "OD1" and "OD2" are the carboxyl oxygens of aspartate residues. Atoms "OG" and "OG1" are the hydroxyl oxygen atoms of serine and threonine, respectively. See Figure 5 for a schematic of some of these bonds.

Table 2.8

Key Hydrogen Bonds, Model 2[a,b]

H-Bond		Statone X-Ray	Md-Equilibrated Structure	Perturbed, λ=1	Perturbed, λ=0
N of Val 325	OG1 of THR 217	+	+	+	-
N of Val 326	OD1 of ASP 77	+	+	-	-
	OD2 of ASP 77	-	-	+	-
O of Val 326	N of ASP 77	+	+	+	+
	N of GLY 76	+	+	+	+
N of Statone(ine)	O of GLY 215	+	-	-	+
	OG1 of THR 216	+	+	-	-
O2 of Statone(ine)	N of GLY 76	+	+	+	+
N2 of Statone(ine)	OG of SER 36	-	-	-	+
Statone(ine) O	OD1 of ASP 33	+	-	-	-
	OD2 of ASP 33	+	-	-	-
	OD1 of ASP 213	+	+	+	+
	OD2 of ASP 213	+	+	-	+
	OG1 of THR 216	-	+	+	-
Statone(ine) OXT (H in statine)	OD1 of ASP 33	+	+	+	+
	OD2 of ASP 33	+	+	+	+
	OD1 of ASP 213	-	-	-	-
	OD2 of ASP 213	-	-	-	-
	OG of SER 36	-	-	-	+
OD1 of ASP 33	OG of SER 36	+	+	+	+
OD2 of ASP 33	N of GLY 35	+	+	+	+
OD2 of ASP 213	OG1 of THR 216	+	+	-	+
OD2 of ASP 33	OD1 of ASP 213	+	-	-	-

(a) The MD-equilibrated structure was obtained from 19 picoseconds of molecular dynamics equilibration of the energy-minimized statone X-ray structure. The "Perturbed, λ=0" structure is the result of a complete perturbation from the statone to the statine state. The structure labeled "λ=1" is the result of a perturbation from the statone to the statine state and back to the statone state.

(b) In this table, atoms "O" and "N" are peptide carboxyl oxygens and peptide amide nitrogens, except for water residues (where "O" is the water oxygen) and the statone residue, where atom "O" is the non-perturbed hydroxyl oxygen. Atom OXT is the oxygen atom of the perturbed hydroxyl group in the statone residue. Atoms N2 and O2 are the nitrogen and oxygen of the peptide bond of the N-methyl group in the statone residue. Atoms "OD1" and "OD2" are the carboxyl oxygens of aspartate residues. Atoms "OG" and "OG1" are the hydroxyl oxygen atoms of serine and threonine, respectively. See Figure 6 for a schematic of some of these bonds.

Table 2.8

See the caption to Table 7 for explanation of atom names.

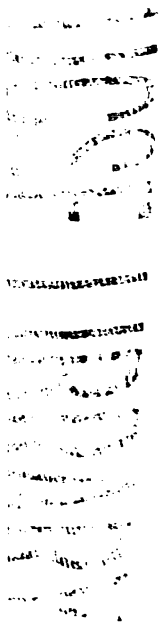


Table 2.9

Key Hydrogen Bonds, Model 3[a,b]

H-Bond		Statone X-Ray	Md-Equilibrated Structure	Perturbed, λ=1	Perturbed, λ=0
N of Val 325	OG1 of THR 217	+	+	+	+
N of Val 326	OD1 of ASP 77	+	+	+	+
	OD2 of ASP 77	-	-	-	-
O of Val 326	N of ASP 77	+	+	+	+
	N of GLY 76	+	+	+	+
N of Statone(ine)	O of GLY 215	+	+	+	+
	OG1 of THR 216	+	+	-	-
O2 of Statone(ine)	N of GLY 76	+	+	+	+
N2 of Statone(ine)	OG of SER 36	-	-	-	-
Statone(ine) O	OD1 of ASP 33	+	-	+	-
	OD2 of ASP 33	+	+	+	+
	OD1 of ASP 213	+	+	+	+
	OD2 of ASP 213	+	+	+	+
	OG1 of THR 216	-	+	-	+
Statone(ine) OXT (H in statine)	OD1 of ASP 33	+	+	+	+
	OD2 of ASP 33	+	+	-	+
	OD1 of ASP 213	-	-	-	-
	OD2 of ASP 213	-	-	-	-
	OG of SER 36	-	-	-	-
OD1 of ASP 33	OG of SER 36	+	+	+	-
OD2 of ASP 33	N of GLY 35	+	+	+	+
OD2 of ASP 213	OG1 of THR 216	+	+	+	+
OD2 of ASP 33	OD1 of ASP 213	+	+	+	+

(a) The MD-equilibrated structure was obtained from 19 picoseconds of molecular dynamics equilibration of the energy-minimized statone X-ray structure. The "Perturbed, λ=0" structure is the result of a complete perturbation from the statone to the statine state. The structure labeled "λ=1" is the result of a perturbation from the statone to the statine state and back to the statone state.

(b) In this table, atoms "O" and "N" are peptide carboxyl oxygens and peptide amide nitrogens, except for water residues (where "O" is the water oxygen) and the statone residue, where atom "O" is the non-perturbed hydroxyl oxygen. Atom OXT is the oxygen atom of the perturbed hydroxyl group in the statone residue. Atoms N2 and O2 are the nitrogen and oxygen of the peptide bond of the N-methyl group in the statone residue. Atoms "OD1" and "OD2" are the carboxyl oxygens of aspartate residues. Atoms "OG" and "OG1" are the hydroxyl oxygen atoms of serine and threonine, respectively. See Figure 6 for a schematic of some of these bonds.

Table 2.9

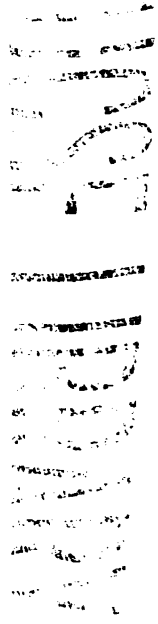
See the caption to Table 7 for explanation of atom names.

FIGURE CAPTIONS FOR CHAPTER TWO

Figure 2.1

Structures of the inhibitor peptides containing: (a)difluorostatine; (b)difluorostatone; (c)the hydrated form of difluorostatone.

Figure 2.2



(a-c)Structures used in quantum-mechanical derivation of point charges: (a)6-31G* charges of difluorostatone and difluorostatine; (b)STO-3G charges of difluorostatone and difluorostatine; (c)STO-3G charges of isovaleryl residue.

(d-f)Quantum-mechanically calculated point charges used for: (d)difluorostatine (e)hydrated difluorostatone; (f)isovaleryl group. The atoms marked with an asterisk have charges calculated with the 6-31G* basis set; the other charges were calculated with the STO-3G basis set.

(g-h)Charges used in calculations involving the smaller perturbed group (Models 2 and 3).

(i) Charges used for the aspartic acid residue.

Figure 2.3

Example of dihedrals assigned the same torsional parameters.

Figure 2.4

Complexes used in the derivation of 6-12 parameters for fluorine (a) and hydroxyl hydrogen (b).

Figure 2.5

(a)Thermodynamic cycle for the calculation of the difference in free energy of interaction between two inhibitors.
(b)Thermodyanmic cycle for the calculation of the free energy of hydration relative to the gas phase.

Figure 2.6

Key hydrogen bonds between inhibitor peptide and penicillopepsin.
(a) Difluorostatone residue.
(b) Remainder of inhibitor peptide.

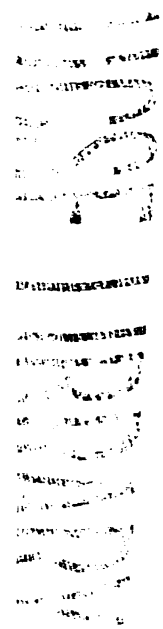


Figure 2.7

Stereoviews of the environment around the hydroxyls of the statone/statine residue and Asp 33 and 213 of penicillopepsin, crystal structure and calculations using Model 1.

(a) The crystal structure of the difluorostatone/penicillopepsin complex. (Hydrogen atoms were placed randomly with reasonable constraints on bond lengths and angles, and should not be considered respresentative of the proper hydrogen bonding).

(b) A snapshot of the same structure after the addition of waters, energy minimization, and 19 picoseconds of equilibration by molecular dynamics.

(c) A snapshot of the structure resulting from mutation of the molecular dynamics-equilibrated structure to the statine (Lambda=0) form (atom OXT is now a hydrogen).

(d) A snapshot of the structure resulting from mutation of structure (c) back to the statone (Lambda=1) form.

Figure 2.8

Stereoviews of the environment around the hydroxyls of the statone/statine residue and Asp 33 and 213 of penicillopepsin, Model 2.

(a) A snapshot of the structure after equilibration by molecular

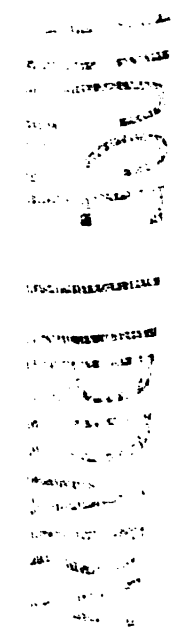
dynamics.

(b) A snapshot of the structure resulting from mutation of the molecular dynamics-equilibrated structure to the statine (Lambda=0) form (atom OXT is now a hydrogen).

(c) A snapshot of the structure resulting from mutation of structure (c) back to the statone (Lambda=1) form.

Figure 2.9

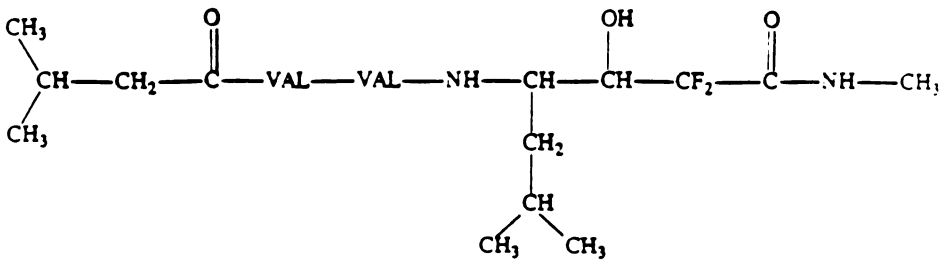
Stereoviews of the environment around the hydroxyls of the statone/statine residue and Asp 33 and 213 of penicillopepsin, Model 3.



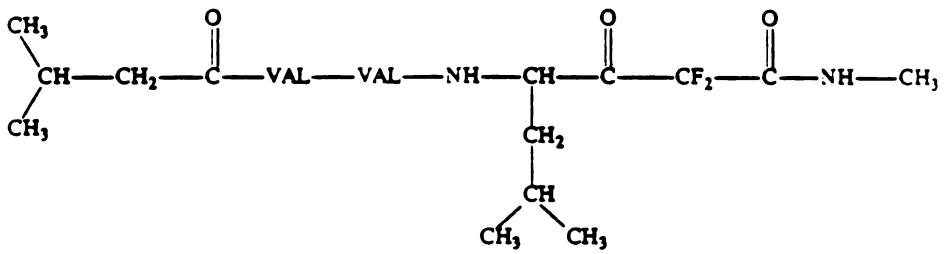
(a) A snapshot of the structure after equilibration by molecular dynamics.

(b) A snapshot of the structure resulting from mutation of the molecular dynamics-equilibrated structure to the statine (Lambda=0) form (atom OXT is now a hydrogen).

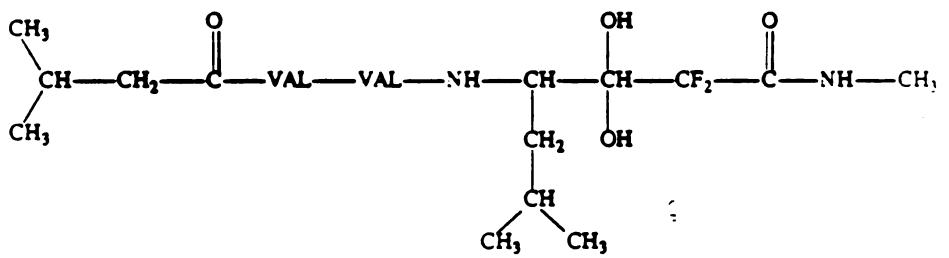
(c) A snapshot of the structure resulting from mutation of structure (c) back to the statone (Lambda=1) form.



(A)



(B)



(C)

Figure 2.1

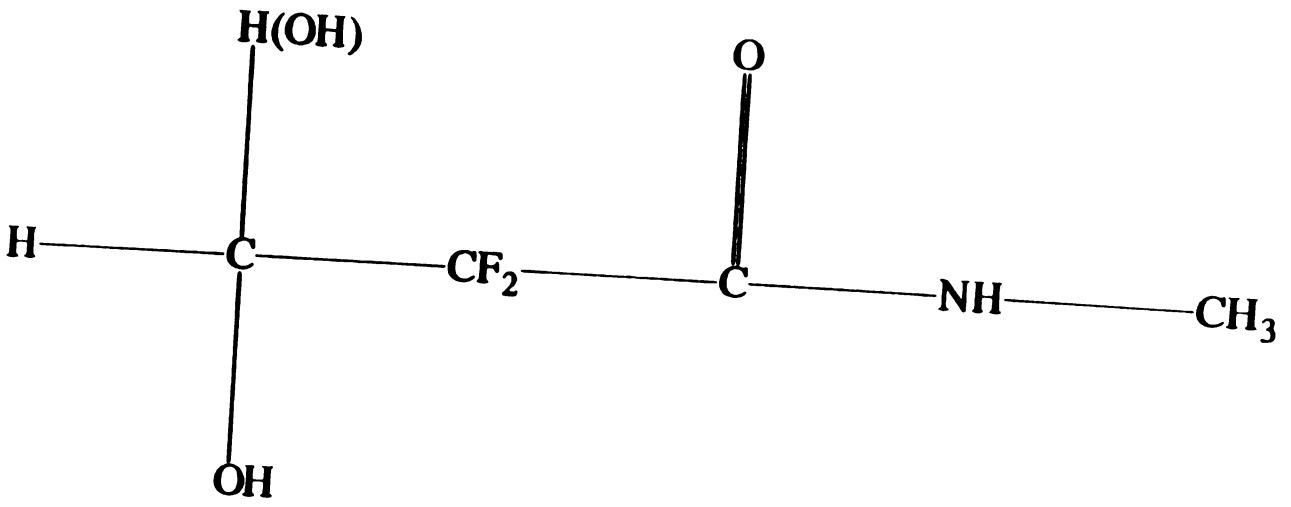


Figure 2.2(a)

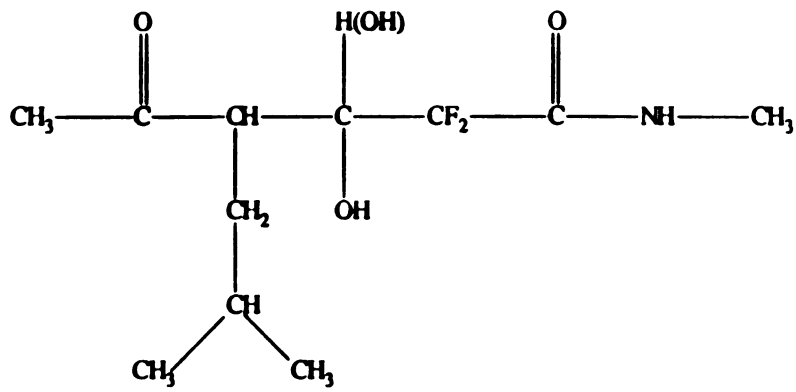


Figure 2.2(b)

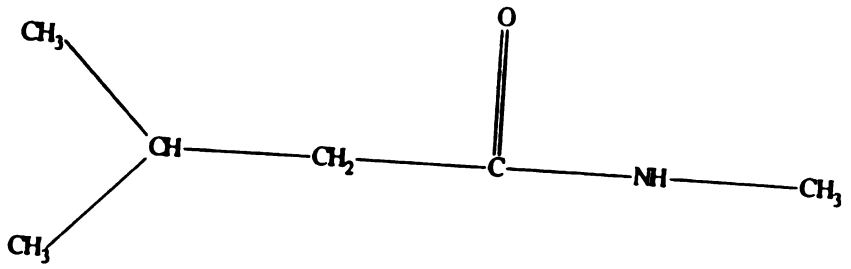


Figure 2.2(c)

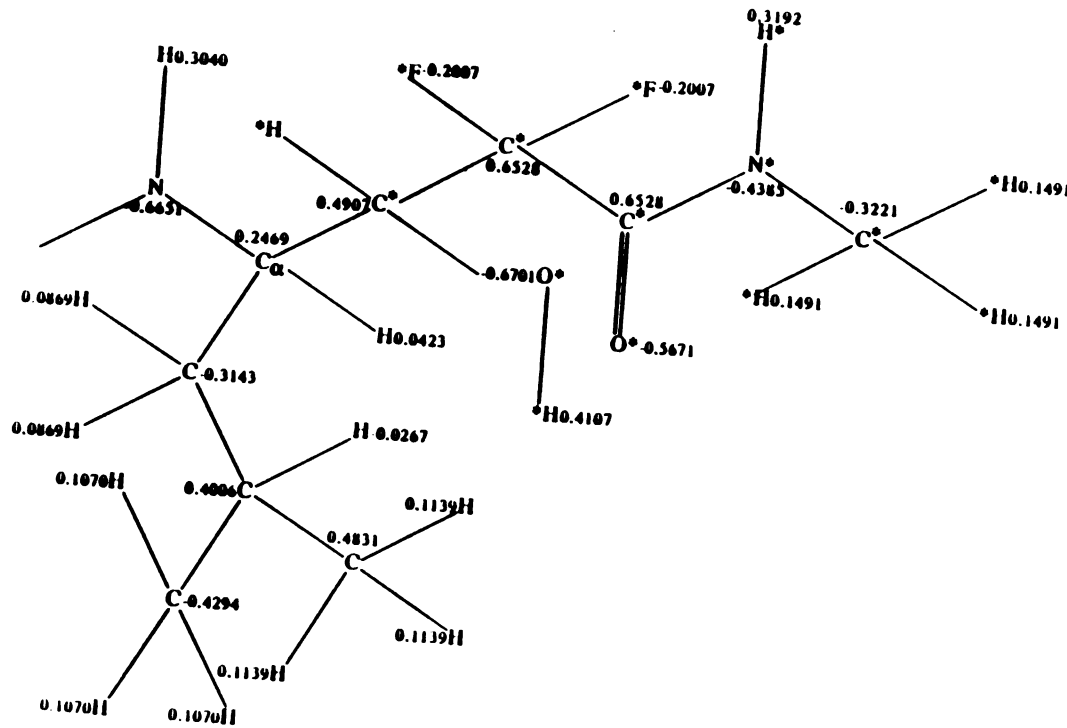

0.3192
H*
Ho.3040
*F-0.2007
*F-0.2007
*H
C*
0.6528
N*
-0.4383
*Ho.1491
N
-0.6651
0.4907C*
0.6528
C*
-0.3221
C*
0.2469
Cα
-0.6701O*
*Ho.1491
*Ho.1491
0.0869H
Ho.0423
O*-0.5671
C-0.3143
*Ho.4107
0.0869H
H-0.0267
0.1070H
0.4006C
0.1139H
0.4831
C
C-0.4294
0.1139H
0.1139H
0.1070H
0.1070H

Figure 2.2(d)

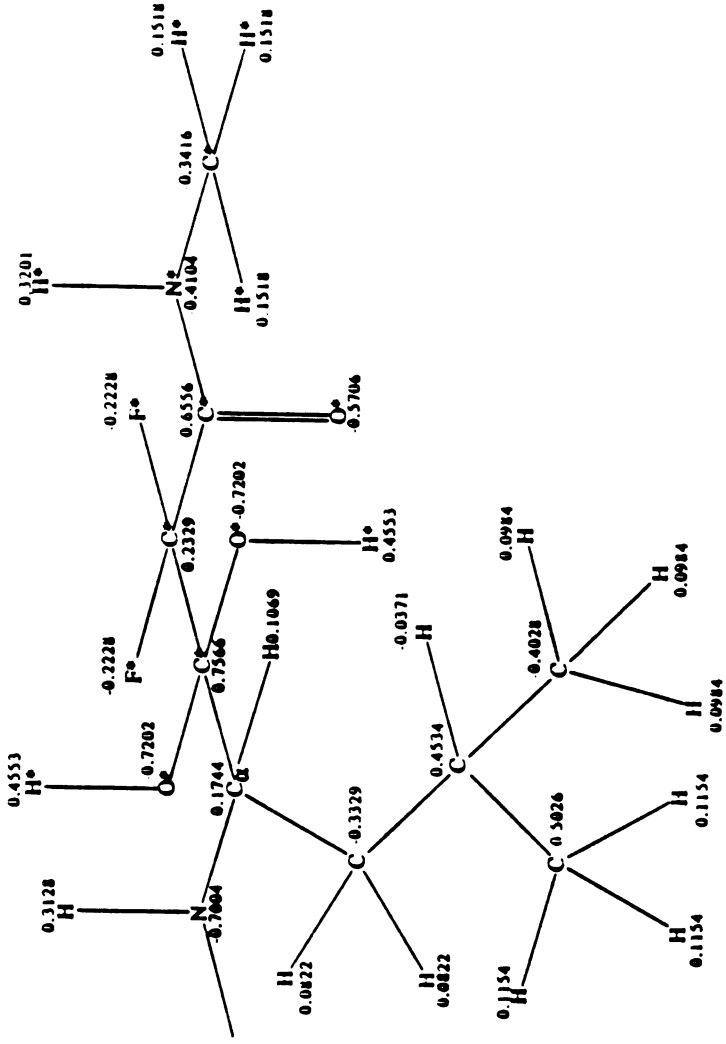


Figure 2.2(e)

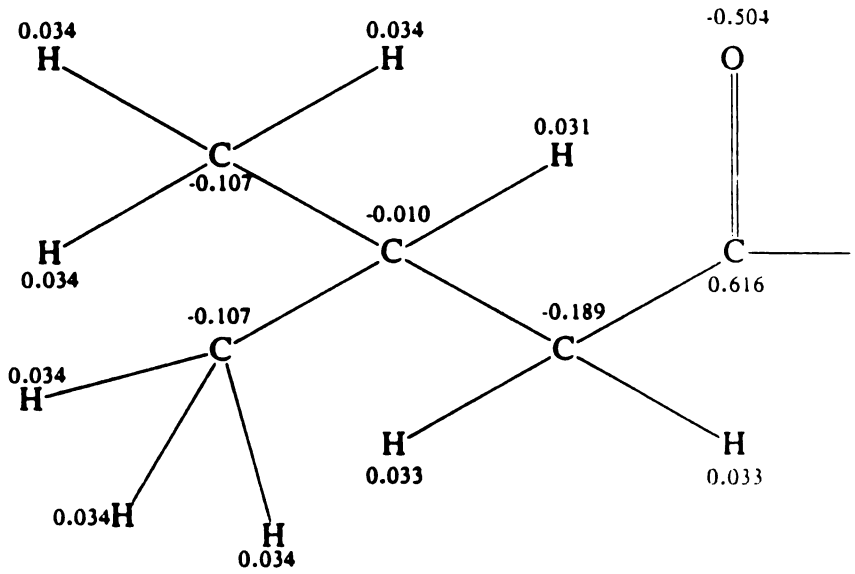

0.034
H
0.034
H
-0.504
O
0.031
H
C
-0.107
-0.010
C
C
H
0.034
0.616
-0.107
C
-0.189
C
0.034
H
H
0.033
H
0.033
0.034H
H
0.034

Figure 2.2(f)

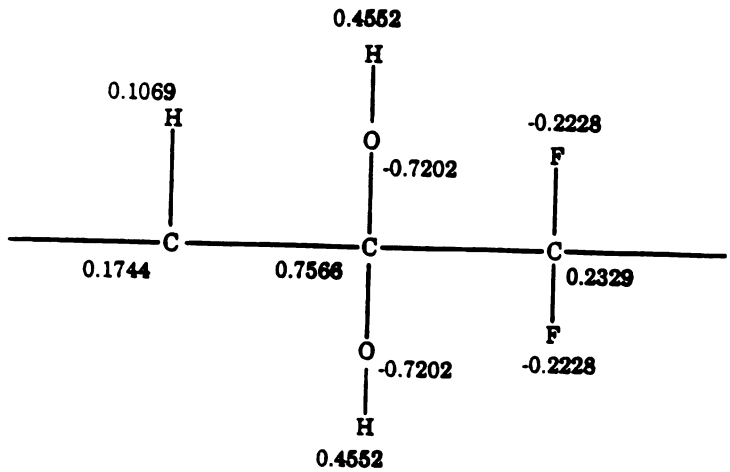

0.4552
H
0.1069
H
O
-0.7202
-0.2228
F
C
C
C
0.1744
0.7566
0.2329
F
O
-0.7202
-0.2228
H
0.4552

Figure 2.2(g)

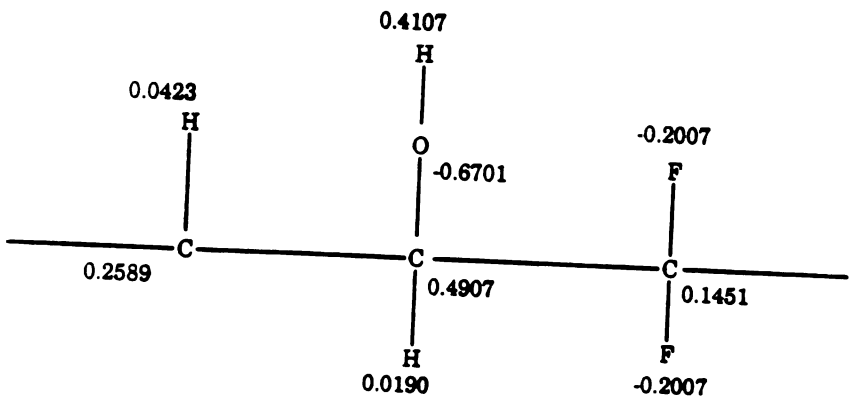

0.4107
H
0.0423
H
O
-0.6701
-0.2007
F
C
C
C
0.2589
0.4907
0.1451
H
F
0.0190
-0.2007

Figure 2.2(h)

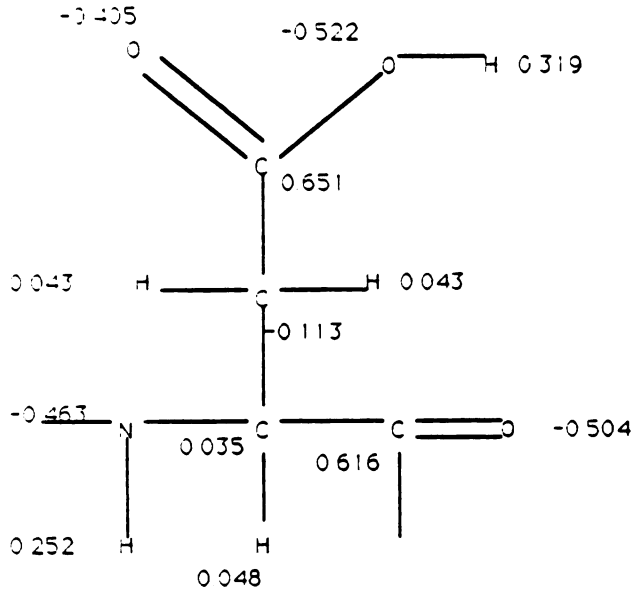
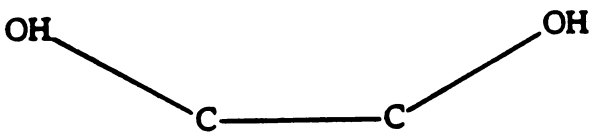

-0 405
O
-0 522
O
H 0 319
C
0 651
0 043
H
C
H 0 043
0 113
-0 463
N
C
C
O
-0 504
0 035
0 616
0 252
H
H
0 048

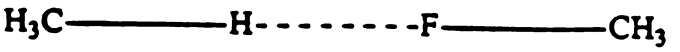
Figure 2.2(i)

(A)

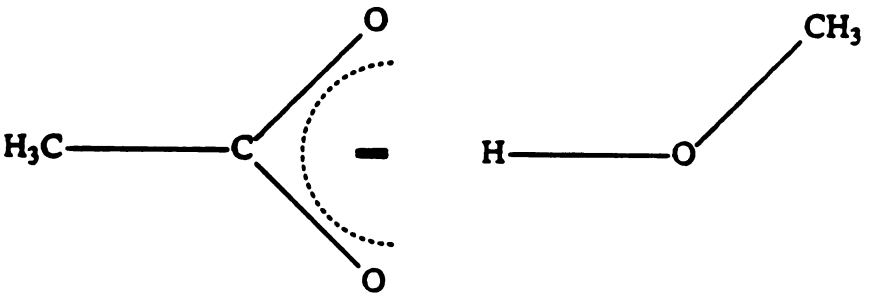


(B)

Figure 2.3



(A)



(B)

Figure 2.4

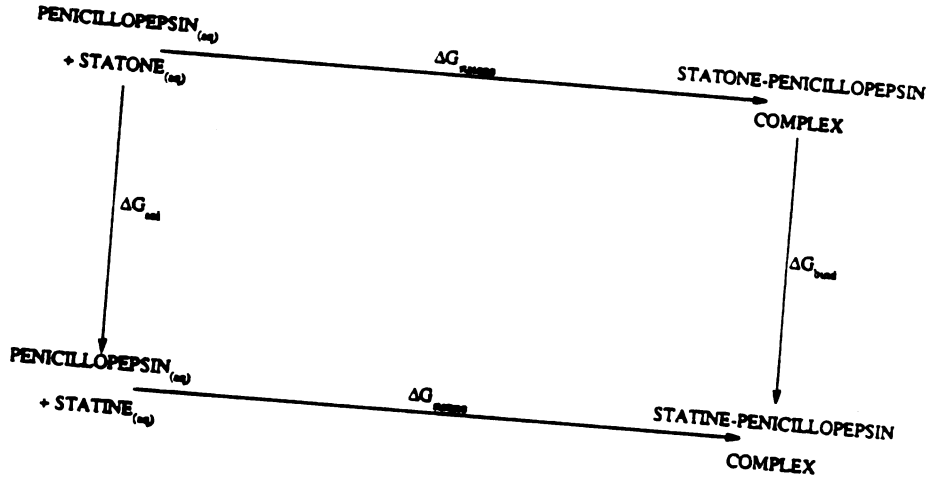

PENICILLOPEPSIN(aq)
+ STATONE(aq)
ΔG bound
STATONE-PENICILLOPEPSIN
COMPLEX
ΔG aq
ΔG bound
PENICILLOPEPSIN(aq)
+ STATINE(aq)
ΔG bound
STATINE-PENICILLOPEPSIN
COMPLEX

Figure 2.5(a)

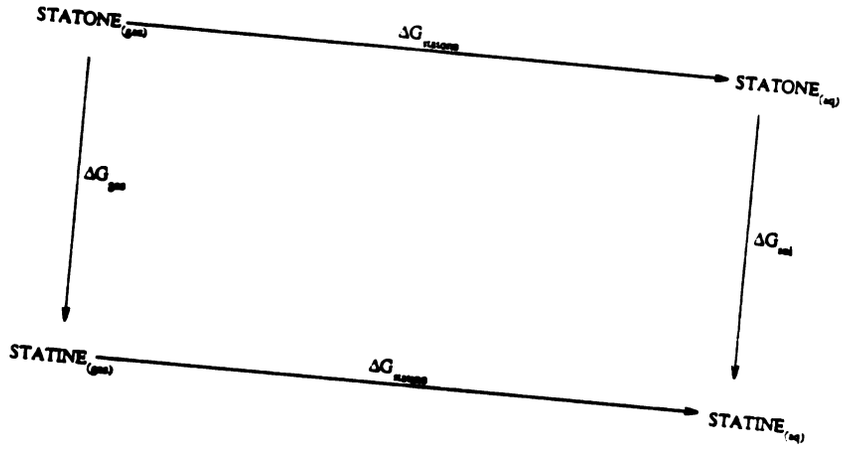


Figure 2.5(b)

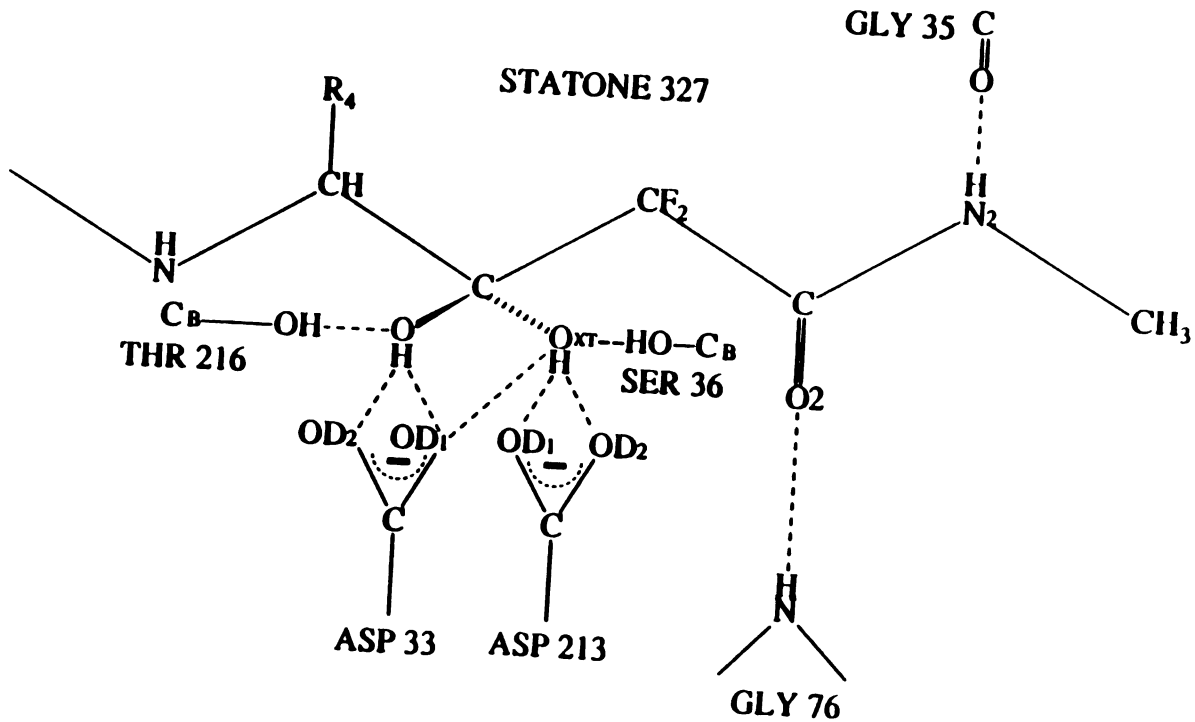

GLY 35 C
O
STATONE 327
R4
CH
CE2
H
N2
H
N
C
C
CB
OH
O
OXT--HO–CB
CH3
THR 216
H
H
SER 36
O2
OD2
OD1
OD1
OD2
C
C
H
N
ASP 33
ASP 213
GLY 76

Figure 2.6(a)

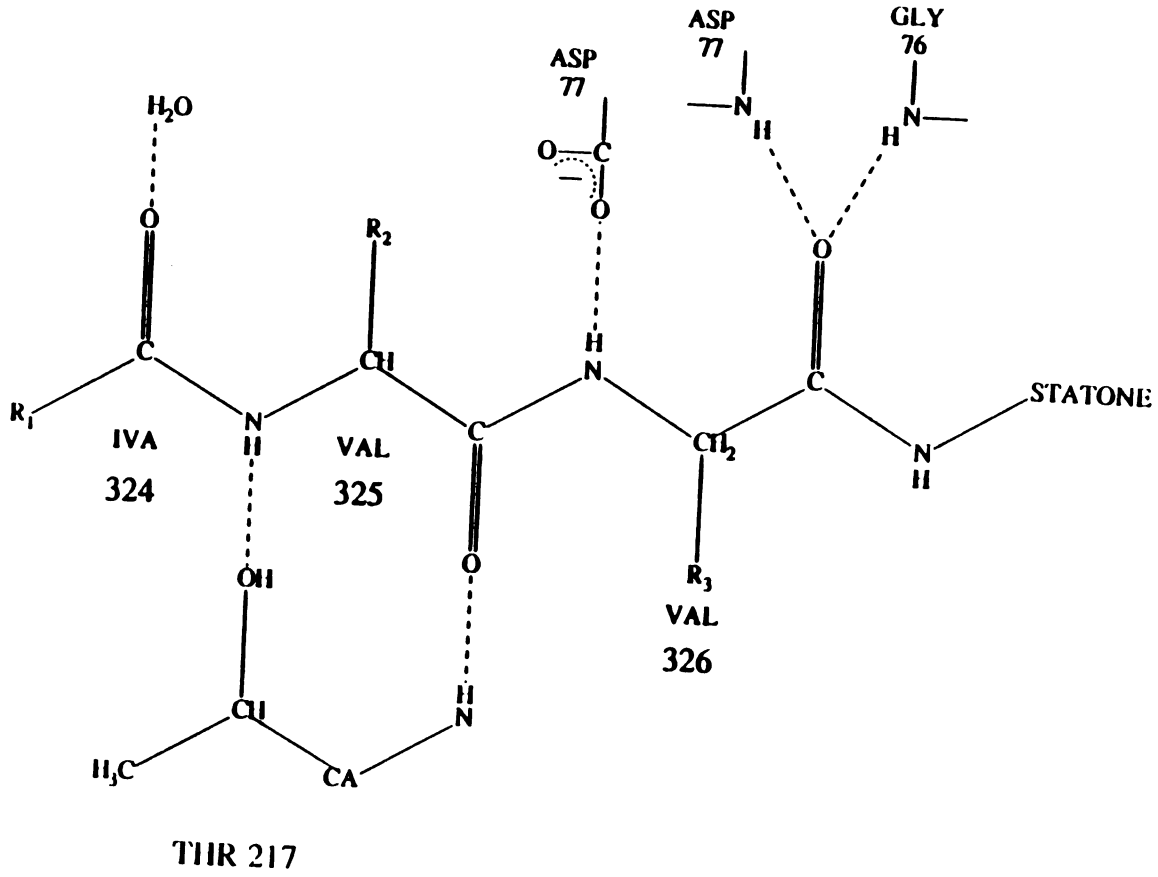

ASP
77
GLY
76
ASP
77
H_2O
R_2
STATONE
R_1
IVA
324
VAL
325
R_3
VAL
326
THR 217

Figure 2.6(b)

dist 0 #0:33@OD2 #0:327@O 2.65
dist 1 #0:213@OD2 #0:327@O 2.64
dist 2 #0:213@OD1 #0:327@O 2.91
dist 3 #0:33@OD1 #0:327@OXT 2.60
dist 4 #0:33@OD2 #0:327@OXT 3.28
dist 5 #0:33@OD1 #0:36@OG 2.61
dist 6 #0:213@OD2 #0:216@OG1 2.95
dist 7 #0:33@OD2 #0:35@N 2.96

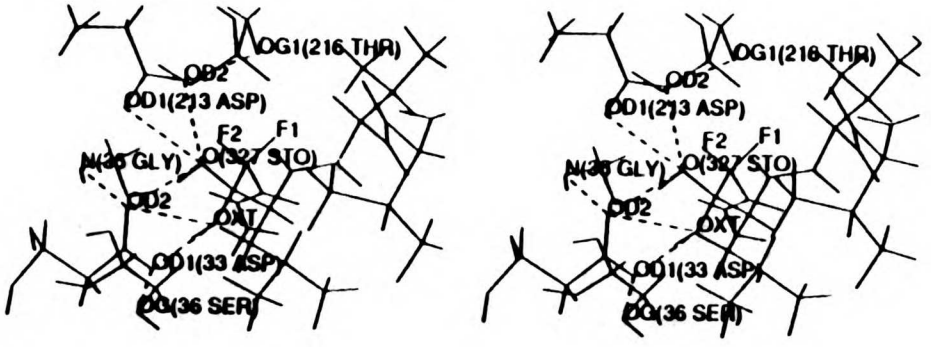


Figure 2.7(a) UCSF MidasPlus

dist 0 #1:33@OD2 #1:327@O	3.68
dist 1 #1:213@OD2 #1:327@O	2.51
dist 2 #1:213@OD1 #1:327@O	3.75
dist 3 #1:33@OD1 #1:327@OXT	3.87
dist 4 #1:33@OD2 #1:327@OXT	2.52
dist 5 #1:33@OD1 #1:36@OG	3.59
dist 6 #1:213@OD2 #1:216@OG1	2.66
dist 7 #1:33@OD2 #1:35@N	2.97

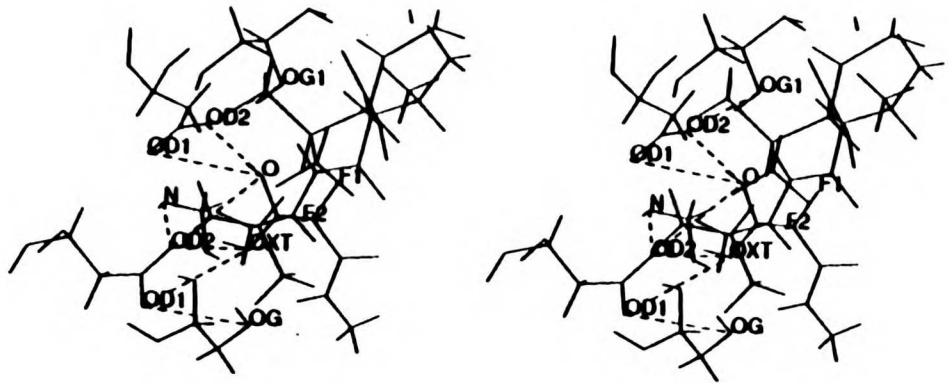


Figure 2.7(b)

UCSF MidasPlus

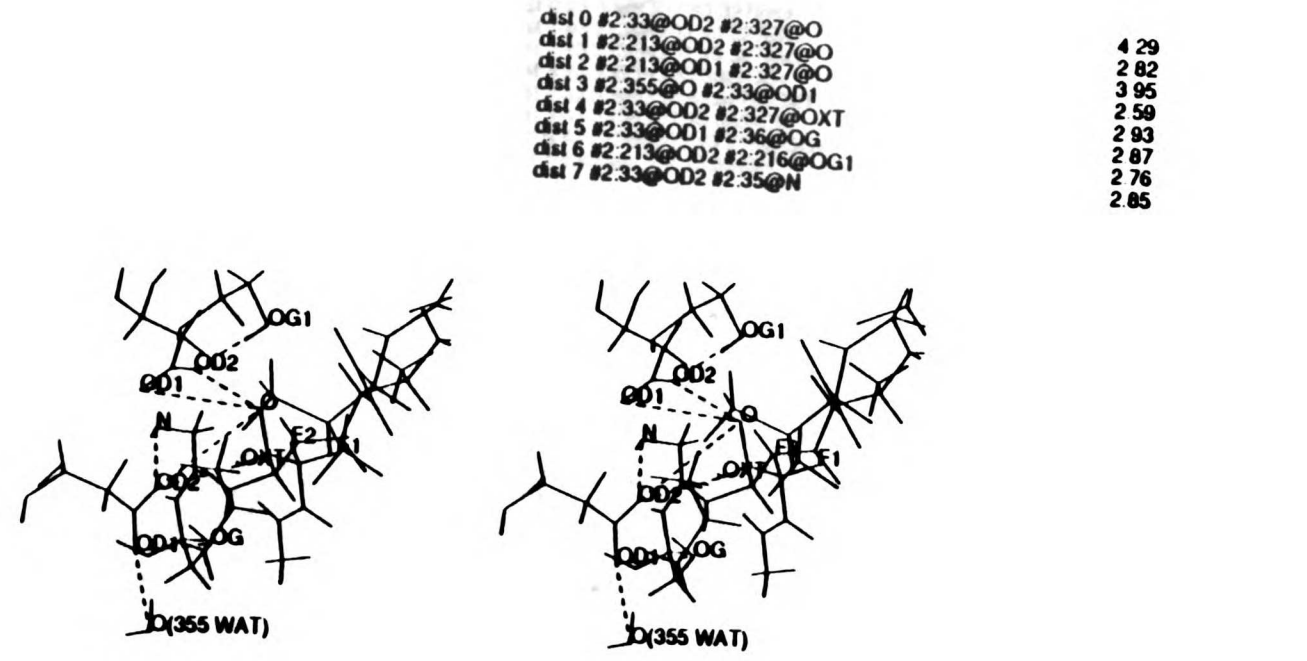


Figure 2.7(c)

UCSF MidasPlus

dist 0 #3:33@OD2 #3:327@O	4.33
dist 1 #3:213@OD2 #3:327@O	2.81
dist 2 #3:213@OD1 #3:327@O	3.70
dist 3 #3:355@O #3:33@OD1	2.73
dist 4 #3:33@OD2 #3:327@OXT	4.08
dist 5 #3:33@OD1 #3:36@OG	3.64
dist 6 #3:213@OD2 #3:216@OG1	2.67
dist 7 #3:33@OD2 #3:35@N	2.86

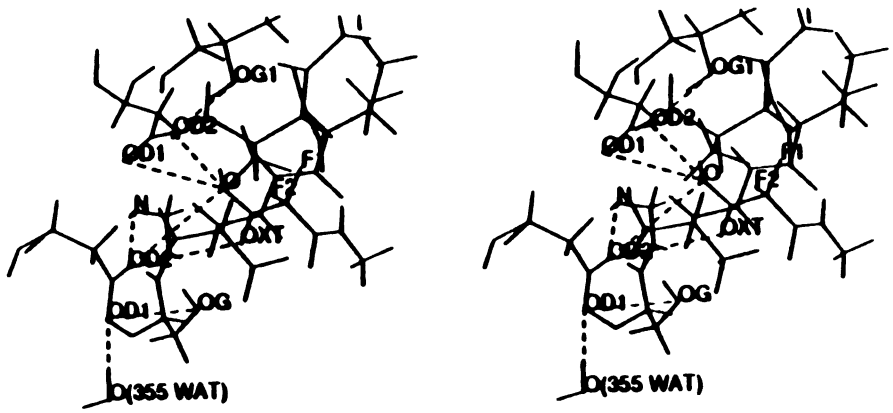


Figure 2.7(d)

UCSF MidasPlus

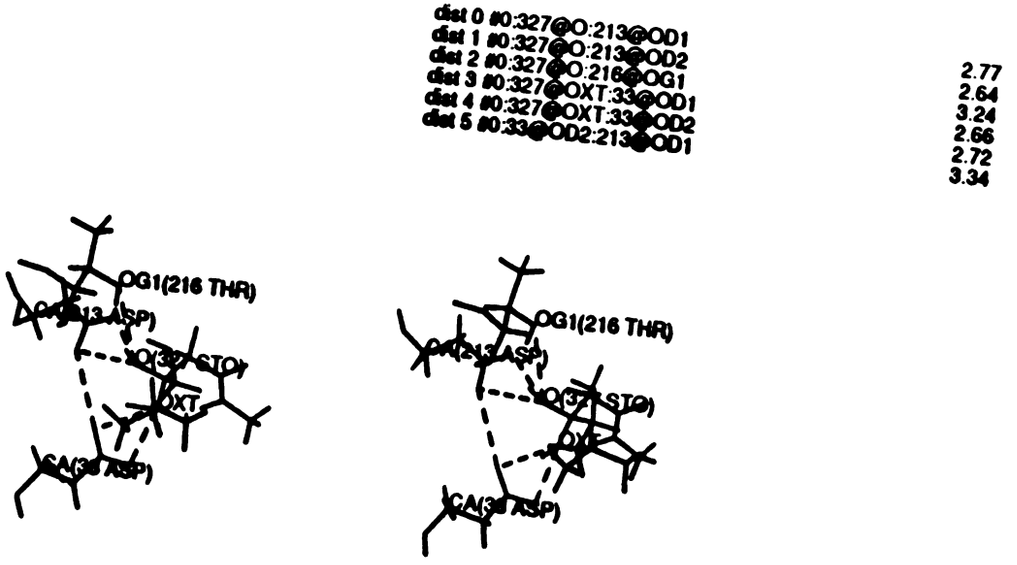


Figure 2.8(a)

UCSF MidasPlus

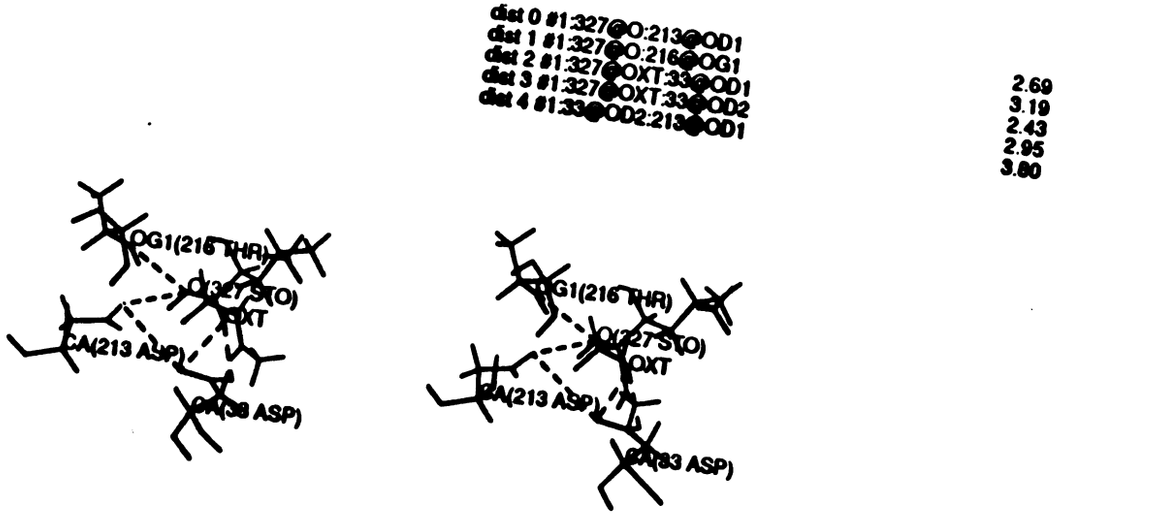


Figure 2.8(b)

UCSF MidasPlus

dist 0 #2:327@O:213@OD1	3.07
dist 1 #2:327@O:213@OD2	2.65
dist 2 #2:327@OXT:33@OD1	3.56
dist 3 #2:327@OXT:33@OD2	2.72
dist 4 #2:327@OXT:36@OG	3.00
dist 5 #2:33@OD2:213@OD1	4.10

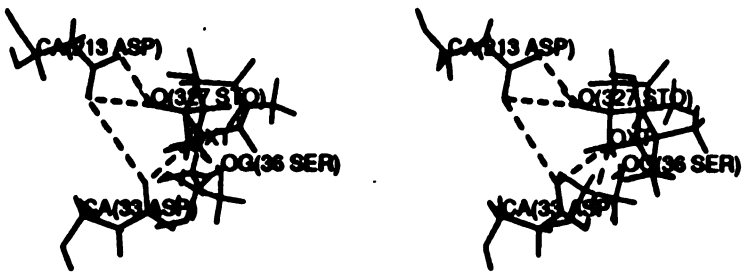


Figure 2.8(c)

UCSF MidasPlus

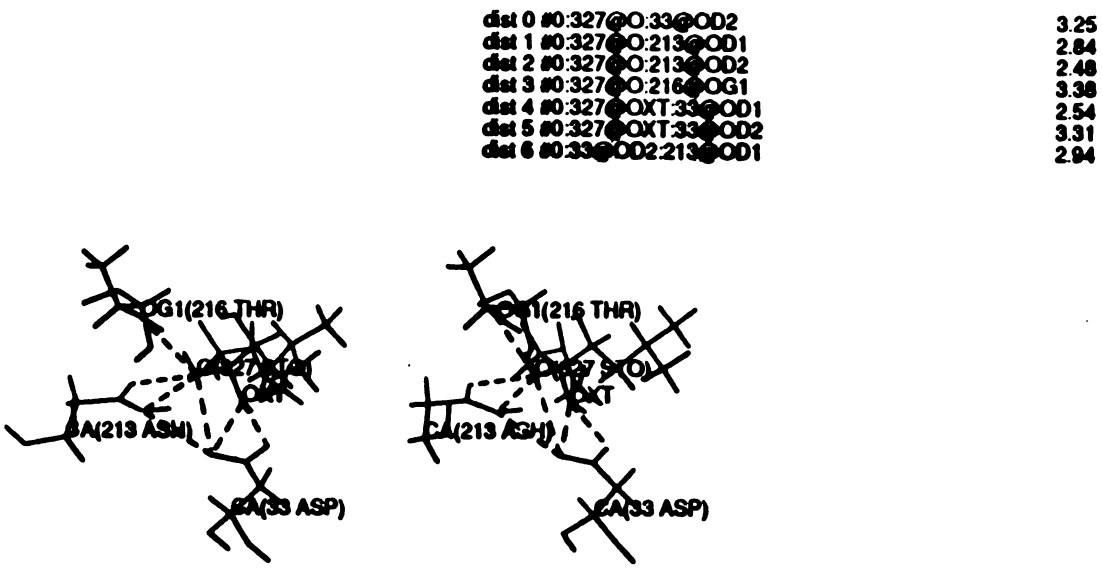


Figure 2.9(a) UCSF MidasPlus

dist 0 #1:327@O:33@OD1 3.11
dist 1 #1:327@O:33@OD2 2.85
dist 2 #1:327@O:213@OD1 2.70
dist 3 #1:327@O:213@OD2 3.59
dist 4 #1:327@OXT:33@OD1 2.51
dist 5 #1:33@OD2:213@OD1 2.87

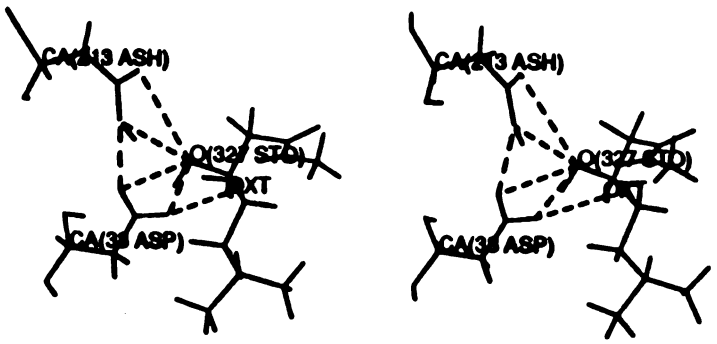


Figure 2.9(b)

UCSF MidasPlus

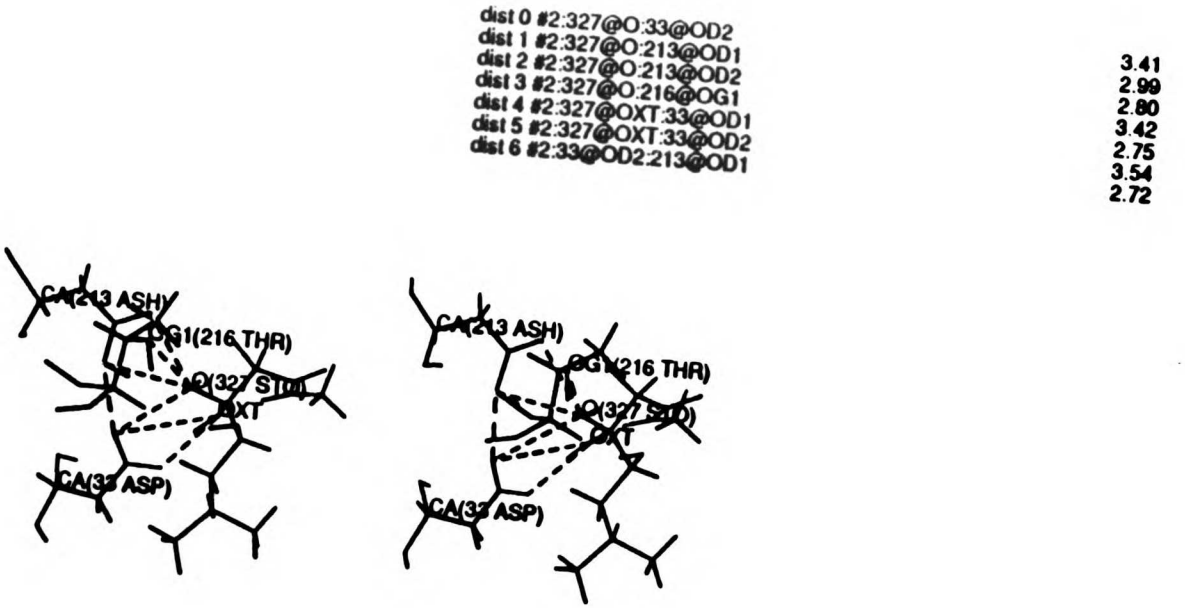

dist 0 #2:327@O:33@OD2
dist 1 #2:327@O:213@OD1
dist 2 #2:327@O:213@OD2
dist 3 #2:327@O:216@OG1
dist 4 #2:327@OXT:33@OD1
dist 5 #2:327@OXT:33@OD2
dist 6 #2:33@OD2:213@OD1
3.41
2.99
2.80
3.42
2.75
3.54
2.72
CA(213 ASH)
OG1(216 THR)
O(327 STO)
OXT
CA(33 ASP)
CA(213 ASH)
OXT
CA(33 ASP)
UCSF MidasPlus

Figure 2.9(c)

CHAPTER THREE

ON THE DERIVATION OF FLUORINE AND HYDROGEN ATOM PARAMETERS USING LIQUID SIMULATIONS

This chapter will be published in a future issue of the Journal of Computational Chemistry, as "Derivation of Fluorine and Hydrogen Atom Paramters Using Liquid Simulations", by Craig A. Gough, Stephen, E. DeBolt, and Peter A. Kollman, Copyright (c) 1992 John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.

INTRODUCTION

In view of the widespread use of molecular mechanics, molecular dynamics, and free energy perturbation calculations for the design and optimization of new drugs and enzyme inhibitors, it is essential that the parameters used in these calculations reproduce intermolecular interactions as accurately as possible. In particular, the parameters involved in nonbonded interactions are of the utmost importance, since it is primarily these parameters that primarily determine the strength of binding of ligands to enzyme active sites or other receptors. These nonbonded interactions include both electrostatic interactions and van der Waals interactions.

In the application of molecular mechanical and dynamical calculations to macromolecules, the electrostatic interactions are usually calculated using atom-centered point charges; these charges are usually determined using quantum-mechanical calculations and are different for each different molecule or residue of interest. The van der Waals parameters, on the other hand, are determined for a limited set of atom types, and are expected to be more transferable between molecules containing the same atom type. This is because the van der Waals parameters are dominated by the filled inner shell of electrons and are not as variable as the electrostatic terms, which are dominated by movement of valence electrons.

The Weiner *et al.* force field[1,2] contains parameters for amino acids and nucleotides, but not for all atom types found in potentially interesting drugs. In particular, as of yet there have been no fluorine parameters derived to be specifically compatible with this force field. The molecular mechanics program MM2[3,4] contains parameters for fluorine, but it is desirable to have a self-consistent set of parameters for a given force field.

The correct parameterization of the fluorine atom is extremely important for the design of new drugs, since the replacement of a hydrogen atom by a fluorine atom or the replacement of a methyl group by a trifluoromethyl group is often used to control the electronic properties of a molecule[5] , or to substitute an extremely stable carbon-fluorine bond for a labile carbon-hydrogen bond[6]. The similarity in size between a fluorine atom and a hydrogen atoms allows such replacements to be almost isosteric.

In addition to the aforementioned uses of fluorine substitution in drug design, another heretofore unexplored possibility exists; namely, that of using the peculiar properties of the trifluoromethyl group to fine-tune the nonbonded interactions of an inhibitor with its target protein and with water. The trifluoromethyl group is polar, and can thus interact electrostatically with both water and with a region of electrostatic potential gradient in a protein active site. However, aqueous solubility data reveal that fluorocarbons are somewhat hydrophobic[7]. This solubility data suggests that replacing a methyl group of an enzyme inhibitor by a trifluoromethyl group could result in a net increase in the binding affinity of an inhibitor for an enzyme active site if a region of favorable electrostatic potential gradient were located at or near the active site. That is, the trifluoromethyl group could have a significantly negative binding free energy, while the free energy of desolvation would be positive but with a smaller magnitude than the binding free energy, due to the hydrophobicity of the trifluoromethyl moiety. Of course, the favorable electrostatic interactions of the trifluoromethyl group with water would be lost, but the overall free energy of desolvation and binding could possibly be negative. This can be contrasted with the situation in which a hydroxyl group is added to an inhibitor: the free energy of desolvation and the free energy of hydrogen bonding in an enzyme active site are about equal and opposite, resulting in no net gain of bind-

ing affinity. The use of computer simulations to test this idea would require accurate fluorine van der Waals parameters.

In recent years, a great deal of effort has been put into the derivation of molecular mechanical van der Waals parameters using Monte Carlo calculations on pure liquids[8,9]. In these simulations, the parameters are fine-tuned to reproduce the experimental density and enthalpy of vaporization of the liquid. The parameters obtained from these liquid simulations include multi-body effects implicitly, giving an effective two-body potential that reproduces some multi-body effects. Some progress has been made[9] in the development of a new all-atom force field for AMBER using Monte Carlo simulations of pure liquids; in particular, new parameters for tetrahedral carbon and for hydrocarbon hydrogen have been derived. These parameters have been shown to be significantly superior to the older parameters for calculation of the enthalpy of vaporization and density of the hydrocarbons and of the free energy of hydration of hydrocarbons

With this in mind, we set out to determine van der Waals parameters for the fluorine atom by performing molecular dynamics calculations on a system consisting of pure tetrafluoromethane, using the tetrahedral carbon parameters of Spellmeyer et al.[9].

A second goal of this project was to use molecular dynamics simulations to determine what van der Waals parameters were required for the hydrogen of trifluoromethane, in order to accurately reproduce the experimental enthalpy of vaporization, ΔH_{vap}, and molar volume, V_M. Since fluorine is an extremely electronegative atom, we reasoned that a hydrogen bonded to a carbon bearing fluorines would likely have a great deal of its electron density withdrawn towards that carbon, resulting in an effective van der Waals radius that would be less than that of a hydrocarbon hydrogen.

We reasoned that it would be necessary to used a somewhat smaller R* on the hydrogen of trifluoromethane in order to accurately simulate intermolecular electrostatic interactions involving this hydrogen. The question to be answered is: just how small must this R* be in order to give good results? The extreme case would be to set R* equal to 0, as it is for hydrogens bonded to oxygen, as in the TIP3P water potential[10]. However, it is more likely that R* will be smaller than the hydrocarbon value of 1.49 Å[9]. An analogous case is that of the hydrogens attached to methanol: molecular dynamics simulation of methanol have shown that an R* smaller than that of pure hydrocarbons appears to be required to accurately simulate the liquid properties[11].

We show in this paper that an R* even smaller than that used for methanol is necessary, and that therefore it is possible to estimate what R* is needed for carbons bearing electron-withdrawing groups, by considering the relative electronegativites of these groups; that is, the R* of hydrogens bonded to a carbon bearing several fluorines must be smaller than that for hydrogens bonded to a carbon bearing an oxygen-containing moiety. These results correspond very well to results obtained from *ab initio* calculations of the interaction of water with trifluoromethane[12].

METHODS

The energy minimizations and molecular dynamics trajectories were calculated with the program AMBER, which uses a potential energy function of the following form:

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

In this potential function, $A_{ij} = \varepsilon_{ij} r^{*12}_{ij}$ and $B_{ij} = 2\varepsilon_{ij} r^{*6}_{ij}$. r^*_{ij} is the internuclear separation of minimum energy, and is equal to $R^*_{ii} + R^*_{jj}$ for an ij atom pair. The R* described in the remainder of this paper is R^*_{ii} for an atom type i (fluorine or hydrogen, in this case), and is equal to one half of r^*_{ii}.

The equilibrium bond angles, bond lengths, angle bending force constants, and bond stretching force constants used are given in Table 3.1.

The Spellmeyer *et al.* tetrahedral carbon van der Waals parameters[9] were used for all calculations; these parameters are given in Table 3.2.

The atomic partial charges used on all the atoms of CF_4 and CF_3H were calculated, as described previously[13], by fitting the electrostatic potential produced by point charges to that of an electrostatic potential derived from *ab initio* quantum mechanical calculations. The quantum-mechanical calculations were single-point calculations, using the experimental geometry[14,15]. The 6-31G* basis set was used in all cases. The charges used are given in Table 3.3.

Both simulations involved determining R* and ε for fluorine and hydrogen by varying these quantities to give molar volumes and enthalpies of vaporization, calculated from the simulations, that matched the experimental values of these quantities. The molar volume, V_M, was calculated from the size of the periodic box containing 125 molecules. The enthalpy of vaporization per mole, ΔH_{vap}, was calculated by subtracting

the potential energy of the liquid phase from that of a single-molecule gas phase calculation using the same parameters (correcting for the number of molecules in a mole) and adding RT (the gas constant times the absolute temperature). That is,

$$\Delta H = \Delta E + \Delta PV \approx \Delta E + \Delta nRT = \Delta E + \mathrm{RT}$$

where the vapor phase has been assumed to behave as an ideal gas.

The experimental enthalpy of vaporization of CF_4 is available from two sources. One of these experimental values was measured at the boiling point (at one atmosphere pressure) of 145 degrees K[16], and we found no heat capacity data that would enable us to calculate the enthalpy of vaporization at another temperature. Therefore, the molecular dynamics simulation was run at this temperature, and we varied the van der Waals parameters to duplicate this enthalpy of 3.01 kcal/mole to within 5%. There was initially some trepidation on our part as to whether the system would show begin to evaporate at this temperature, but evaporation did not occur. Considering the loose temperature coupling involved in molecular dynamics simulations, this is not surprising. After the CF_4 simulations had been completed, a second set of experimental values for ΔH_{vap} was published by Rubio *et al.*[17]. ΔH_{vap} for CF_4 was reported as 2.91 kcal/mole and 2.80 kcal/mole at 140 K and 150 K, respectively, at the equilibrium vapor pressure of CF_4. A linear interpolation between these two values to 145 K gives a ΔH_{vap} of 2.86 kcal/mole, well within 5% of our calculated value (see the Results section).

We initially found the experimental molar volume of CF_4 of 54.8 $cm^3 mol^{-1}$ in the literature at 145 K and a pressure that was not given in the reference[18]. (The experimental pressure was the equilibrium vapor pressure of CF_4 at the given temperature inside a closed system). We assumed that the molar volume was not very sensitive to pressure,

and that this molar volume would approximately describe the system at one atmosphere. The later work of Rubio *et al.*[17] gives the molar volume of CF_4 as 53.8 cm^3mol^{-1} and 55.7 cm^3mol^{-1} at 140 and 150 K, respectively; this results in an interpolation to 54.8 at 145 K, identical to the value used as discussed above.

We initially found the experimental molar volume of CF_3H in the literature only at 173 K, where it is equal to 46.1 cm^3mol^{-1} [19]. The CF_3H simulations therefore aimed at reproducing the molar volume and the ΔH_{vap} at 173 K. The later work of Rubio *et al.*[17] reports values of 46.3 and 47.2 cm^3mol^{-1} at 170 and 180 K, respectively; a linear interpolation gives 46.6 cm^3mol^{-1} at 173 K. Our calculated value is within 5% of both of these values.

We initially found a literature value of 3.994 kcal/mol for the experimental enthalpy of vaporization of CF_3H at one atmosphere and 191 K (the normal boiling point)[20]. The enthalpy of vaporization at 173 K was determined from this value, using the heat capacity of CF_3H in the gas and liquid phases, which were relatively constant from 173 to 191 K[20]. ΔH_{vap} at 173 K was calculated using the thermodynamic cycle shown in Figure 3.1. The enthalpy of vaporization of CF_3H at 173 K can be calculated as follows:

$$\Delta H_{vap,\,173K} = \Delta H_{vap,\,191K} + \Delta H_{liq,\,173\text{–}191} + \Delta H_{gas,\,191\text{–}173} = \Delta H_{vap,\,191K} + C_{P,liq}\Delta T - C_{P,gas}\Delta T$$

The heat capacity of liquid trifluoromethane varies only slightly in the between 173 and 191 K, from 20.2 $cal\ mol^{-1}K^{-1}$ to 20.6 $cal\ mol^{-1}K^{-1}$ [20]; the average value of 20.4 $cal\ mol^{-1}K^{-1}$ was used. The heat capacity of trifluoromethane in the gas phase also varies very little in this temperature range; a value of 9.6 $cal\ mol^{-1}K^{-1}$ [20] was used. The

enthalpy of vaporization of trifluoromethane at 191 K is 3.994 kcal/mol[20] . Using these data, the enthalpy of vaporization at 173 K can thus be calculated to be 4.2 kcal/mole.

The more recent work of Rubio *et al.*[17] reports the ΔH_{vap} of CHF_3 to be 4.28 and 4.17 kcal/mole at 170 and 180 K, respectively; linear interpolation to 173 K gives a ΔH_{vap} of 4.25 kcal/mole, within 5% of our calculated value.

This simulation was run using the customary methods of periodic boundary conditions[21] to avoid edge effects.

The simulated system consisted of a box of 125 molecules of CF_4. A simple three-dimensional (5 x 5 x 5) cubic lattice was used to prepare the initial solvent configurations. The center of mass of a molecule (CF_4 or CHF_3) was placed at each of the 125 lattice points in the all positive octant. The spacings between lattice points were computed so that the initial volume of the solvent "box" corresponded to the expected final equilibrium volume of the media. Each molecule was displaced a random distance (within 0.1 angstrom) from its lattice point along each of the three axes and rotated about each axes through a random angle (0 to 360 degrees). This was done in order to break the lattice symmetry and facilitate more rapid progress toward thermodynamic and structural equilibrium during initial molecular dynamics.

The prepared solvent boxes were then minimized. The CF_4 box was minimized to a gradient of 0.5 $kcal\ mol^{-1}\text{Å}^{-1}$, with pairlist regeneration every 25 steps, and a nonbonded cutoff of 6.0 Å. The CHF_3 box was minimized to a gradient of 0.1 $kcal\ mol^{-1}\text{Å}^{-1}$, with pairlist regeneration every 40 steps, and a nonbonded cutoff of 8.0 Å. Amber3a was used, in which the first atom only is used as a criterion in the search for interacting residue pairs. The minimization of each system was only performed once, and the coordinates from these minimizations were used to begin the molecular dynamics runs for

every set of parameters used. For the Amber MD system topology, one molecule of the 125 was considered a solute (due to logistical requirements of the program only), with the other 124 as solvents.

All molecular dynamics calculations were performed using Amber3a, using either constant volume or constant pressure; the constant pressure calculations were carried out using Berendsen's algortihm for pressure coupling[21]. The nonbonded pairlist, including only those nonbonded interactions within 8 Å, was updated every 40 dynamics steps.

The same equilibration protocol was followed for each value of R* and ε tried. After the system was minimized, 5 ps of constant-volume molecular dynamics was performed, with the box size determined by the experimental molar volume. This was followed by a further 5 ps of constant-pressure molecular dynamics at one atmosphere pressure, before any determination of the volume or enthalpy of vaporization was attempted (other than monitoring whether or not evaporation was occurring). Following this, a 10 ps constant-pressure data collection run was performed, taking averages of the potential energy and box volume every 1 ps to get an approximate determination of the proximity of the molar volume and enthalpy of vaporization to the desired values. One 10 ps single molecule gas phase simulation was run, using the same equilibrium bond length, bond angle, and bond-bending force constant, to get a gas phase potential energy for the calculation of the enthalpy of vaporization. Since there are no nonbonded interactions in this gas-phase simulation, this gas phase potential energy will be independent of the R* and ε of fluorine, and the same gas-phase potential energy can be used in all of the ΔH_{vap} calculations.

Once R* and ε for fluorine were determined from the CF_4 simulation, these values of the van der Waals parameters were used for the fluorine atoms in the trifluoromethane simulation.

Equilibration of the system consisting of a periodic box of 125 trifluoromethane molecule proved to be much more difficult. Initial equilibration involved the same protocol used for tetrafluoromethane; that is, energy-minimization, followed by 5 ps of constant-volume molecular dynamics at the target temperature (173 degrees K in this case), followed by 5 ps of constant-pressure dynamics at this temperature. Several values of R* and epsilon were tried for the hydrogen of CF_3H, all of which gave molar volumes that were much larger than the experimental value. These values in fact were all greater than the molar volume of CF_4, which is physically unrealistic since CF_3H has a smaller molecular volume than CF_4 and has significant electrostatic intermolecular attractions. The potential energy of the liquid was also too high to match the experimental ΔH_{vap} (i.e., the magnitude of ΔH_{vap} would be too small). These physically unrealistic values of the molar volume and of ΔH_{vap} were obtained over a wide range of values of the van der Waals parameters of the hydrogen: R* varied from 0 to 1.29 Å, and ε values of 0 or 0.015 were used.

We attempted to find a more stable configuration of the system by heating the system to 250 K and then cooling it slowly back to 173 K, all at constant volume. When this protocol was followed by a constant-pressure dynamics run at 173 K, the molar volume was once again too great, and the potential energy too high, even though this simulation was carried out with R* and ε for hydrogen equal to 0.

A successful protocol for finding a more stable configuration of the system involved starting a constant-volume simulation at 189 K (two degrees below the boiling

point), beginning with the final coordinates and velocities of a constant-volume dynamics run at 190 K. This run was one of the intermediate simulations when cooling from 250 K, as described in the previous paragraph, with the hydrogen R* and ε equal to 0. This 189 K simulation was run for 25 ps. Within 0.1 ps (100 dynamics steps) of beginning this run, the kinetic energy became very large, and the potential energy became very large and positive (2070 kcal/mole). Over a period of 10 ps, the system gradually settled into a stable configuration, with a potential energy of approximately -420 kcal/mole. This energy remained more or less constant for another 15 ps. When this system was cooled to 173 K, still at constant volume, the potential energy was -457 kcal/mole, considerably lower than the original potential energy of -378 kcal/mole in the initial constant volume simulations at 173 K. So the system had indeed found a more stable configuration. However, when a constant pressure simulation at 173 K was begun using the coordinates and velocities from the constant volume 173 K run in the more stable configuration, the molar volume was again found to be much too large, even though R* and ε for hydrogen were equal to 0.

The source of this problem was found to be the C-F bond length used in the simulation. Up to this point, we had been using 1.38 A as the equilibrium length of the C-F bond, a value used for all C-F bonds in the MM2 molecular mechanics program. However, the experimental C-F bond length in trifluoromethane has been determined as 1.332 Å[14]. When this bond length was used, the molar volume of the system was sensitive to the value of R* used for hydrogen, and it was possible to vary R* to achieve a molar volume corresponding to experiment. The enthalpy of vaporization was still relatively insensitive to the van der Waals parameters used (probably since electrostatic interactions dominate), but was close to the experimental value for all of the R*-ε com-

binations tried.

The protocol for performing a simulation with the new bond length involved first allowing all of the bonds to be flexible; that is, the SHAKE algorithm was turned off and the bonds were represented by a harmonic potential rather than being rigidly fixed (the C-F bond force constant used was 367 $kcal\ mol^{-1}$ $Å^{-2}$). We initially tried running 1 ps of constant volume dynamics with flexible bonds, followed by 25 ps of constant pressure dynamics with flexible bonds, using an $R_{eq,C-F}$ of 1.38 throughout, and with R* and ε of hydrogen set to 0.

However, bond flexibility alone was not enough to achieve the desired molar volume, even with the zero hydrogen van der Waals parameters. With SHAKE still turned off, we began another simulation, using an $R_{eq,C-F}$ of 1.332 and zero hydrogen van der Waals parameters, and starting with the final coordinates from the 1 ps constant-volume run with SHAKE turned off, as described in the previous paragraph. This simulation gave a molar volume, as an average over 50 ps, of 42.4 $cm^3 mol^{-1}$. This is somewhat lower than the experimental molar volume of 46.1, which is to be expected, since the extreme values of R*=0 and ε=0 were used for hydrogen; more physically realistic values of R* and ε gave a V_M in accord with experiment.

It was now possible, using an $R_{eq,C-F}$ of 1.332 Å, to vary R* and ε of the hydrogen to attempt to achieve values of V_M and ΔH_{vap} close to experiment. The parameters used and thermodynamics quantities obtained are described in the Results section. For all of these values, the starting coordinates and velocities were those of the constant-volume, 173 K simulation using flexible bonds, which was obtained from the more stable configuration obtained as described previously in this section. The V_M and ΔH_{vap} are averages over 50 ps. SHAKE was again turned on for these simulations, as we wished

to parameterize the system with rigidly constrained bonds. There were some initial large fluctuations in the energy as SHAKE was switched on, but the system always settled down to a reasonable potential energy and volume.

Once an optimal R* and ε were obtained, an attempt was made to again use a C-F bond length of 1.38 Å, using several value of R* and ε, in the hope that the system had reached a more stable configuration which would allow a smaller V_M with this bond length. None of these attempts were successful; the system expanded to an unrealistically large volume when the larger bond length was used, and showed signs of evaporating; i.e., a slow, steady increase in volume was seen throught the simulation.

RESULTS AND DISCUSSION

Table 3.4 shows the values of R* and ε of fluorine which were used in attempts to match the calculated ΔH_{vap} and V_m of CF_4 with the experimental values.

The initial R* value of 1.55 Å is that determined earlier[22] via an *ab initio* calculation of the optimal interaction distance between the fluorine of fluoromethane and a hydrogen of methane; the ε value of 0.08 kcal/mole was taken from the molecular mechanics program MM2[3,4]. It is obvious that these parameters do not reproduce the experimental values of the liquid properties very well, and the final values of R* and ε which do reproduce these properties satisfactorily are quite different from the quantum-mechanically derived parameters.

As might be expected, the molar volume of CF_4 is approximately proportional to R*. Also, ΔH_{vap} is approximately proportional to ε, since the energy of interaction of

CF_4 molecules is primarily the dispersion energy. The protocol for varying R* and ε thus involved keeping one of the two parameters constant and varying the other to try to reproduce the desired expermental quantity. Of course, there is some lowering of the potential energy as the molar volume decreases with R*, so the two parameters were not completely separable in their effect on the V_m and ΔH_{vap}; it was necessary to take this into account during the parameterization process.

The R* and ε finally decided upon for tetrafluoromethane were 1.75 Å and 0.061 kcal/mole. These reproduce the expermental V_m and ΔH_{vap} very well; there is essentially 0% deviation between the calculated and experimental V_m, while the difference between the calculated and experimental ΔH_{vap} is only 2%.

Table 3.5 shows the values of R* and ε tried for the hydrogen of CHF_3. The ε value of 0.015 kcal/mole was that found by Spellmeyer[9] and was kept constant. The R* values of 1.49 and 1.29 Å that were tried represent the R* values for, respectively, a hydrocarbon hydrogen derived by Monte Carlo simulations of pure methane[9] and the non-hydroxyl hydrogens of methanol, derived via molecular dynamics simulations of pure methanol[11]. The extreme case of zero R* and ε was also tried.

It can be seen that an R* of 1.49 Å is too large; when this R* is used, the molar volume is significantly larger than the experimental value. An R* of 1.29 Å also gives a V_m that is too large. The extreme values of zero for R* and ε give a reasonable ΔH_{vap}, but the molar volume is too small.

The best fit to experiment occurs when an R* of 1.21 Å and an ε of 0.015 kcal/mole are used; the difference between the calculated and experimental molar volume is only 2%. This is a very satisfying result, since, as was discussed in the Introduction of this paper, one would expect the electron density of a hydrogen atom

attached to a carbon bearing very electronegative atoms to be shifted toward the carbon, allowing closer interactions between this hydrogen and other molecules, and this shift of electron density should be more pronounced if the carbon in question is bonded to more electronegative atoms. Thus, the hydrogen R* should be smaller if fluorines are bonded to the carbon bearing the hydrogen than if an oxygen is bonded to the carbon. The steady decrease in the R* necessary for accurate simulations, as one goes from a hydrocarbon to methanol to a fluorine-bearing carbon, supports the hypothesis of the Introduction.

The accompanying paper by Veenstra *et al.*[12] discusses quantum-mechanical calculations intended to develop hydrogen van der Waals parameters for methyl hydrogens adjacent to electronegative groups. For trifluoromethane, an R* of 1.15-1.20Å was also obtained. It is very encouraging that these two very different approaches give similar R* values.

The calculated enthalpy of vaporization of CHF_3 is consistently lower than the experimental value, regardless of the van der Waals parameters used. The calculated ΔH_{vap} is from 1% to 7% lower than the experimental ΔH_{vap}, but does not seem to be proportional to R* once one goes below a certain value of R*; the ΔH_{vap} values are all approximately the same for all of the R* values less than 1.49 Å. This is physically reasonable, since much of the intermolecular potential energy would be due to relatively longer-range electrostatic interactions; electrostatic interactions are in general much more important than dispersion for this dipolar molecule. It is satisfying that our quantum-mechanically calculated atomic partial charges reproduce the enthalpy of vaporization so well. There was some concern on our part that since the 6-31G* basis set often overestimates dipole moments, the electrostatic energy of the simulation might

be inaccurate, but apparently this basis set does not overestimate the polarization of the electron density distribution when fluorine is a substituent. The dipole moment calculated using our atomic partial charges is 1.77 Debyes, and the dipole moment of the calculated wave function is 1.71 Debyes, neither of which is very different from the experimental dipole moment, which has been measured in the gas phase as 1.59-1.65 Debyes[23]. One would indeed expect the dipole moment of the liquid to be somewhat greater than the gas phase value, due to polarization.

In the Methods section of this paper, we described how it was necessary to use the experimental value of 1.332 Å for the C-F bond length of fluoroform, in order to achieve a stable liquid configuration that did not evaporate. This sensitivity to bond length may be due to the fact that the fluoroform simulation was performed at a temperature only 18 degrees below the boiling point. Although these potential sensitivities in the stability of the trajectory are a subject of concern, once a stable trajectory was achieved for CHF_3 with the correct C-F bond length parameter, we could derive the parameter dependence reported in Table 3.5. We thus recommend that the experimental C-F bond length be used when available.

CONCLUSIONS

The results of the CHF_3 liquid simulation, the methanol simulations of DeBolt and Kollman[11] and of the quantum-mechanical calculations mentioned in the accompanying paper[12] are very interesting in that they show a logical relationship between the size of R* of hydrogen and the electronegativity of the atoms attached to the carbon bearing

the hydrogen in question. For purposes of simulation, it should thus be possible to determine in a predictable manner just what R* value should be used for a hydrogen in a given environment. Use of hydrogen parameters derived in this manner should give better simulation results than the use of the same parameters for a hydrocarbon hydrogen in all environments.

REFERENCES FOR CHAPTER THREE

(1) Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghia, C.; Alagona, G.; Profeta, S.,Jr.; Weiner, P.; *J. Am. Chem. Soc.* 1984, **106**, 765.

(2) Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A.; *J. Comp. Chem.* 1986, **7**, 232.

(3) MM2: Molecular Mechanics II. Quantum Chemistry Program Exchange, Program No. 423.

(4) Meyer, A.Y.; Allinger, N.L.; *Tetrahedron* 1975 **31**, 1971.

(5) The Pharmacological Basis of Therapeutics, A.G. Gilman; L.S. Goodman; T.W. Rall; F. Murad. MacMillan Publishing Company, New York. 1985.

(6) *ibid.*; pp. 1267.

(7) Hine, J.; Mookerjee, P.K.; *J. Org. Chem.*1984, **40**, 292.

(8) For example, see Jorgensen, W.L.; C.J. Swenson; *J. Am. Chem. Soc.* 1985 **107**, 569; and Jorgensen, W.L.; J.D. Madura; C.J.Swenson; *J. Am. Chem. Soc.* 1984 **106**, 6638.

(9) Spellmeyer, D; Kollman, P.A., results in preparation, found for the hydrogen, R^*=1.486 and ε=0.015; subsequent refinement suggests ε=0.0157 kcal/mole leads to slightly better liquid hydrocarbon properties, but given the low density of C-H groups, this is unlikely to significantly affect our results for trifluoromethane given in Table 3.5.

(10) Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L.; *J. Chem. Phys.* 1983, **79**, 926.

(11) DeBolt, S.E.; Kollman, P.A.; unpublished results.

(12) Veenstra, D.; Ferguson, D.; Kollman, P.A.; *J. Comp. Chem.*, previous paper.

(13) Singh, U.C.; Kollman, P.A. *J. Comp. Chem.* 1984, **5**, 129.

(14) Baird, N.C.. *Can. J. Chem.* 1983, **61**, 1567.

(15) Harmony, M.D.; Laurie, V.W.; Kuczkowski. R.H.; Schwendeman, R.H.; Ramsay, D.A.; Lovas, F.J.; Lafferty, W.J.; Maki, A.G.; *J. Phys. Chem. Ref. Data* 1979 **8**, 619.

(16) Selected Values of Chemical Thermodynamic Properties. F.D. Rossini, D.D. Wagman, W.H. Evans, S. Levine, and I. Jaffe. U.S. Government Printing Office, Washington. 1961. Series Title: Circular of the National Bureau of Standards; 500.

(17) Rubio, R.G.; Zollweg, J.A.; Palanco, J.M.G.; Calado, J.C.G.; Miller, J.; Streett, W.B.; *J. Chem. Eng. Data* 1991 **36**, 171.

(18) Terry, M.J.; Lynch, J.T.; Bunclark, M.; Mansell, K.R.; Staveley, L.A.K. *J. Chem. Thermodynamics* 1969, **1**, 413.

(19) "Handbook of Data on Organic Compounds", Volume 5. R.C. Weast and J.C. Grasselli, editors. CRC Press, Inc., Boca Raton, Florida. 1989.

(20) Valentine, R.H.; Brodale, G.E.; Giauque, W.F. *J. Phys. Chem.* 1962, **66**, 392.

(21) Computer Simulation of Liquids. M.P. Allen and D.J. Tildesley. Oxford Science Publications, Oxford, UK, 1987.

(22) Gough, C.; Kollman, P.A. In preparation.

(23) Tables of Experimental Dipole Moments. A.L. McLellan. W.H. Freeman and Co., San Francisco. 1963.

Table 3.1

Bond Stretching and Bending Parameter for Bonds Involving Fluorine		
Bond	K_R (in kcal mol^{-1} $Å^{-2}$)	R_o (in Å)
CT - F of CF_4	367	1.38
CT - F of CHF_3	367	1.332
Angle	**K_o (in kcal mol^{-1} rad^{-2})**	**θ_o (in degrees)**
F-CT-HC	35.0	109.5
F-CT-F	77.0	109.1

Table 3.2

6-12 Parameters Used for Tetrahedral Carbon		
	R* (in Å)	ε (in kcal/mol)
CT	1.91	0.1094

Table 3.3

Atomic Point Charges of Tetrafluormethane and Trifluoromethane (in fractions of an electron)			
Molecule	Charge on Carbon	Charge on Fluorine	Charge on Hydrogen
CF_4	0.756	-0.189	-
CHF_3	0.644	-0.230	0.046

Table 3.4

Variation of V_M and ΔH_{vap} of CF_4 with van der Waals parameters of Fluorine[a]

R*[b]	ε[c]	V_M[d]	ΔH_{vap}[c]
1.55	0.08	42.8	4.03
1.60	0.08	45.4	3.94
1.70	0.08	50.6	3.88
1.70	0.10	49.6	4.73
1.75	0.08	53.3	3.84
1.77	0.063	56.2	3.03
1.73	0.063	53.7	3.09
1.75	0.063	54.8	3.07
1.75	0.061	55.2	2.98

[a] The experimental values of V_M and ΔH_{vap} are 54.8 cm^3mol^{-1} and 3.01 or 2.86 kcal/mole. [b] In Angstroms. [c] In kcal/mole. [d] In cm^3mol^{-1}.

Table 3.5

Variation of V_M and ΔH_{vap} of CHF_3 with van der Waals parameters of Hydrogen[a]

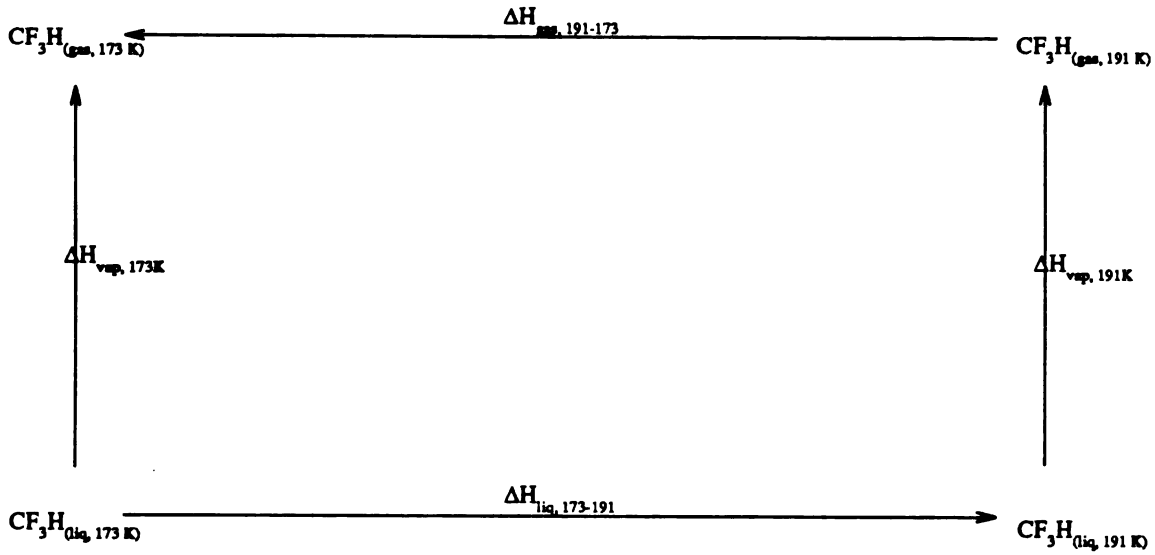
R^{*b}	ε^c	$V_M{}^d$	$\Delta H_{vap}{}^c$
0.00	0.000	43.7	4.15
1.29	0.015	47.8	4.13
1.49	0.015	48.0	3.91
1.21	0.015	46.0	4.11

[a] The experimental values of V_M and ΔH_{vap} are 46.1 cm^3mol^{-1} or 46.6 cm^3mol^{-1}; and 4.19 kcal/mole or 4.25 kcal/mole. [b] In Angstroms. [c] In kcal/mole. [d] In cm^3mol^{-1}.

Figure 3.1

Thermodynamic cycle used for the calculation of the enthalpy of vaporization of CHF_3 at 173 K.

Figure 3.1


$\Delta H_{gas,\ 191\text{-}173}$
$CF_3H_{(gas,\ 173\ K)}$
$CF_3H_{(gas,\ 191\ K)}$
$\Delta H_{vap,\ 173K}$
$\Delta H_{vap,\ 191K}$
$\Delta H_{liq,\ 173\text{-}191}$
$CF_3H_{(liq,\ 173\ K)}$
$CF_3H_{(liq,\ 191\ K)}$

CHAPTER FOUR

CALCULATIONS OF THE RELATIVE FREE ENERGIES OF AQUEOUS SOLVATION OF SEVERAL FLUORINATED DERIVATIVES OF METHANE, AND OF ETHANOL AND ETHANE: A TEST OF THE BOND-POTENTIAL OF MEAN FORCE CORRECTION

INTRODUCTION

The Free Energy Perturbation (FEP) method has seen extensive use in the last few years for the prediction of the relative free energy of desolvation and binding of inhibitors to enzyme active sites. However, the FEP technology has not yet advanced to the state of being automatically reliable for this purpose, and new modifications to the methodology are still being discovered to be necessary for the accurate prediction of free energy changes due to subtle modifications of the compounds in question[1].

Many of the drug and inhibitor calculations performed on complex systems can not easily be used to discover and remedy any remaining defects in FEP methodology and parameterization for molecular dynamics calculations. This is because it is often difficult to separate problems of methodology and parameterization from sampling problems, which may also lead to inaccurate free energy results. For this reason, it is essential to perform test case simulations involving as few degrees of freedom as possible, but that still are representative of a certain class of compounds. In this paper, we explore the calculation of the free energies of aqueous solvation of a series of fluorocarbons, with the express purposes of testing the suitability of our fluorine parameters for hydration free energy calculations, and of testing the ability of the FEP method to accurately calculate subtle changes between two hydrophobic compounds. These changes involve the mutation of hydogen atoms into fluorines, and vice versa, and are among the smallest changes possible.

Another reason to use fluorine-to-hydrogen substitutions as a test case for the FEP method is the importance of such substitutions for drug design. Fluorine is almost isosteric with hydrogen, and therefore substitution of fluorine for hydrogen can be used to

control the electronic properties of a molecule[2] or to replace a labile carbon-hydrogen bond with a stable carbon-fluorine bond[3].

The replacement of a hydrogen by a fluorine has some additional consequences that may be useful in drug design. The solubility in water of a given perfluorocarbon is general lower than that of the hydrocarbon from which it is derived[4,5], while the solubility of organic compounds containing both hydrogen and fluorine decreases with the number of fluorines[4]. A positive change in heat capacity ΔC_P is observed upon dissolution of fluoroalcohols in water, which is in general about twice as large in magnitude as the ΔC_P accompanying dissolution of the corresponding hydrocarbon alcohols[6]. This positive ΔC_P is characteristic of an increase in hydrophobicity, as defined by several authors[7], and suggests that a given fluorinated compound is more hydrophobic than the corresponding molecule in which hydrogens take the place of fluorines. The solubility of a given fluoroalcohol in water is also less than that of the corresponding hydrocarbon alcohol.

The significance of this increased hydrophobicity of fluorocarbons compared to hydrocarbons is that replacing several hydrogens with fluorines may decrease the water solubility of a given enzyme inhibitor. Thus, the free energy for desolvation of an inhibitor and binding to a protein could be made more negative by, for example, replacing a methyl group with a trifluoromethyl group, if it was known that the trifluoromethyl group would be bound in a region of favorable electrostatic potential gradient in the protein. The electrostatic interactions between the trifluoromethyl group and the protein, and between the trifluoromethyl group and solvating water, could be comparable; but the solvated inhibitor might be destabilized by the decrease in entropy of the solvated waters (the "hydrophobicity" of the trifluoromethyl group), thus providing an

additional driving force for ligand-protein association.

It is a challenging test of the FEP method to accurately predict the subtle changes in water solubility that are found in comparing hydrocarbons and fluorocarbons.

There are several issues of parameterization and methodology to be investigated by these FEP calculations. One is whether the electrostatic interactions between fluorocarbons and water can be accurately simulated. A reasonable set of quantum-mechanically derived charges commonly involves using a derivation of these charges based on electrostatic potential calculations at the *ab initio* level with a 6-31G* basis set. This basis set leads to a wave function and charges that overestimate the gas-phase dipole moment of the molecule in question. However, these charges are suitable for calculations on molecules in aqueous solution, since one expects the dipole on the molecule to be enhanced by polarization from water molecules[8]. Thus, the 6-31G* charges provide a good representation of the dipole moment in aqueous solution of molecules containing only carbon, oxygen, nitrogen, and hydrogen. Since fluorine is less polarizable than oxygen or nitrogen, it is possible that the 6-31G* charges would thus overestimate the solution dipole moment of polar fluorocarbons, and give electrostatic contributions to the relative free energies of solution that were too large. In particular, if the 6-31G* dipole moment were too large, one might expect the magnitude of the free energy change to be too great when perturbing from a non-polar methane derivative (methane or tetrafluoromethane) to one of the polar fluorinated derivatives.

We also wish to test the usefulness of our recently derived parameters for fluorine with simulations of aqueous solubility. Recently, van der Waals parameters for fluorine, to be used in the molecular dynamics simulation of fluorine-containing compounds, have been derived using molecular dynamics simulations of pure tetrafluoromethane[9].

In addition, this paper will demonstrate the importance of a recent development in the free energy perturbation methodology used in AMBER[10]: the bond potential of mean force (bond PMF) correction. This correction is applied when perturbing between two species with different bond lengths. This bond PMF method was described and tested in a recent paper[1], In this earlier paper, it was clearly shown that the bond PMF is very important when perturbing between two hydrophobic species which have identical topologies; i.e., they have the same number of atoms.

The FEP calculations presented in this paper clearly reinforce the conclusion of reference 1 that it is necessary to use the bond PMF correction when perturbed bonds whose lengths are changing are constrained with the SHAKE algorithm. The particular case of interconverting the nonpolar hydrophobic solutes methane and tetrafluoromethane gives a very clear demonstration of this precisely because the number of atoms does not change during the simulation; rather, hydrogens are mutated into fluorines, and electrostatics do not complicate the picture. In addition, the perturbation is small enough such that adequate sampling of conformational space should be obtained within practical simulation times.

We also present in this paper an expanded explanation of just why the bond-PMF method works, and why the old method of calculating free energy differences between hydrophobic species is inadequate in cases involving changes in length of constrained bonds.

In addition, the importance of the bond PMF method for FEP calculations involving the contribution of changing bond lengths to a change in the dipole moment of the molecule in question, and thus to the electrostatic component of the free energy, will be discussed.

Perturbations in which the total number of atoms does not change, but in which there are significant changes in bond lengths, such as the methane to tetrafluoromethane transformation, are especially relevant to the bond-PMF methodology. It was shown in reference 1 that the bond-PMF method is especially important when bond lengths to disappearing atoms are being shrunk during a perturbation, but that if an atom is mutated to nothing without the shrinking of the bond, reasonable results can be obtained without a bond-PMF calculation. However, simulations of the CH_4 to CF_4 type cannot be performed without a change in bond length, and therefore the bond-PMF calculation must be performed.

METHODS

The energy minimizations and molecular dynamics trajectories were calculated with the program AMBER[10], which uses a potential energy function of the following form:

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

All of the parameters needed are found in the Weiner *et al.* force field[11,12] except for parameters involving fluorine, hydrocarbon carbon and hydrogen van der Waals parameters[13], and the point charges of the solute molecules.

The point charges of the molecules used were calculated using the electrostatic potential fitting routine developed by Singh & Kollman[14], in which point charges on the atoms are adjusted to give an electrostatic potential that fits that calculated quantum mechanically. The charges of each methane derivative were calculated via a fixed-point calculation, using the 6-31G* basis set and the experimentally determined geometry[15,16]. The calculated charges are shown in Table 4.1.

The standard AMBER parameter files do not contain parameters for fluorine, and it was therefore necessary to derive them or obtain them them from other sources. The force constants and equilibrium bond angles for bond stretching and bending potential energy terms were taken from the parameter list for MM2[17,18] and used without modification, after being converted into the slightly different energy expressions used in AMBER; these parameters are listed in Table 4.2. The equilibrium bond distances used were the experimentally measured gas phase bond distances[16], with the exception of tetrafluoromethane, where a bond length of 1.38 Å was used, since the fluorine parameters were derived from simulations of pure CF_4 using a bond length of 1.38 Å[9].

The van der Waals 6-12 parameters for fluorine were derived from simulations of pure tetrafluoromethane[9], in which R* and ε of fluorine were adjusted to duplicate the experimental molar volume and enthalpy of vaporization. The van der Waals radius used for the hydrogen attached to a fluorinated methane derivative is smaller than that usually used for a hydrocarbon hydrogen, in keeping with the results of Gough *et al.*[9], Veenstra *et al.*[19], and DeBolt and Kollman (private communication), which show that a hydrogen bonded to a carbon bearing one or more electronegative atoms requires a smaller van der Waals radius if molecular dynamics and molecular mechanics calculations are to accurately simulate experiment and *ab initio* quantum-mechanical

calculations. The 6-12 parameters used for fluorine, tetrahedral carbon, and hydrogen attached to a fluorine-bearing carbon are given in Table 4.3.

The difference in the free energy of solvation between two solutes can be calculated using the thermodynamic cycle of Figure 4.1, rather than by "growing in" the entire solute molecule in the solvent. That is,

$$\Delta\Delta G = \Delta G_{sol,B} - \Delta G_{sol,A}$$

$$= \Delta G_{aq} - \Delta G_{gas}$$

ΔG_{aq} and ΔG_{gas} are easily calculable by leaving the inhibitor in the gas or solution environment and perturbing the potential energy function. For the calculations described in this paper, no intramolecular energy terms are included in the calculation of the free energy, and thus ΔG_{gas} is assumed to be zero, or, more precisely, $\Delta G_{intermolecular}$ is assumed to be identical in the gas phase and solution.

The free energy calculations were carried out using the "windowing" method described previously[20]. The two states between which we wish to calculate a free energy are defined as the λ=1 and λ=0 states. For example, in the perturbation of fluoromethane to methane, fluoromethane and methane are the λ=1 and λ=0 states, respectively. Each of these states has appropriate molecular mechanical parameters, such as atomic partial charges and van der Waals radii, assigned to it. In addition, when using the bond potential of mean force (bond-PMF) correction, each state has a fixed bond length characteristic of the value of λ. The free energy calculation involves performing molecular dynamics on the system and changing λ in a series of steps of size $\delta\lambda$, from one extreme value to the other.

At each intermediate value of λ, or "window", λ_i, the dynamics trajectory evolves under the influence of a Hamiltonian that contains a λ-dependent combination of the

molecular mechanical parameters of the two extreme states; for a given parameter P, having value P^A and P^B corresponding to the two extreme states A and B:

$$P_{\lambda_i} = \lambda P^A + \left[1 - \lambda\right] P^B$$

The free energy for each intermediate window is calculated using the equation:

$$\Delta G = -RT \ln \langle exp\,[-(H_{\lambda_i + \delta\lambda} - H_{\lambda_i})/RT] \rangle_{\lambda_i}$$

where $\langle\rangle_{\lambda_i}$ indicates an equilibrium average at $\lambda = \lambda_i$. At each window, both a "forward" ΔG and a "backward" ΔG can be calculated, corresponding to a negative or positive $\delta\lambda$ in the equation given above. A comparison of the sums of these forward and backward calculations gives a crude measure of the uncertainty inherent in the calculation (although errors estimated in this manner are certainly underestimates of the true statistical uncertainty[21]).

Each ΔG_{aq} was calculated twice, once by perturbing from the λ=1 state to the λ=0 state; and once by perturbing from the λ=0 state back to the λ=1 state, starting with the results of a 5 ps equilibration from the coordinates from the previous perturbation. It was thus possibly to have a check on the uncertainty involved in the calculation by using the hysteresis between these two calculations as a standard deviation.

The molecular dynamics trajectory from which the free energy was calculated was run using a time step of one femtosecond. Each window, as described above, consisted of a 500 fs equilibration period; and a 500 fs data collection period, in which the actual calculation of the ensemble average, used in determining ΔG, took place. A free

energy perturbation in one direction took 40 ps. For the CF_4 to CH_4 mutation, we also ran a second set of simulations in which the perturbation in each direction took 80 ps.

All simulations were performed at 298 K and one atmosphere, using the algorithms of Berendsen[22] for maintaining constant temperature and pressure. A time step of 1 femtosecond was used. The calculations were performed using periodic boundary conditions[23]. The box size was approximately 21 Å on a side, and contained from 285 to 324 water molecules, depending on the size of the solute. The initial water coordinates were obtained from a snapshot of a Monte Carlo simulation of water. The TIP3P[24] water potential was used. This collection of water molecules was placed in the desired location around the solvent, and any water molecule with either an oxygen within 2.4 Å, or a hydrogen within 1.4 Å, of a solute atom was removed from the structure.

All calculations were carried out using a cutoff distance of 8 Å for the non-bonded interactions.

RESULTS AND DISCUSSION

The calculated and experimental free energies are given in Table 4.4. In general, the free energies were reproduced quite well, and the hysteresis between the "forward" and "backward" simulations is quite small. The bond-PMF methodology improved the agreement of the calculated free energy with the experimental free energy in all cases.

In the case of the perturbations involving at least one species with a non-zero dipole moment, the bond-PMF method improves the calculated free energy, but the results are qualitatively reasonable even without the bond-PMF calculation.

In all but one of the perturbations, the ΔG is underestimated if the bond-PMF method is not used. As will be discussed below, this is likely due to an underestimate of the increase in free energy of hydration due to the larger size of the species containing the greater number of fluorines.

This size effect seems to be less important in the case of the fluoromethane to methane perturbation. One explanation for this is that water-fluorine hydrogen bonding is more important for fluoromethane than for the other species; this is consistent with our ongoing analysis of the structure of the waters solvating fluorocarbons[25].

The bond-PMF is absolutely necessary to get the right sign for the CF_4 to CH_4 perturbation. This can be understood by examining the importance of entropy in the free energies of hydration of the two solutes. The negative entropy of hydration of a nonpolar solute is believed to be due to a reduction of the accessible volume of conformational space available to the solvating waters: in order for each water molecule to form four good hydrogen bonds with its neighbors, the water must give up some of its rotational freedom, since it is impossible for it to form hydrogen bonds in the direction of space occupied by the nonpolar solute[26]. The negative entropy change upon hydration will be greater in magnitude for larger solutes, since a solute of larger surface area will force more water molecules into giving up some of their rotational freedom in order to retain four hydrogen bonds. The negative entropy of hydration is often referred to as the "cavity" term in the free energy of hydration, as defined within the formalism of scaled particle theory[27,28].

It is not feasible to accurately calculate the ΔS of hydration using molecular dynamics; this has been attempted but results in very large standard deviations[29]. However, the negative entropies of hydration will be included within our calculated free

energies.

The increase in cavity size upon perturbing methane into tetrafluoromethane is due to two changes accompanying the transformation of hydrogens into fluorines: the increase in volume due to the larger van der Waals radius of fluorine, and the increase in volume due to the C-F bond being longer than the C-H bond. This latter contribution is neglected entirely if the FEP calculation is performed using SHAKE without performing the PMF-type calculation.

The size of the hydrophobic cavity formed by CF_4 and CH_4 depends upon both the van der Waals radius of the constituent atoms (hydrogen versus fluorine), and the length of the bonds (C-H bonds versus C-F bonds). A C-H bond length is 1.09 Å, versus 1.38 for the C-F bonds in CF_4, so the bond length difference make a significant difference in the cavity size.

Since the free energy difference between CF_4 and CH_4 would be expected to be almost entirely due to the different sizes of this cavity and a difference in dispersion interaction, rather than having a significant electrostatic component, it is essential that this cavity size be included accurately in the FEP calculation.

To see why the bond-PMF contribution makes such a difference to the calculation of the difference in the free energy due to the difference in the cavity size, consider the equation used to calculate the free energy difference between two states A and B:

$$G_B - G_A = -RT \ln \langle exp\,[-(H_B - H_A)/RT] \rangle_A$$

where $\langle\rangle_A$ indicates an equilibrium average using a Hamiltonian characteristic of state A. If the bond lengths are fixed via the SHAKE algorithm, and no bond-PMF correction calculation is performed, then no bond term appears in the Hamiltonian of state A or

B^1.

To understand fully the importance of including the bond-PMF contribution, it is necessary to understand that:

$$<exp[-(H_B - H_a)/RT]>_A = <(P_B)/(P_A)>_A$$

where P_B and P_A are the unnormalized probability of the system characteristic of state B or A, respectively,being in a given region of phase space. That is,

$$P_A = P_A(\mathbf{r}), \text{ and } P_B = P_B(\mathbf{r})$$

where **r** is a point in configuration space.

These probabilities will be accurate only if the Hamiltonian accurately represents all of the factors that are affecting the phase space distribution of a given system; that is, only if the interatomic distances r_{ij} used in calculating H_A and H_B are truly representative of each of these states. If SHAKE is used to constrain the bond lengths, this is not true, as the bond lengths are not determined by a term in the Hamiltonian. If SHAKE is used, and the bond-PMF correction is not calculated, the Boltzmann factor $exp[-H_B]/RT$ does not accurately represent the probability of state B being in a given region of phase space, since neither H_A nor H_B contain any term relating to the bond length; the bond length is determined by the SHAKE algorithm rather than the Hamiltonian. The r_{ij}'s used in calculating H_B will not be those characteristic of the equilibrium bond length of state B, but rather will be characteristic of state A, unless some way of taking account of the correct equilibrium bond length of state B is provided. In fact, if only the bond lengths, and none of the parameters in the Hamiltonian, differed between states A and B, $exp[-H_B]/RT$ and $exp[-H_A]/RT$ would be *identically equal* if the bond-PMF

calculation was not done.

In the bond-PMF method[1], the calculation of H_B and H_A is performed using interatomic distances r_{ij} corrected to the appropriate lengths for states B and A, respectively. Thus, $exp[-H_B]/RT$ at a given point in a trajectory accurately describes the probability of being at that point in a phase space corresponding to the H_B.

If the bonds were not constrained with the SHAKE algorithm, the different bond lengths would show up as different r_{eq} values in the potential energy expression. But this may lead to the sampling problems discussed in reference 1.

It should be noted that when SHAKE is used, standard thermodynamic integration will not correctly calculate ΔG between two states with different bond lengths either. In thermodynamic integration, the free energy is calculated as an integral over λ, from 0 to 1, of $\langle \delta H/\delta\lambda \rangle$; if SHAKE is used, $\delta H/\delta\lambda$ will have no term representing the bond, and thus $\langle \delta H/\delta\lambda \rangle$ will not reflect the different bond lengths.

It appears that a significant part of the difference in the effective cavity size in the CF_4 to CH_4 perturbation is due to the difference in bond length rather than the change in van der Waals radius. The erroneous sign of the the free energy change (positive rather than negative when perturbing from CF_4 to CH_4 without performing the bond-PMF calculation) is likely due to the decrease in solute-solvent dispersion interactions when fluorines are mutated to hydrogens, without the accompanying negative free energy change that would accompany the correct inclusion of cavity size.

Even though the CF_4 to CH_4 calculations using the bond-PMF method give a qualitatively correct result, neither one gives a ΔG_{sol} that quantitatively reproduces the experimental value. The ΔG from the longer simulation is more reliable in terms of pre-

cision, as indicated by the very small hysteresis. However, what accounts for the approximately 0.8 kcal difference between this calculated value and the experimental value? Since electrostatics does not play a significant role in this ΔG_{sol}, the simulation should be very sensitive to the van der Waals parameters of methane and tetrafluoromethane. In particular, the van der Waals parameters of methane may need further refinement; this is currently being investigated by others in this group.

The free energy of perturbing methane to "nothing" in aqueous solution was determined by the simulations of Pearlman and Kollman[1] to be ~ -2.6 kcal/mole, as compared to the experimental value of ~ -1.9 kcal/mole. This may be indicative of a problem with the parameters used for methane; this error is in the direction that would account for methane being less stable in aqueous solution than it should be relative to tetrafluoromethane; if the parameters account for all of the 0.7 kcal/mole error, and the parameters used for tetrafluoromethane are optimal, one would expect a ΔG_{sol} of (-0.4 - 0.7) = -1.1 for the CF_4 to CH_4 perturbation, much closer to the experimental value of -1.2 kcal/mole.

In the case of the perturbation calculations in which at least one of the species has a non-zero dipole moment, the electrostatic terms seem to dominate such that ΔG is qualitatively correct even without the bond-PMF calculation. However, when the bond-PMF correction is not used, the solubility of the species containing the greater number of fluorines is consistently overestimated. This suggests that the previous discussion relating to cavity size has some bearing for these molecules also; the greater water-accessible surface area of the species with the greater number of fluorines is being underestimated when the bond-PMF calculation is not included.

However, the bond-PMF method should improve the calculation of electrostatic effects on the ΔG also, at least in cases where the perturbation involves changes in bond lengths between atoms with significant charges. Free energy contributions due to geometry are important not just for correctly calculating the entropy effect due to hydrophobic hydration, but will also show up in long-range electrostatic interactions, such as those between a polar solute and solvent. To see this, consider a dipole, consisting of two partial charges δ+ and δ– separated by a distance r_A, in a cavity (of low dielectric constant), and surrounded by a medium of higher dielectric constant. If the distance between the partial charges δ+ and δ– were to be increased, say, to r_B, without altering the value of the charges, a negative free energy change would occur; i.e., $G_B - G_A < 0$.

If, however, one attempted to do a FEP calculation on a physically similar system (say, a dipolar system with SHAKEn bond length surrounded by waters), and any change in the van der Waals interactions between the dipolar system and the environment was omitted from the calculation, $G_B - G_A$ would be calculated to be *exactly zero*, using the FEP equation, if the PMF correction was not included. To see this, recall that the charges are not changing, and thus if the PMF correction is omitted, E_B is identically equal to E_A at every step of the trajectory. When the PMF correction is used, however, the two calculations $E_B(r_{ij})$ and $E_A(r_{ij})$ involve different r_{ij}'s for each point in the trajectory of the water molecules. Here, electrostatic effects are reflected in the different regions of accessible phase space of the water molecules. Thus, the PMF correction is important not only for correctly calculating changes in cavity size, but is also necessary for the correct inclusion of electrostatic effects that are due to changes in the relative geometry of molecular charges (as distinct from changes in the values of the

charges themselves), if the changes in the geometry are due to SHAKEn bond lengths. In the vast majority of real molecular simulations, of course, both the charges and their relative geometry will be changing, but the above thought experiment suggests that there will be some error in a FEP calculation if there is any significant change in a SHAKEn geometry and the bond-PMF correction is not used.

It should be noted that in the case of perturbing, say, a hydroxyl group to a hydrogen atom in aqueous solution, the bond-PMF method may be important if one shrinks the oxygen-hydrogen bond during the perturbation, rather than just perturbing the charges and van der Waals parameters while leaving the oxygen-hydrogen bond length intact. This is true even though there is no great change in the size of the solute, as was the case both in the calculations discussed in reference 1 and in the tetrafluoromethane to methane perturbation discussed above. This is due to the fact that the loss of hydrogen bonding between the hydroxyl and the solvating waters, as calculated by the FEP perturbation expression, will depend on the change in geometry as well as the change in charges. Since hydrogen bonds can to a certain extent be considered dipole-dipole interactions, the previous discussion of the relevance of the bond-PMF correction to the calculation of the free energy of charging or uncharging a dipole is relevant.

Consider the case in which a hydroxyl is perturbed to a hydrogen. If the oxygen-hydrogen bond is shrunken to zero length or to a very small value during the perturbation, it will be necessary to use the bond-PMF method. To see this, note that the calculation involving shrinking could be done in two steps: first, the bond length is shrunken, and this contribution would be zero if SHAKE is used and the bond-PMF method is not used; second, the charges would be perturbed. But this latter contribution would be very small, as it would involve the disappearance of two charges very close

together, which would not create a good hydrogen-bonding geometry. So the ΔG would be expected to be smaller in magnitude than the correctly calculated value. Thus, the contribution of the bond-PMF method to perturbations involving large electrostatic changes may be just as important as its contribution to primarily steric perturbations.

To test this, we performed a perturbation from ethanol to ethane in aqueous solution using three different conditions. In all cases, the hydroxyl oxygen and hydrogen were perturbed to a hydrogen and a "dummy" atom, having zero charge and van der Waals parameters. In one simulation, the hydroxyl oxygen-hydrogen bond was kept at the initial value of 0.96 Å. In the other two simulations, this bond was shrunken to a final value of 0.2 Å, once including the bond-PMF contribution and once without including this contribution.

The results are striking and are shown in Table 4.5. When the bond is shrunken, ΔG is too small by about 3 kcal/mole if the bond-PMF contribution is not included. This is indeed the sign of the error expected from the discussion above; the magnitude of the ΔG is smaller than it should be. When the bond is not shrunken and the bond-PMF calculation is not used, ΔG is only slightly smaller than it should be, in all likelihood due to the error in not carrying out a PMF correction for the change in length of the C-O bond as it is perturbed to a C-H bond. The difference is small, likely because of the opposite signs of the van der Waals and electrostatic contributions to ΔG from the C-O to C-H perturbation. That is, the decrease in size of the system as a C-O bond is shrunken to a C-H bond makes a negative contribution to the free energy of hydration, while the electrostatic effect of this change makes a positive contribution. This type of compensation is apparently seen in in the fluoromethane to methane perturbation also; presumably this is the reason that the ΔG for that perturbation has approximately the

same value whether or not the bond-PMF correction is calculated. However, this compensation will not always be present, and therefore the bond-PMF correction should be calculated if bond lengths are changing.

The ethanol-to-ethane results relate to an earlier calculation of ours (Chapter Two of this dissertation), in which one hydroxyl of a *gem*-diol was perturbed to a hydrogen; this calculation gave a ΔG_{sol} that was of the correct sign, but several kcal/mol in magnitude smaller than the 7 kcal/mol expected for the disappearance of a hydroxyl group. Although this simulation involved a very large solute, the conformational space of which probably could not be sampled adequately during the simulation time, it now seems that at least part of the reason for the small magnitude of the ΔG_{sol} is that the O-H bond was shrunk during the simulation, and the bond-PMF contribution was not included. Since the large size of the system being simulated probably precludes convergence of the free energy within a reasonable simulation time, we did not repeat the calculation using the bond-PMF calculation. However, the results of the ethanol-to-ethane simulation suggest that even a calculation of sufficient length would give a ΔG whose magnitude would be significantly smaller than the experimental value unless the bond-PMF correction were included or the hydroxyl oxygen-hydrogen bond were not shrunken during the simulation.

CONCLUSIONS

This work has shown that the FEP method can be used to accurately estimate the changes in solubility when hydrogens are replaced by fluorines. In view of the small magnitude of this change, particularly in the perturbation between the non-polar species

CF_4 and CH_4, our success is encouraging for the use of the FEP method for drug design, in particular, for testing the novel use for substitution of hydrogen by fluorine discussed in the introduction of this paper. We have also demonstrated that our current electrostatic models deal with fluorocarbons quite well, and that our fluorine parameters, derived by simulations of pure CF_4, reproduce relative free energies of interaction with water quite well. In addition, we have shown the importance of the use of the bond-PMF method on these systems.

The discussion in this paper has also demonstrated that the free energy perturbation expression is not valid unless both the Hamiltonians of the systems at both endpoints of the perturbation are descriptive of *all* of the factors affecting the phase space of both systems. If this is not true, there must be some correction for this omission. In the case of the bond-PMF calculation, the correction is that the interatomic distances used in the calculation of the potential energy are calculated according to the fixed bond lengths characteristic of the two endpoints of the perturbation. This should be kept in mind in any future calculations in which constraints on a perturbed system are included in a FEP calculation, if the constrained geometry changes during the perturbation.

Also, we have made a prediction that the difference in free energy of hydration between difluoromethane and trifluoromethane is 1.0 ± 0.1 kcal/mole, with difluoromethane being more soluble. At the present time, the aqeuous solubility of difluoromethane has not been published.

REFERENCES FOR CHAPTER FOUR

(1) Pearlman, D.A.; Kollman, P.A. *J. Chem. Phys.*1991, **94**, 4532.

(2) The Pharmacological Basis of Therapeutics, A.G. Gilman; L.S. Goodman; T.W. Rall; F. Murad. MacMillan Publishing Company, New York. 1985.

(3) *ibid.*; pp. 1267.

(4) Hine, J.; Mookerjee, P.K.; *J. Org. Chem.*1984, **40**, 292.

(5) Wen, W.-Y.; Muccitelli, J.A. *J. Solution Chem.*1979, **8**, 225.

(6) Rochester, C.H.; Symonds, J.R. *J. Chem. Soc., Faraday Transactions* 1973, **1**, 1577.

(7) Dill, K.A. *Biochemistry*1990, **29**, 7133, and references therein.

(8) Kuyper, L.F.; Hunter, R.N.; Ashton, D; Merz, K.M., Jr.;Kollman, P.A.; *J. Phys. Chem.*1991, **95**, 6661.

(9) Gough, C.A.; DeBolt, S.E.; Kollman, P.A. Submitted to *J. Comp. Chem.*

(10) Pearlman, D.A.; Case, D.A.;Caldwell, J.C.; Seibel, G.L.; Singh, U.C.; Weiner, P.; Kollman, P.A. (1991), AMBER 4.0, University of California, San Francisco.

(11) Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghia, C.; Alagona, G.; Profeta, S.,Jr.; Weiner, P.; *J. Am. Chem. Soc.* 1984, **106**, 765.

(12) Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A.; *J. Comp. Chem.* 1986, **7**, 232.

(13) Sun, Y.; Spellmeyer, D.; Pearlman, D.A.; Kollman, P; *J. Am. Chem. Soc.*, in press.

(14) Singh, U.C.; Kollman, P.A. *J. Comp. Chem.* 1984, **5**, 129.

(15) Harmony, M.D.; Laurie, V.W.; Kuczkowski. R.H.; Schwendeman, R.H.; Ramsay, D.A.; Lovas, F.J.; Lafferty, W.J.; Maki, A.G.; *J. Phys. Chem. Ref. Data* 1979 **8**, 619.

(16) Baird, N.C. *Can. J. Chem.*1983, **61**, 1567.

(17) MM2: Molecular Mechanics II. Quantum Chemistry Program Exchange, Program No. 423.

(18) Meyer, A.Y.; Allinger, N.L.; *Tetrahedron* 1975 **31**, 1971.

(19) Veenstra,D.; Ferguson,D.;Kollman, P.A. Submitted to *J. Comp. Chem.*

(20) Singh, U.C.; Brown, F.K.; Bash, P.A.; Kollman, P.A. *J. Am. Chem. Soc.* 1987, *109*, 1607.

(21) Pearlman, D.A.; Kollman, P.A. *J. Chem. Phys.*1989, **90**, 2460.

(22) Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, A.; Haak, J.R. *J. Chem. Phys.*1984, **81**, 3684.

(23) Computer Simulation of Liquids. M.P. Allen and D.J. Tildesley. Oxford Science Publications, Oxford, UK, 1987.

(24) Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L.; *J. Chem. Phys.* 1983, **79**, 926.

(25) Gough, C.A.; Kollman, P.A. Work in progress.

(26) Rossky, P.J.; Karplus, M. *J. Am. Chem. Soc.*1979, **101**, 1913, and references therein.

(27) Reiss, H.*Adv. Chem. Phys.*1965, **9**, 1.

(28) Pierotti, R. A. *J. Phys. Chem.* 1965, **69**, 281.

(29) Fleischman, S.H.; Brooks, C.L., III. *J. Chem. Phys.* 1987, **87**, 3029.

Table 4.1

Atomic Point Charges of Methane and Fluorinated Derivatives (in fractions of an electron)			
Molecule	Charge on Carbon	Charge on Fluorine	Charge on Hydrogen
CF_4	0.756	-0.189	-
CHF_3	0.644	-0.230	0.046
CH_2F_2	0.390	-0.246	0.051
CFH_3	0.011	-0.251	0.080
CH_4	-0.484	-	0.121

Table 4.2

Bond Stretching and Bending Parameters for Bonds Involving Fluorine		
Bond	K_R (in kcal mol^{-1} $Å^{-2}$)	R_o (in Å)
CT - F of CF_4	367	1.38
CT - F of CHF_3	367	1.33
CT - F of CH_2F_2	367	1.357
CT - F of CFH_3	367	1.38
Angle	K_θ (in kcal mol^{-1} rad^{-2})	θ_o (in degrees)
F-CT-HC	35.0	109.5
F-CT-F	77.0	109.1

Table 4.3

6-12 Parameters Used for Tetrahedral Carbon, Fluorine, and Hydrogen		
Atom Type	R* (in Å)	ε (in kcal/mol)
CT	1.91	0.1094
F	1.75	0.061
HC of Methane	1.49	0.015
HC of Fluoromethane and Difluoromethane	1.29	0.015
HC of Trifluoromethane	1.21	0.015

Table 4.4

Relative Free Energies (in kcal/mole) of Aqueous Hydration of Fluorinated Methane Derivatives

Perturbation	Bond-PMF Calculated?	ΔG^{a}_{calc}	$\Delta G^{b}_{experimental}$
CF_4 to CH_4	no	0.77 ±0.06	-1.2
CF_4 to CH_4	yes	-0.8 ±0.3	-1.2
CF_4 to CH_4^c	yes	-0.44 ±0.01	-1.2
CFH_3 to CHF_3	no	0.3 ±0.3	1.0
CFH_3 to CHF_3	yes	0.7 ±0.1	1.0
CFH_3 to CH_4	no	2.4 ±0.3	2.2
CFH_3 to CH_4	yes	2.3 ±0.2	2.2
CHF_3 to CH_2F_2	no	-0.71 ±0.07	-
CHF_3 to CH_2F_2	yes	-1.0 ±0.1	-
CHF_3 to CF_4	no	1.5 ±0.1	2.4
CHF_3 to CF_4	yes	2.19 ±0.05	2.4

(a) The value reported is the average value in the "forward" direction, as given in the first column. Each value represents the average of a forward (0 to 1) and a reverse (1 to 0) simulation. Each direction involved a 40 psec simulation, unless otherwise noted. The error bars represent the difference between the calculations in the two directions.

(b) From reference 4.

(c) This calculation was run over 80 psec.

Table 4.5

Free Energy Difference (in kcal/mole) between Ethanol and Ethane in Aqueous Solution[a]

Hydroxyl O-H Bond Shrunk?	Bond-PMF Calculated?	ΔG^{b}_{calc}
no	no	6.6 ±0.1
yes	no	4.1 ±0.2
yes	yes	7.1 ±0.1

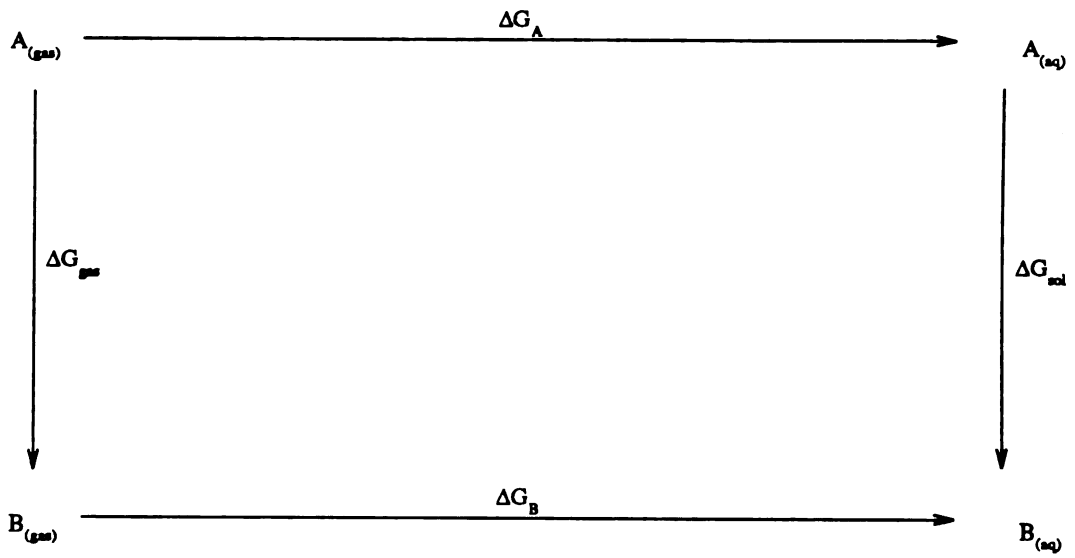
(a) The experimental value is 7.0 kcal/mole.

(b) The value reported is the average value in the "forward" direction (ethanol to ethane) between a "forward" and a "backward" calculation. The error bars represent the difference between these two calculations.

Figure 4.1

Thermodynamic cycle used for the calculation of difference in solvation free energies between two species A and B.

Figure 4.1



CHAPTER FIVE

A MOLECULAR DYNAMICS ANALYSIS OF THE STRUCTURE OF WATER SOLVATING FLUOROMETHANE, TRIFLUOROMETHANE, AND TETRAFLUOROMETHANE

INTRODUCTION

Fluorocarbons display interesting solubility properties that are not yet fully understood. The effect of replacing one or more hydrogens of a hydrocarbon with fluorines depends upon the number of these replacements[1].

Table 5.1 shows the solubility data for several halogenated derivatives of methane. It can be seen that the replacement of one hydrogen by a fluorine increases the water solubility by 2.2 kcal/mole, but three such substitutions increases the solubility by only 1.2 kcal/mole, and replacing all of the hydrogens with fluorines actually results in a decrease in solubility of 1.2 kcal/mole. Replacing hydrogens with chlorines produces a different trend: trichloromethane is *more* soluble than chloromethane, and tetrachloromethane is more soluble than methane, though less soluble than trichloromethane. Replacing hydrogens with bromines gives rise to a similar trend; the more hydrogens replaced with bromines, the more soluble the species is, and the increased solubility of tribromomethane relative to bromomethane is even greater than that of trichloromethane relative to chloromethane.

Why do the fluorinated derivatives of methane show such a different trend from the other halogenated derivatives? These trends can be understood if one recalls that fluorine is considerably less polarizable than the other halogens. Though the fluorinated derivatives of methane are more polarizable than methane itself[2], the increased dispersion interaction with water resulting from substitution of hydrogens with fluorines does not compensate for the increase in the size of the hydrophobic "cavity" caused by the increase in van der Waals radii and bond lengths. Bromine and chlorine, on the other hand, are more polarizable, leading to relatively greater dispersion interaction with water, relative to fluorine. This polarizability also leads to the possibility of enhanced

electrostatic interactions with water, due to polarization by the solvating waters.

It thus seems that some fluorocarbons, though polar, have some characteristics of hydrophobic compounds, and stand out among halogenated hydrocarbons in this respect. While the exact molecular mechanism of the hydration of hydrophobic solutes is subtle[3] computer simulations of hydrophobic solutes in aqueous solution[4], suggest that the negative entropy of hydration associated with hydrophobicity is due to a restriction of the conformational space of the waters adjacent to a hydrophobic solute. That is, the properties of the waters adjacent to a hydrophobic solute are due to an absence of a hydrogen-bonding moiety in the region occupied by the hydrophobic cavity. This absence restricts the librational freedom of the solvating waters, as they are more limited, relative to the waters in bulk water, in the number of conformations they can adopt while still maintaining four good hydrogen bonds with neighboring waters. A larger cavity restricts the conformational freedom of more waters. A solute such as methanol, on the other hand, has a hydroxyl group that can form three good hydrogen bonds with waters (Monte Carlo simulations of methanol in water show three hydrogen bonds on average[5]), and therefore the waters adjacent to the hydroxyl do not have their conformational mobility restricted to the extent that they would be by a hydrophobic solute of approximately the same size, such as ethane.

The structure of the waters solvating fluorocarbons has not been determined experimentally. Crystallographic studies of clathrate hydrates show waters in cage-like arrangements around the fluorinated derivatives, similar in structure to the waters in the clathrate hydrate of methane[6]. In these structures, each water molecule is hydrogen-bonded to four nearest-neighbor water molecules, with the O-O-O angles deviating only 3 or 4 degrees from the tetrahedral structures found in ice. No hydrogen bonding of

waters to fluorines is seen. However, these results are for a solid phase and may not be relevant to liquid solutions of fluorocarbons.

One of the aims of this paper is to determine if the water structure around hydrated fluorocarbons during molecular dynamics simulations shows the same features observed in molecular dyanamics simulations of the waters around moieties commonly considered to be hydrophobic, such as methane and methyl groups[5]. The question can be asked: What is the effect of fluorocarbons on the structure of the solvating waters? Are there specific, directional water-fluorine interactions that stabilize the more soluble of the fluorocarbons, or more general dipole-water interactions, and how do these interactions depend upon the number of fluorines attached to a given carbon atom?

The low solubility of tetrafluoromethane is not surprising, as it lacks a net dipole moment. However, the dipole moment of trifluoromethane and fluoromethane are very similar, and are similar to the dipole moment of methanol[7] (see Table 2 for dipole moments). This suggests that if there are specific water-fluorine interactions such as hydrogen bonding, they do not substitute for water-water hydrogen bonds to the same extent as hydrogen bonds between water and a hydroxyl group. The data also suggests that there may be a trend in the type of solute-water interactions as more fluorines are added; the lower solubility of trifluoromethane relative to fluoromethane suggests that the number or the "quality", or both, of the water-fluorine hydrogen bonds is decreased as the number of fluorines bonded to a carbon is increased from one to three.

Our purpose in this paper is to examine the hydrogen-bonding, if any, between fluorocarbons and water, using the results of molecular dynamics simulations. This is the key to determining if the similarities in the hydration thermodynamics of fluorocarbons and hydrocarbons can be explained in terms of the absence, in the region of the

solute, of any hydrogen-bonding moiety. If indeed there is no hydrogen-bonding to some or all of the fluorinated derivatives of methane, the properties of water near these species should be similar, in a manner typical of decreased librational mobility of the water, to the water surrounding methane and other hydrophobic species.

We wish to use trifluoromethane as a test case for the trifluoromethyl group in general, to study the "hydrophobicity" of a dipolar moiety with three fluorines attached to a single carbon. The case of trifluoromethane and the trifluoromethyl group has practical interest, for if the structure of the solvating waters suggests that a trifluoromethyl group is indeed characteristic of waters solvating a hydrophobic solute, this group is thus both polar and hydrophobic. This can be made use of in the design of enzyme inhibitors, since adding a trifluoromethyl group to an enzyme inhibitor may decrease its affinity for water due to the decreased entropy of the water, while the electrostatic interactions of the group with an enzyme active site may be at least as strong as its electrostatic interactions with water, if there are no specific hydrogen-bonding type interactions with water molecules.

We have chosen to examine the fluorinated derivatives of methane for which experimental relative free energies of hydration have been reproduced by our free energy calculations[8], in order that we can be confident that our force field parameters for fluorine accurately represent interactions with water.

METHODS

All simulations were performed with AMBER 4.0[9], which uses a potential of the form:

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

All of the parameters needed are found in the Weiner *et al.* force field[10,11] except for parameters involving fluorine, and the point charges of the solute molecules.

The point charges of the molecules used were calculated using the electrostatic potential fitting routine developed by Singh & Kollman[12], in which point charges on the atoms are adjusted to give an electrostatic potential that fits that calculated quantum mechanically. The charges of each methane derivative were calculated via a fixed-point calculation, using the 6-31G* basis set and the experimentally determined geometry[13,14]. The calculated charges are shown in Table 3.

The bond-stretching and bond-bending parameters, shown in Table 4, were taken from the molecular mechanics program MM2[15,16]. The van der Waals parameters for fluorine, shown in Table 5 were derived, as described elsewhere[17], by performing molecular dynamics simulations on pure tetrafluoromethane and adjusting these parameters until the calculated molar volume and enthalpy of vaporization agreed with the experimental values within 5%. The van der Waals parameters for tetrahedral carbon were those derived from Monte Carlo simulations of liquid methane[18]. The van der Waals parameters for the hydrogens attached to the fluorocarbons are those used for the hydrogens attached to methanol in the pure liquid simulations of DeBolt and Kollman (private communication). The van der Waals radius of these hydrogens must be smaller because of the electron-withdrawing nature of oxygen and fluorine.

The calculations were performed using the standard method of periodic boundary conditions[19]. A constant temperature of 298 K and a constant pressure of 1 atmosphere were maintained using the temperature- and pressure-coupling algorithms of Berendsen *et al.*[20].

Each periodic box contained one solute molecule and between 285 and 305 waters. The boxes were prepared by inserting the solute among a collection of waters equilibrated by the Metropolis-Monte Carlo method[19], and then removing all waters not having at least one atom within 10 Å of a solute atom. Also removed were all waters whose oxygen was within 2.4 Å of a solute atom, or which had a hydrogen within 1.4 Å of a solute atom. These conditions resulted in periodic boxes approximately 21 Å on a side.

Each periodic box was minimized to a gradient of 0.1 $kcal\ mol^{-1}Å^{-1}$, with pairlist regeneration every 40 steps, and a nonbonded cutoff of 8.0 Å.

Before any structural analysis was begun, each solute-solvent system was equilibrated by 5 picoseconds of dynamics. For the purposes of the structural analyses, 20 ps of dynamics were run on each of the solutes in a box of water. Coordinates were saved each 20 fs. Each radial distribution function and angle distribution was thus an average of 1000 snapshots of the trajectory. In all of the molecular dynamics calculations, the nonbonded pairlist, including only those nonbonded interactions within 8 Å, was updated every 40 dynamics steps.

The free energy calculations were carried out using the "windowing" method described previously[21]. The two states between which we wish to calculate a free energy are defined as the λ=1 and λ=0 states. For example, in the perturbation of fluoromethane to methane, fluoromethane and methane are the λ=1 and λ=0 states,

respectively. Each of these states has appropriate molecular mechanical parameters, such as atomic partial charges and Van der Waals radii, assigned to it. The free energy calculation involves performing molecular dynamics on the system and changing λ in a series of steps of size $\delta\lambda$, from one extreme value to the other.

At each intermediate value of λ, or "window", λ_i, the dynamics trajectory evolves under the influence of a Hamiltonian that contains a λ-dependent combination of the molecular mechanical parameters of the two extreme states; for a given parameter P, having value P^A and P^B corresponding to the two extreme states A and B:

$$P_{\lambda_i} = \lambda P^A + \left[1 - \lambda\right] P^B$$

The free energy for each intermediate window is calculated using the equation:

$$\Delta G = -RT \ln \langle exp\,[-(H_{\lambda_i + \delta\lambda} - H_{\lambda_i})/RT] \rangle_{\lambda_i}$$

where $\langle\rangle_{\lambda_i}$ indicates an equilibrium average at $\lambda = \lambda_i$.

Each ΔG_{aq} was calculated twice, once by perturbing from the λ=1 state to the λ=0 state; and once by perturbing from the λ=0 state back to the λ=1 state, starting with the results of a 5 ps equilibration from the coordinates from the previous perturbation. It was thus possibly to have a check on the uncertainty involved in the calculation by using the hysteresis between these two calculations as a standard deviation.

RESULTS AND DISCUSSION

Figure 1 shows the fluoromethane carbon-water oxygen radial distribution function. This shows the first hydration shell centered at about 3.8 Å.

Figure 2 shows the fluoromethane F-water hydrogen radial distriubtion function. Although the main peak is centered at about 4.5 Å, there appears to be a shoulder to this peak, which contains waters with a hydrogen within about 2.5 Å. This suggests that some water hydrogens are hydrogen-bonded to the fluorine of fluoromethane. To examine this possibility, the distribution of F-H-O angle θ, as defined in the Methods section, within Figure 4 shows the distribution of the cosine of θ, as defined in Figure 3, for all waters having a hydrogen within 2.5 Å of the fluorine. A large, fairly sharp peak shows that θ is centered at approximately 180 degrees, corresponding to the H-O bond vector pointing directly to the fluorine. There is some spread in the angles, but it is clear that the tendency is for the angles to be distributed about a linear F-H-O orientation. The other, broader peak corresponds to the non-hydrogen-bonded hydrogen of a hydrogen-bonded water. The integrated area under this distribution corresponds to 1.1 waters; there is therefore, on average, one of these hydrogen bonds per fluoromethane molecule. A value of θ of 180 degrees allows the waters to adopt geometries such as those shown in Figure 5(a), 5(b), and 5(c), but not the geometry shown in Figure 5(d).

The possible orientations of the waters in the vicinity of the fluorine can be categorized more exactly. Figure 7 shows the distribution of C-H-O angles, as defined in Figure 6, for the waters that have at least one hydrogen within 2.5 Å of the fluorine of fluoromethane. Again, this angle distribution is centered on 180 degrees; i.e., the carbon, hydrogen, and oxygen tend to be in a straight line. This limits the configurations possible for the waters; a C-H-O angle of 180 degrees means that the configuration in Figure 5(b) is possible, but not the configurations in Figures 5(a) or 5(c). In other

words, the distribution of water configurations is centered on an approximately linear hydrogen bonding configuration, such that the carbon, fluorine, hydrogen, and oxygen are approximately in a straight line. When a fluoromethane-water dimer was energy-minimized in this configuration, the resulting molecular mechanical energy was -3.40 kcal/mole.

Figure 8 shows the cos θ distribution for waters within the main peak of the F-H radial distribution function, between 2.5 and 5.8 Å. In contrast to the corresponding distribution within 2.5 Å, this distribution shows a sharp peak centered at 0 degrees, and a broad peak centered at approximately The O-H bond vector here does not point at the fluorine. This distribution allows waters to adopt a geometry such as that shown in Figure 9(a), but not a geometry such as that shown in Figure 9(b). Thus, the water hydrogens within 2.5 Å are oriented significantly differently from those within the main peak of the F-H radial distribution function.

The fluorine-water hydrogen radial distribution function of trifluoromethane is shown in Figure 10. It differs from the equivalent radial distribution for fluoromethane in that there does not appear to be a shoulder indicating water-fluorine hydrogen bonding.

This lack of hydrogen bonding is confirmed by the distribution of F-H-O angles as defined for fluoromethane above. Figure 11 shows the distribution of F-H-O angles for all waters having at least one hydrogen within 2.5 Å of a fluorine. The integrated area under the curve corresponds to 0.41 water within this distance. No clear trend can be seen in the distribution. There does not appear to be any hydrogen bonding of the type exhibited in the fluoromethane case, since there is not obvious peak corresponding to an angle of 180 degrees.

Figure 12 shows the same angle distribution for those waters having at least one hydrogen in the region between 2.5 and 4.8 Å from a fluorine. The angular distribution here resembles that of fluoromethane in the region between 2.5 and 5.8 Å, and the possible geometries are the same. That is, there are no water-fluorine hydrogen bonds of the type shown in Figure 5(b).

The F-H-O angle distribution for waters within 2.5 Å would not be inconsistent with a trifurcated water-fluorine interaction as shown in Figure 13, in which an O-H bond of a water molecule is pointed approximately toward the carbon of trifluoromethane and is parallel to the carbon-hydrogen bond. In this type of hydrogen bond, the water hydrogen should be approximately equidistant from all three fluorines. To examine this possibility, we examined the cos θ distribution of the angles between the vector connecting the hydrogen and carbon of trifluoromethane, and the vector connecting the hydrogen and oxygen of any water with at least one hydrogen within a given distance of all three fluorines. The angle θ, as defined in Figure 14, would thus be 180 degrees in the case of the trifurcated hydrogen bond shown in Figure 13.

Several distances between 2.5 and 4.0 Å were used. When 4.0Å was used as a cutoff distance, an average of 1.0 water was found with a hydrogen within the cutoff, and the angle distribution, given in Figure 15, shows no preference for a trifurcated bond. The angle distribution is more or less random, with perhaps a slight tendency towards 90 degrees. 0.0 waters were found with a hydrogen within a cutoff of 2.5 Å, while using a cutoff of 3.0 and 3.5 Å showed only 0.02 and 0.18 water within the cutoff, not enough to suggest any type of a hydrogen bond.

To investigate whether this difference in hydrogen bonding is the source of the approximately 1 kcal/mole difference in solubility between fluoromethane and

trifluoromethane, we performed a free energy perturbation (FEP) calculation in which the charges of both hydrated fluoromethane and hydrated trifluoromethane are perturbed to zero, all other force field parameters remaining unchanged; this should give a qualitative idea of the relative importance of electrostatic interactions that make favorable contributions to the aqueous solubility of the two species.

Table 6 shows the ΔG values obtained from these calculations. Removing the charges costs fluoromethane 0.5 kcal/mole more in free energy than it costs trifluoromethane. This provides some support for the argument given above that explains the differece in solubility of the two species in terms of a water-fluorine hydrogen bond that is present in fluoromethane but not in trifluoromethane. Much of this difference is indeed due to a poorer "quality" of electrostatic interactions in trifluoromethane. But the free energy difference due to zeroing charges of fluoromethane and trifluoromethane does not seem to account for all of the 1 kcal/mole, even though the angle distribution suggests a great difference in fluorine-water hydrogen bonding. Why is this so? In other words, why is the magnitude of the free energy change caused by zeroing the charges of trifluoromethane larger than we would suspect for a species that does not have good fluorine-water hydrogen bonds?

To answer this question, we must consider the greater electron-withdrawing effect of three fluorines bonded to carbon, versus one fluorine so bonded. The quantum-mechanically derived charges (Table 3) show a much greater positive charge on the carbon of trifluoromethane than on the carbon of fluoromethane. This suggests that a strong electrostatic interaction between this carbon and a water oxygen is more likely in the case of trifluoromethane; this interaction would likely have a geometry characteristic of a hydrogen bond between a water oxygen and the trifluoromethane hydrogen,

(even though the charge on the trifluoromethane hydrogen is not very great in magnitude), since the water oxygen would approach the carbon of trifluormethane by moving as far as possible from the negatively-charged fluorines.

Figure 16 shows the trifluoromethane H - water O radial distribution function. The peak at approximately 2.5 Å does indeed suggest an interaction of the type suggested in the previous paragraph.

To further investigate this interaction, an analysis of the carbon-hydrogen-oxygen angle theta, where a theta of 180 degrees would correspond to a linear C-H-O hydrogen bond, was performed for the waters within the first peak of the radial distribution function of Figure 16, corresponding to all waters within 2.8 Å of the hydrogen of trifluoromethane. The area under this peak corresponds to 1.2 waters. The results are shown in Figure 17; a distribution about 120 degrees was found. This can be understood by realizing that a great deal of positive charge is actually on the carbon of trifluoromethane; the point charge on the CHF_3 carbon used in these calculations is 0.644. Thus, the water oxygen is oriented such that there is a close interaction between it and both the carbon and the hydrogen of CHF_3, without the oxygen approaching the negatively charged fluorines too closely.

A water-trifluoromethane dimer subjected to energy minimization starting from a configuration in which the C-H-O angle was 120 degrees displayed a final energy of -3.48 kcal/mole. Thus, this interaction is actually slightly stronger than the water-fluoromethane hydrogen bond.

Since much of the 0.5 kcal/mole difference in free energy between hydrated trifluoromethane and hydrated trifluoromethane with zero charges could be due to this interaction between the water oxygen and the carbon and hydrogen of trifluoromethane,

the structural evidence for this hydrogen bond provides further evidence for the poor quality of the fluorine-water electrostatic interactions in the case of trifluoromethane, since zeroing the charges removes both the interactions of waters with the fluorines and with the carbon and hydrogen of trifluoromethane. In other words, if the free energies of zeroing the charges of aqueous solutions of fluoromethane and trifluoromethane could be compared in terms of electrostatic interactions involving only fluorine-water electrostatic interactions, and not the type of hydrogen-bonding to the hydrogen of trifluoromethane shown in Figure 16, the difference in the free energies of zeroing the charges would likely be much greater than that given in Table 6; in this case the free energy of zeroing the trifluoromethane charges would include only the lose of water-fluorine interactions and not the loss of the hydrogen bond shown in Figure 16. The free energy change upon zeroing the charges of trifluoromethane is greater than one would predict for a species with poor fluorine-water hydrogen bonds because the loss of a hydrogen-water hydrogen bond also contributes to this free energy change. It is thus reasonable to conclude that much of the 1 kcal/mole difference in aqueous solubility between fluoromethane and trifluoromethane is due to a poorer quality of hydrogen-bonding between water and the fluorines of the latter.

We also decided to investigate the water structure around tetrafluoromethane, which is believed to be similar to that around methane and other hydrophobic solutes[22]. Figure 18 shows the fluorine-water hydrogen radial distribution function for tetrafluoromethane. Figure 19 shows the distribution of cos θ, as defined in Figure 3, for waters with a hydrogen within 3.5 Å of a fluorine, and Figure 20 shows the cos θ distribution for waters with a hydrogen between 3.5 and 5.5 Å from a fluorine. In neither case is there any evidence of water-fluorine hydrogen bonding; the H-O vectors adopt

orientations of the type shown in Figure 9(a).

CONCLUSIONS

The results described above suggest that trifluoromethane and tetrafluoromethane are solvated by waters having orientations similar to those seen around typically hydrophobic species; there is no evidence of hydrogen bonding to the fluorines of trifluoromethane or tetrafluoromethane. In contrast, one water-fluorine hydrogen bond is seen, on average, in the case of fluoromethane. These results, and the free energy calculations, suggest that the greater aqueous solubility of fluoromethane, as compared to trifluoromethane, is largely due to the presence of the hydrogen bond in fluoromethane and the absence thereof in trifluoromethane. Any electrostatic interactions between waters and the fluorines of trifluoromethane seem to be less stabilizing than those between waters and the fluorine of fluoromethane.

REFERENCES FOR CHAPTER FIVE

(1) Hine, J.; Mookerjee, P.K.; *J. Org. Chem.*1984, **40**, 292.

(2) Theory of Molecular Fluids. C.G. Gray; K.E. Gubbins. Oxford University Press, New York. 1984.

(3) Dill, K.A. *Biochemistry*1990, **29**, 7133.

(4) Rossky, P.J.; Karplus, M. *J. Am. Chem. Soc.*1979, **101**, 1913, and references therein.

(5) Jorgensen, W.L.; Ravimohan, C. *J. Chem. Phys.*1985, **83**, 3050.

(6) Davidson,D.W. "Water: a Comprehensive Treatise", Volume 2. Felix Franks, Editor. Plenum Press, New York, 1973.

(7) Tables of Experimental Dipole Moments. A.L. McLellan. W.H. Freeman and Co., San Francisco. 1963.

(8) Gough, C.A.; Pearlman, D.A.; Kollman, P.A. submitted to *J. Chem. Phys.*.

(9) Pearlman, D.A.; Case, D.A.; Caldwell, J.C.; Seibel, G.L.; Singh, U.C.; Weiner, P.; Kollman, P.A. (1991), AMBER 4.0, University of California, San Francisco.

(10) Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghia, C.; Alagona, G.; Profeta, S.,Jr.; Weiner, P.; *J. Am. Chem. Soc.* 1984, **106**, 765.

(11) Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A.; *J. Comp. Chem.* 1986, **7**, 232.

(12) Singh, U.C.; Kollman, P.A. *J. Comp. Chem.* 1984, **5**, 129.

(13) Baird, N.C.. *Can. J. Chem.* 1983, **61**, 1567.

(14) Harmony, M.D.; Laurie, V.W.; Kuczkowski. R.H.; Schwendeman, R.H.; Ramsay, D.A.; Lovas, F.J.; Lafferty, W.J.; Maki, A.G.; *J. Phys. Chem. Ref. Data* 1979 **8**, 619.

(15) MM2: Molecular Mechanics II. Quantum Chemistry Program Exchange, Program No. 423.

(16) Meyer, A.Y.; Allinger, N.L.; *Tetrahedron* 1975 **31**, 1971.

(17) Gough, C.A.; DeBolt, S.E.. Submitted to *J. Comp. Chem.*

(18) Sun, Y.; Spellmeyer, D.; Pearlman, D.A.; KOllman, P.; *J. Am. Chem. Soc.*, in press.

(19) Computer Simulation of Liquids, M.P. Allen and D.J. Tildesley. Oxford Science Publications, Oxford, UK, 1987.

(20) Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, A.; Haak, J.R. *J. Chem. Phys.*1984, **81**, 3684.

(21) Singh, U.C.; Brown, F.K.; Bash, P.A.; Kollman, P.A. *J. Am. Chem. Soc.*1987, **109**, 1607.

(22) Wen, W.-Y.; Muccitelli, J.A. *J. Solution Chem.*1979, **8**, 225.

Table 5.1

Free Energies (in kcal/mole) of Aqueous Solvation of Halogenated Methane Derivatives Relative to Methane[a]

Solute	$\Delta\Delta G_{sol}$
Fluoromethane	-2.2
Trifluoromethane	-1.2
Tetrafluoromethane	1.2
Chloromethane	-2.5
Trichloromethane	-3.0
Tetrachloromethane	-1.8
Bromomethane	-2.7
Tribromomethane	-4.1

(a) From Reference 1.

Table 5.2

Dipole Moments (in Debyes) of Methanol, Fluoromethane, and Trifluoromethane

Species	Dipole Moment[a]
Methanol	1.7
Fluoromethane	1.8
Trifluoromethane	1.6

(a) From Reference 7.

Table 5.3

Atomic Point Charges of Methane and Fluorinated Derivatives (in fractions of an electron)			
Molecule	Charge on Carbon	Charge on Fluorine	Charge on Hydrogen
CF_4	0.756	-0.189	-
CHF_3	0.644	-0.230	0.046
CFH_3	0.011	-0.251	0.080

Table 5.4

Bond Stretching and Bending Parameters for Bonds Involving Fluorine		
Bond	K_R (in kcal/mol Å^2)	R_o (in Å)
CT - F of CF_4	367	1.38
CT - F of CHF_3	367	1.33
CT - F of CFH_3	367	1.38
Angle	K_θ (in kcal/mol rad^2)	θ_o (in degrees)
F-CT-HC	35.0	109.5
F-CT-F	77.0	109.1

Table 5.5

6-12 Parameters Used for Tetrahedral Carbon, Fluorine, and Hydrogen		
Atom Type	R* (in Å)	ε (in kcal/mol)
CT	1.91	0.1094
F	1.75	0.061
HC	1.29	0.015

Table 5.6

Free Energy Changes (in kcal/mole) of Aqueous Solutions of Fluorinated Methane Derivatives Caused by Zeroing of Charges[a]

Solute	$\Delta\Delta G^{*}_{sol}$
CFH_3	2.4 ±0.3
CHF_3	1.915 ±0.005

(a) The value reported is the average value in the "forward" direction, as given in the first column. Each value represents the average of a forward (0 to 1) and a reverse (1 to 0) siulation. The error bars represent the difference between the calculations in the two directions.

FIGURE CAPTIONS

Figure 5.1

Radial distribution function for the carbon of fluoromethane and the oxygen of water.

Figure 5.2

Radial distribution function for the fluorine of fluoromethane and the hydrogen of water.

Figure 5.3

Definition of the F-H-O angle θ, as used in succeeding figures.

Figure 5.4

Distribution of cos θ, the F-H-O angle defined in Figure 3, for waters having a hydrogen within 2.5 Å of the fluorine of fluoromethane.

Figure 5.5

(a) and (c) Water orientations around the fluorine of fluoromethane that are consistent with the F-H-O angle distribution of Figure 4, but not with the C-H-O angle distribution of Figure 7.

(b) A water orientation around the fluorine of fluoromethane that is consistent with the F-H-O angle distribution of Figure 4 and with the C-H-O angle distribution of Figure 7.

(d) A water orientation that is not consistent with the F-H-O angle distribution of Figure 4 or the or the C-H-O angle distribution of Figure 7.

Figure 5.6

Definition of the C-H-O angle θ, as used in Figure 7.

Figure 5.7

Distribution of cos θ, the C-H-O angle defined in Figure 6, for waters having a hydrogen within 2.5 Å of the fluorine of fluoromethane.

Figure 5.8

Distribution of cos θ, as defined in Figure 6, for waters having a hydrogen at a distance between 2.5 and 5.8 Å from the fluorine of fluoromethane.

Figure 5.9

(a) A water orientation that is consistent with the cos θ distribution of Figure 8. (b) A water orientation that is not consistent with the cos θ distribution of Figure 8.

Figure 5.10

Radial distribution function for fluorine of trifluoromethane and the hydrogen of water.

Figure 5.11

Distribution of cos θ, as defined in Figure 3, for waters having a hydrogen within 2.5 Å of a fluorine of trifluoromethane.

Figure 5.12

Distribution of cos θ, as defined in Figure 3, for waters having a hydrogen at a distance between 2.5 and 4.8 Å from a fluorine of trifluoromethane.

Figure 5.13

Illustration of the orientation involved in a possible trifurcated hydrogen bond between water and trifluoromethane.

Figure 5.14

Definition of the angle θ, the angle between the O-H vector of water and the C-H vector of trifluoromethane.

Figure 5.15

Distribution of cos θ, the angle between the O-H vector of water and the C-H vector of trifluoromethane defined in Figure 14, for waters having a hydrogen within 4.0 Å of all three fluorines of trifluoromethane.

Figure 5.16

Radial distribution function for the hydrogen of trifluoromethane and the oxygen of water.

Figure 5.17

Distribution of cos θ, the C-H-O angle for the hydrogen bond between oxygen waters and the hydrogen of trifluoromethane, for waters with their oxygen atom within 2.8 Å of the hydrogen.

Figure 5.18

Radial distribution function for fluorine of tetrafluoromethane and the hydrogen of water.

Figure 5.19

Distribution of cos θ, as defined in Figure 3, for waters having a hydrogen within 3.0 Å of a fluorine of tetrafluoromethane.

Figure 5.20

Distribution of cos θ, the F-H-O angle defined in Figure 3, for waters having a hydrogen at a distance between 3.5 and 5.5 Å from a fluorine of tetrafluoromethane.

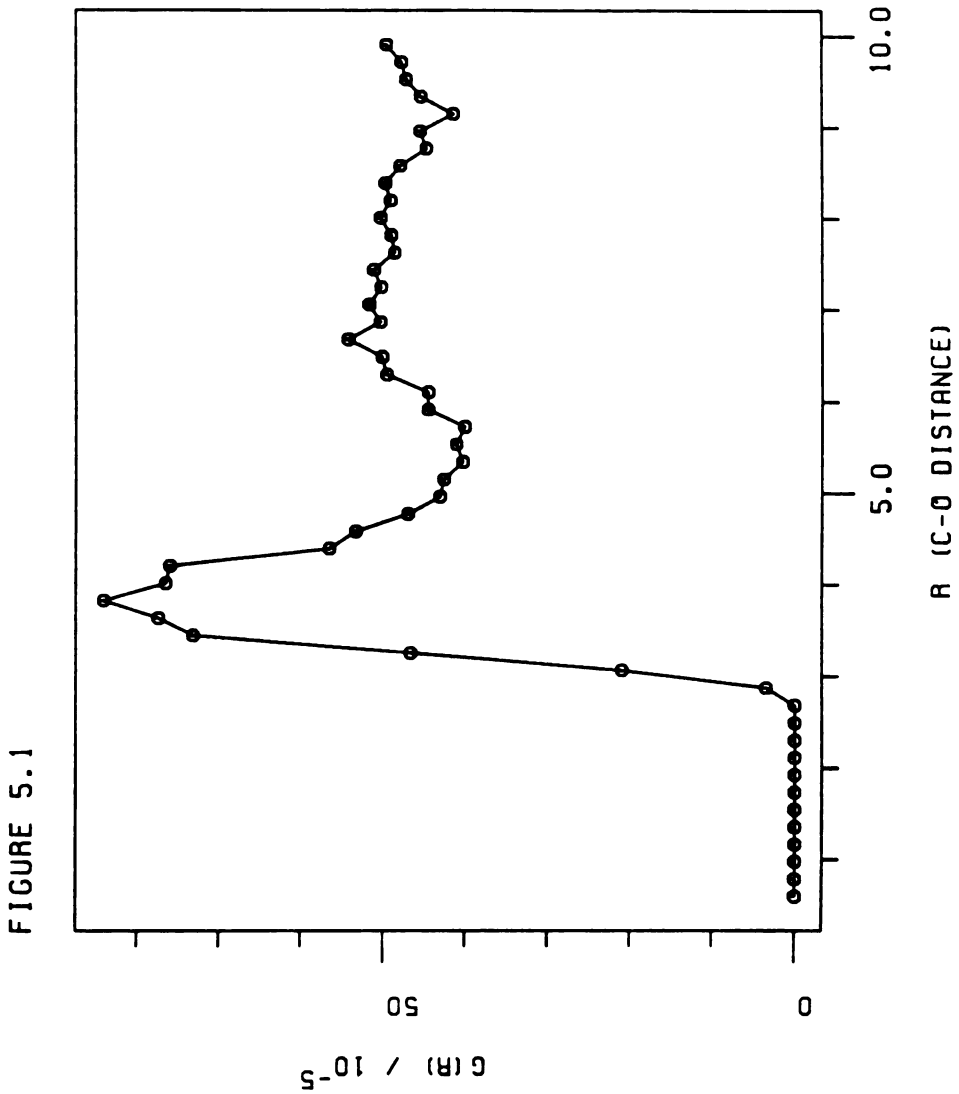


FIGURE 5.1

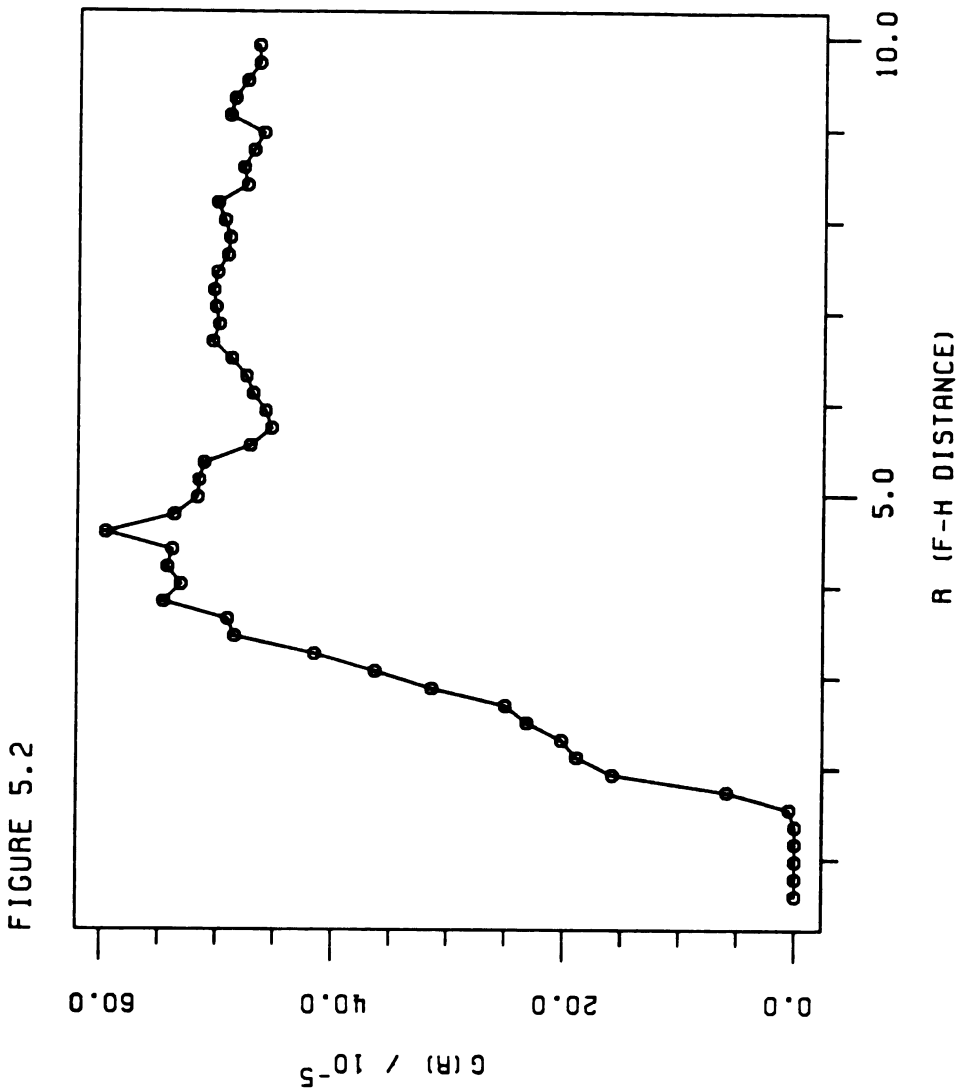


FIGURE 5.2

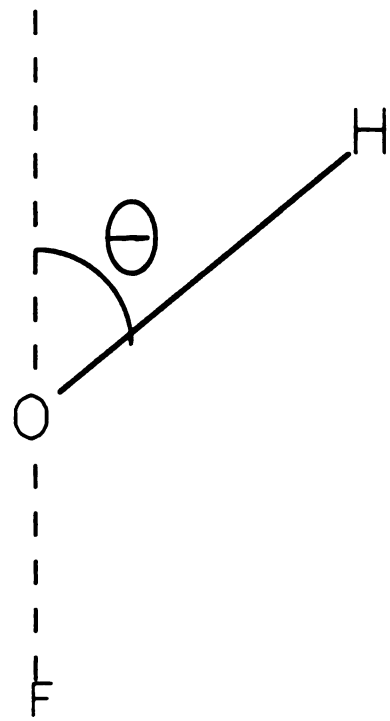


Figure 5.3

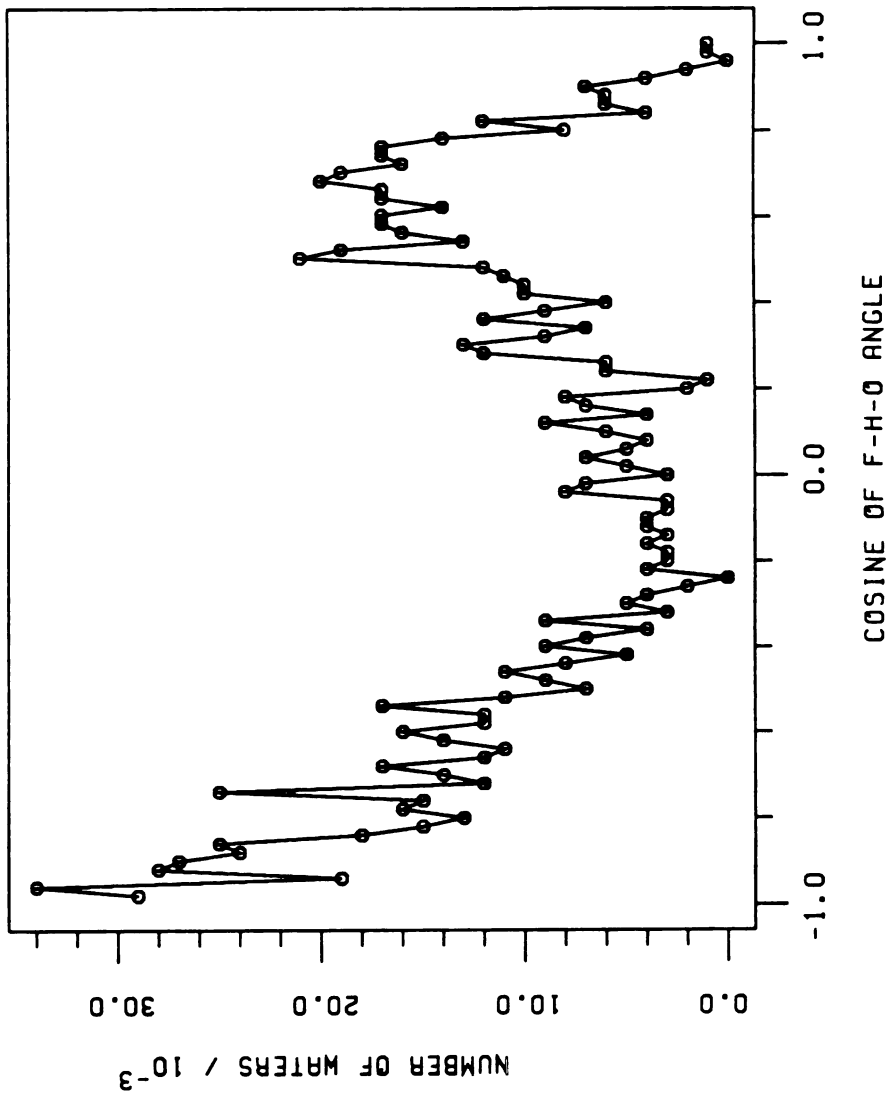

NUMBER OF WATERS / 10^-3
0.0
10.0
20.0
30.0
-1.0
0.0
1.0
COSINE OF F-H-O ANGLE

FIGURE 5.4

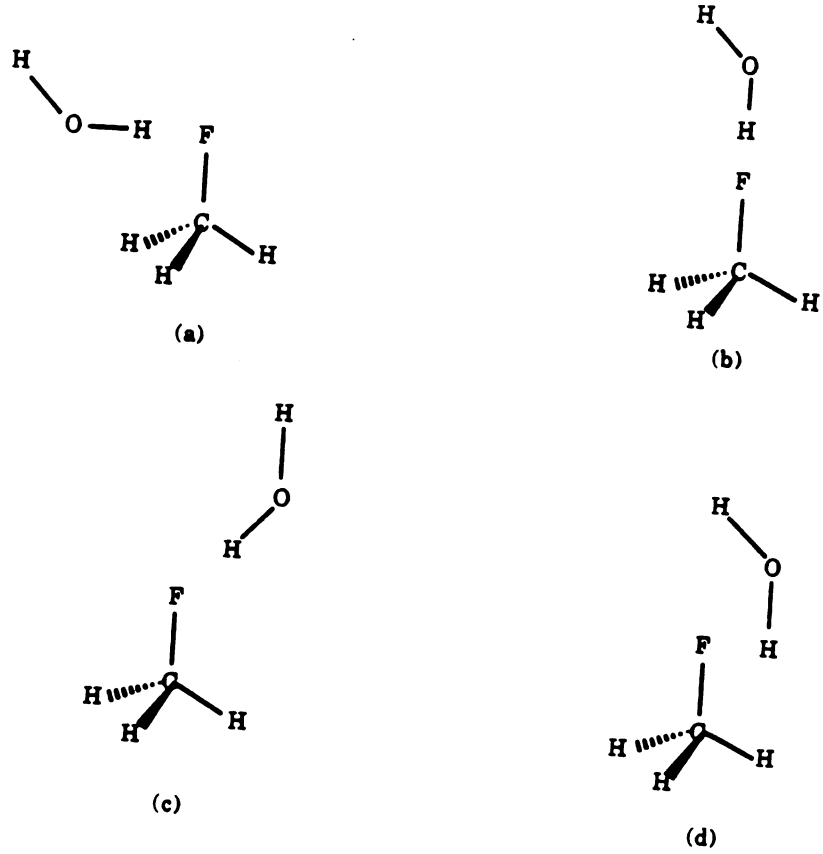


Figure 5.5

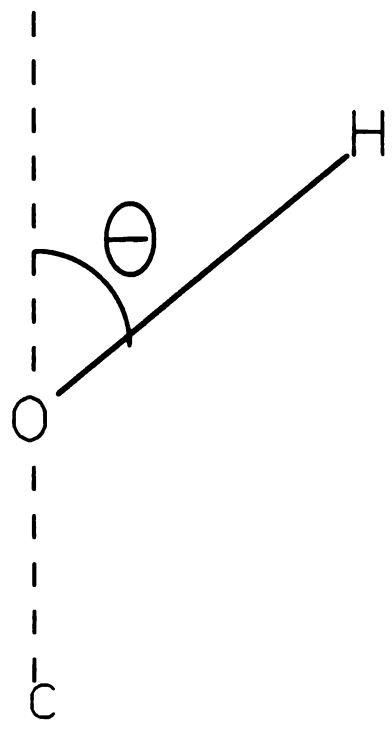


Figure 5.6

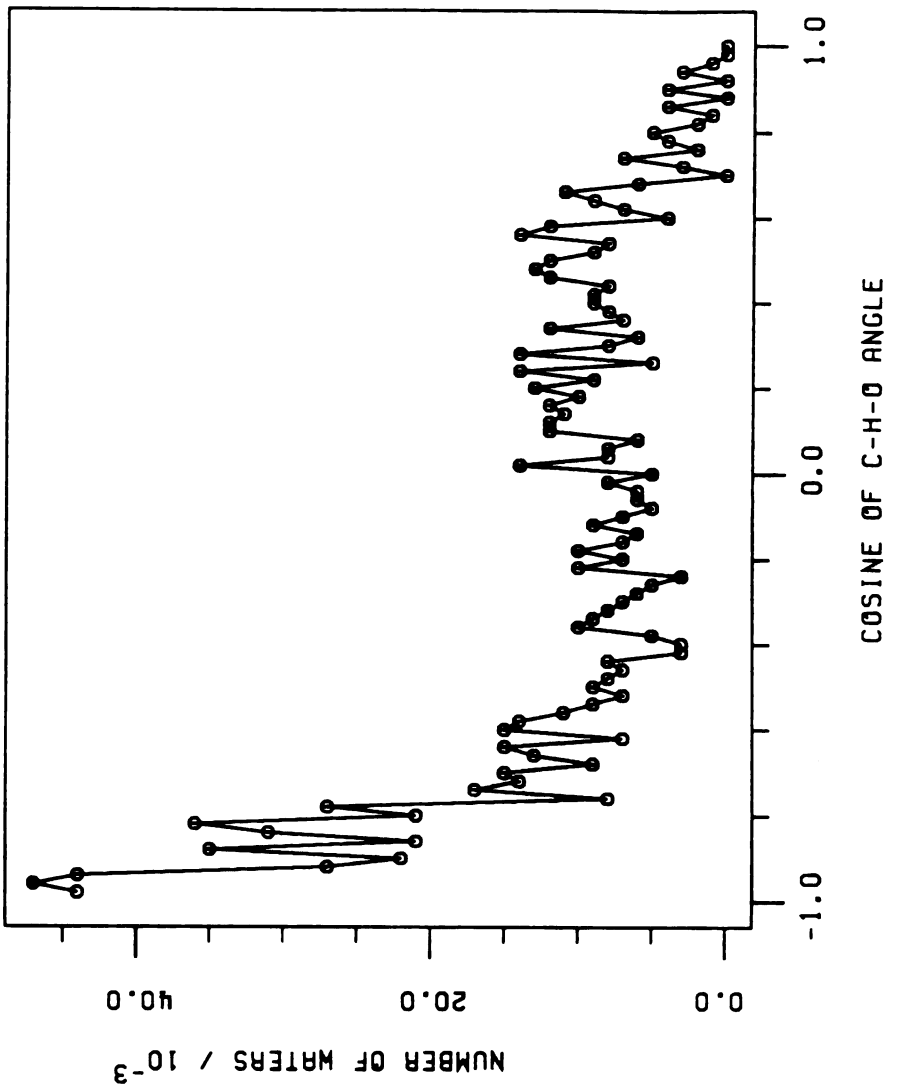


FIGURE 5.7

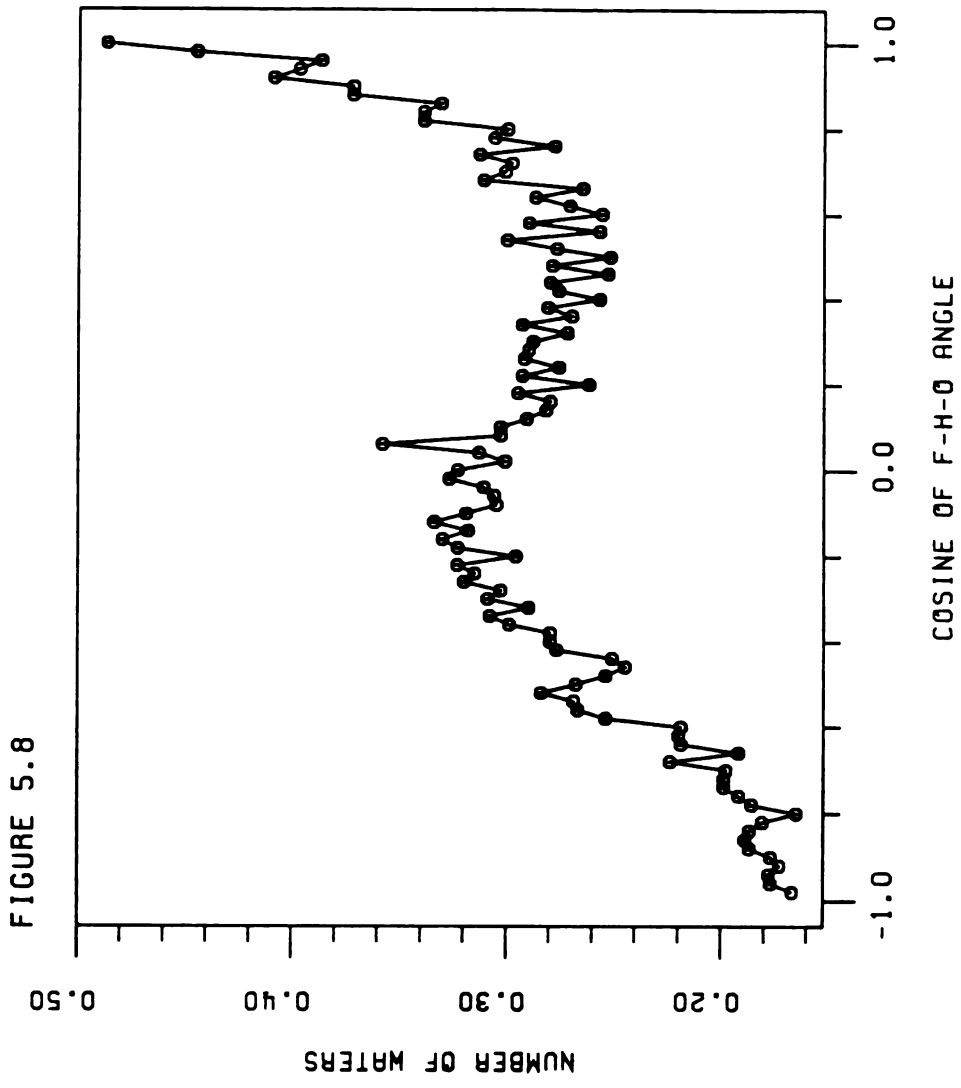
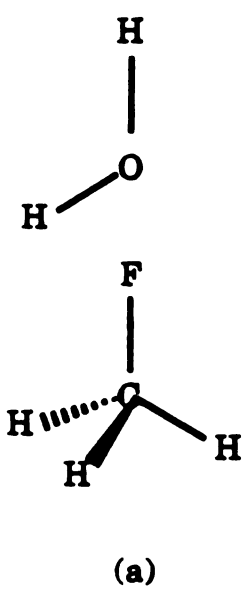
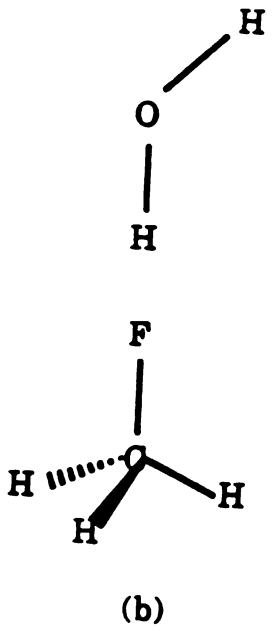


FIGURE 5.8



(a)



(b)

Figure 5.9

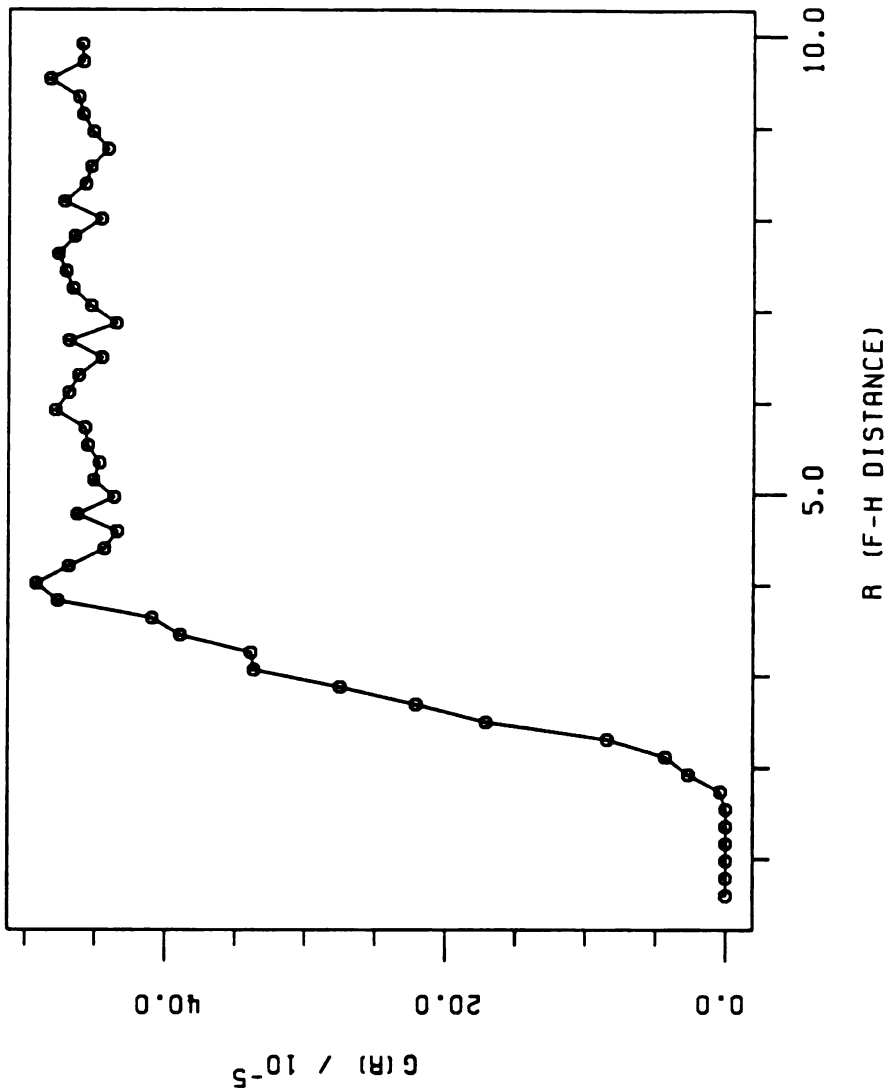


FIGURE 5.10

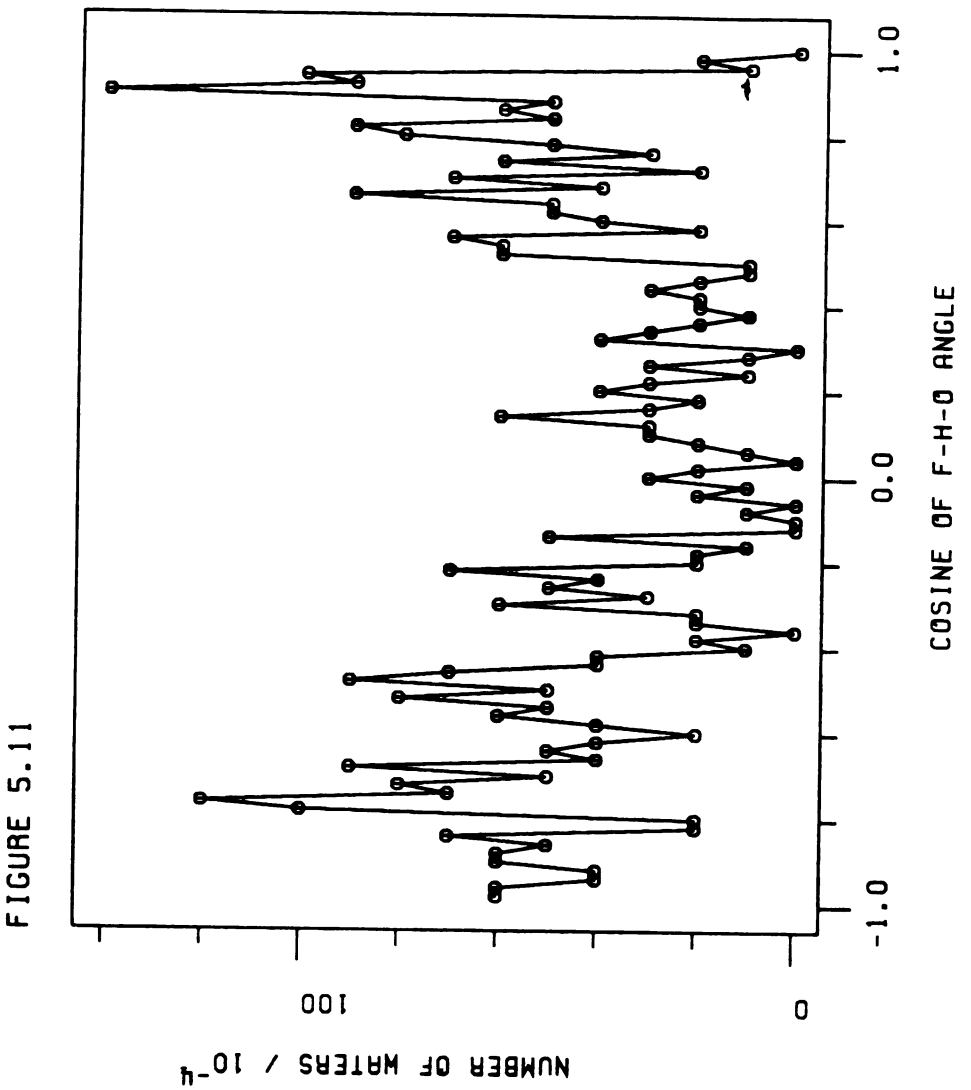


FIGURE 5.11

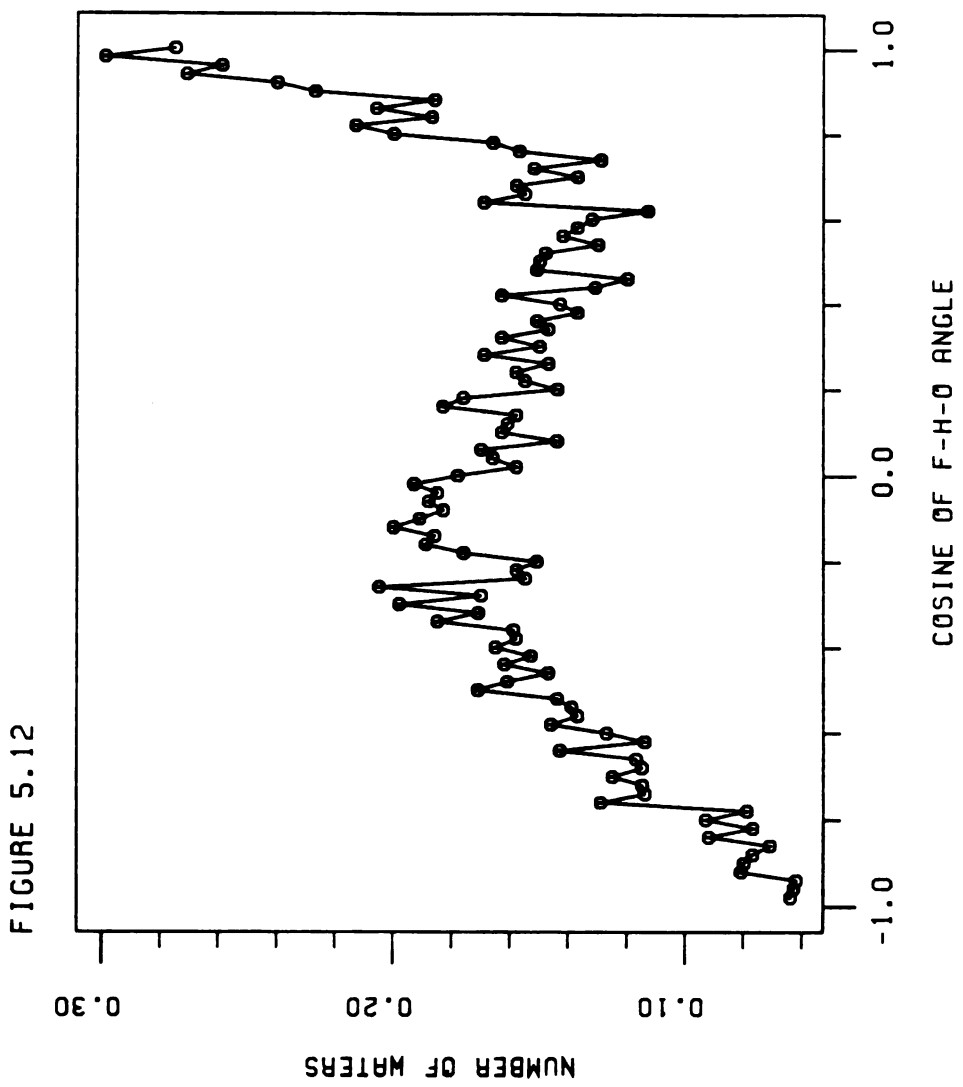


FIGURE 5.12

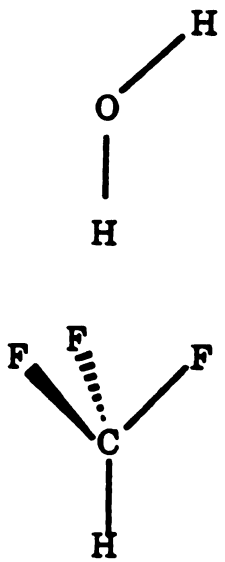


Figure 5.13

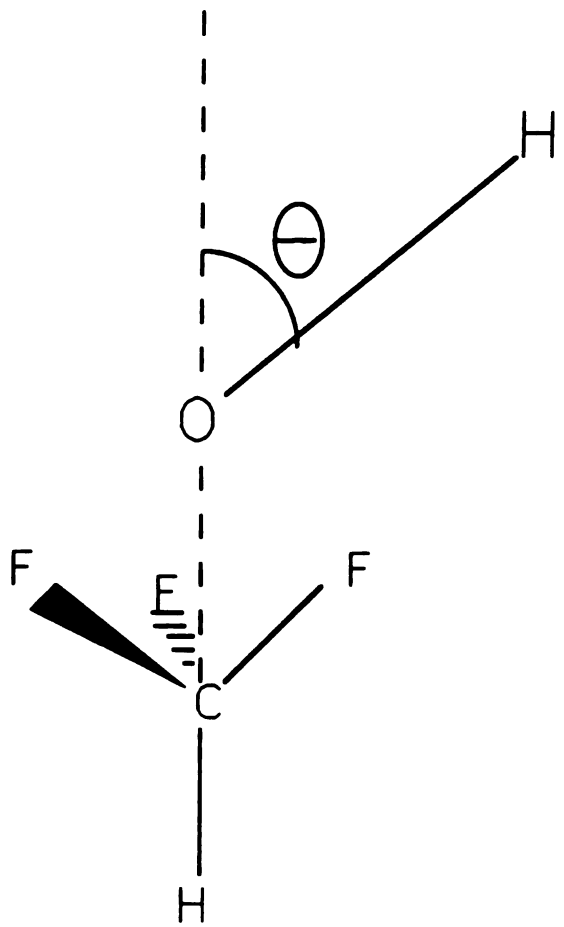


Figure 5.14

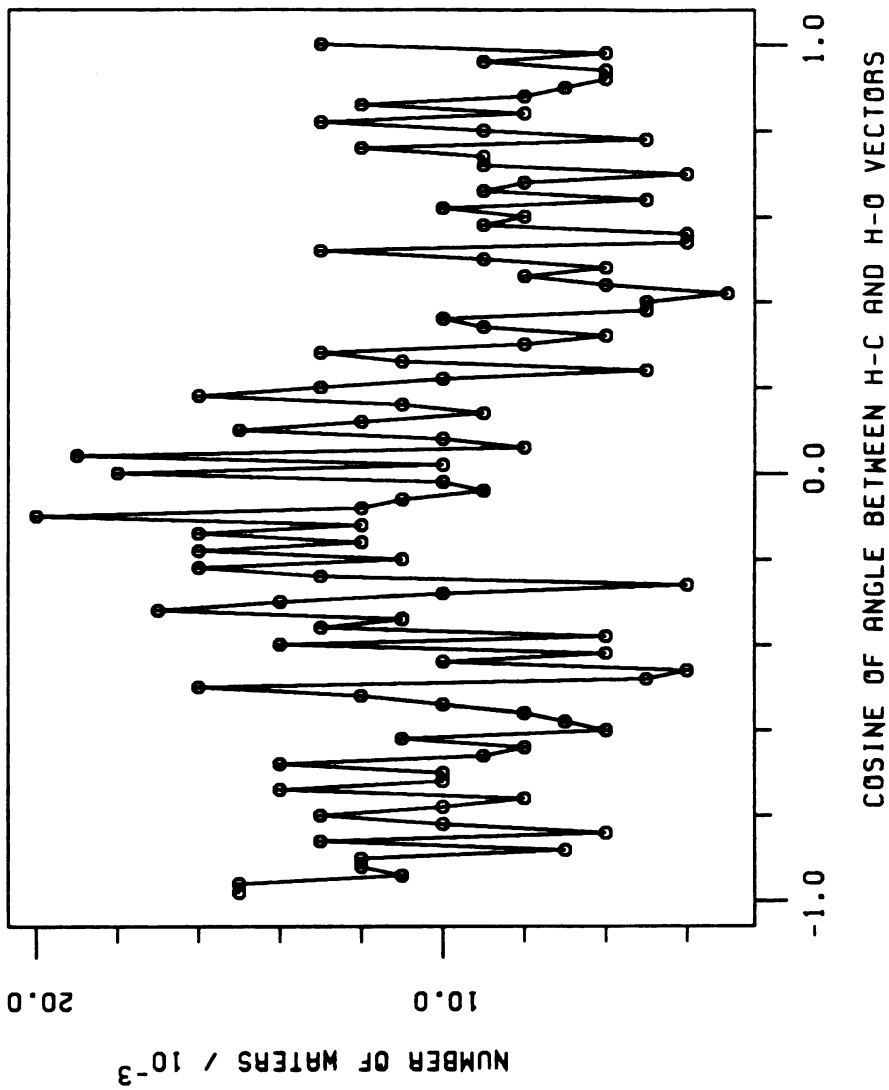

NUMBER OF WATERS / 10^-3
20.0
10.0
-1.0
0.0
1.0
COSINE OF ANGLE BETWEEN H-C AND H-O VECTORS

FIGURE 5.15

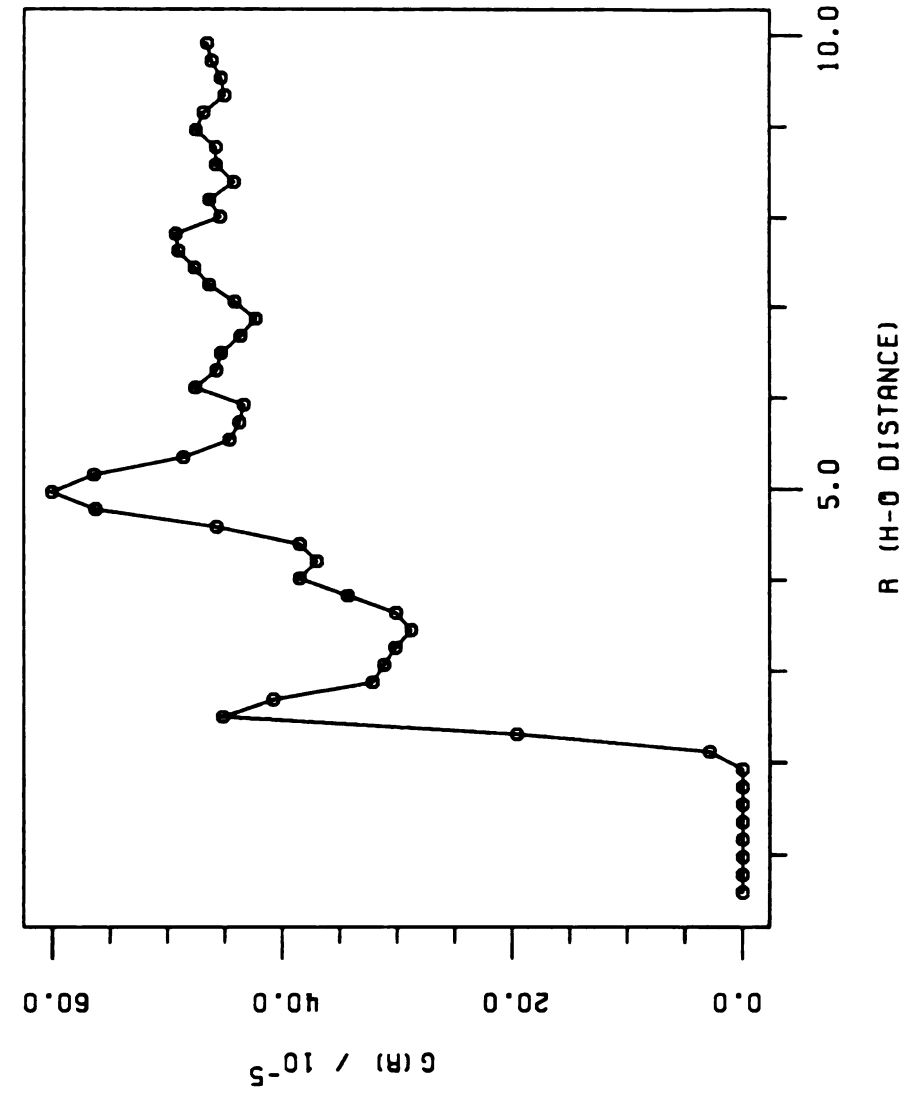


FIGURE 5.16

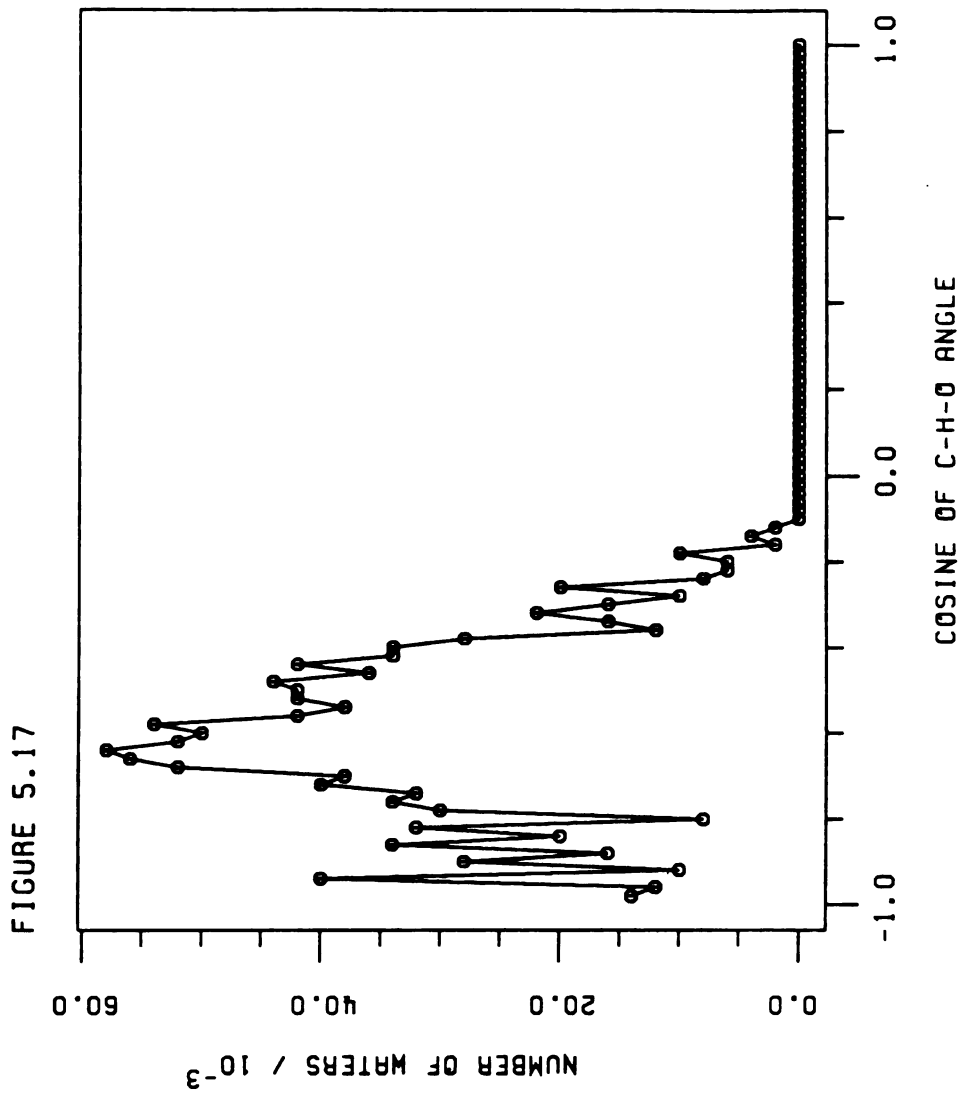

NUMBER OF WATERS / 10^-3
0.0
20.0
40.0
60.0
-1.0
0.0
1.0
COSINE OF C-H-O ANGLE

FIGURE 5.17

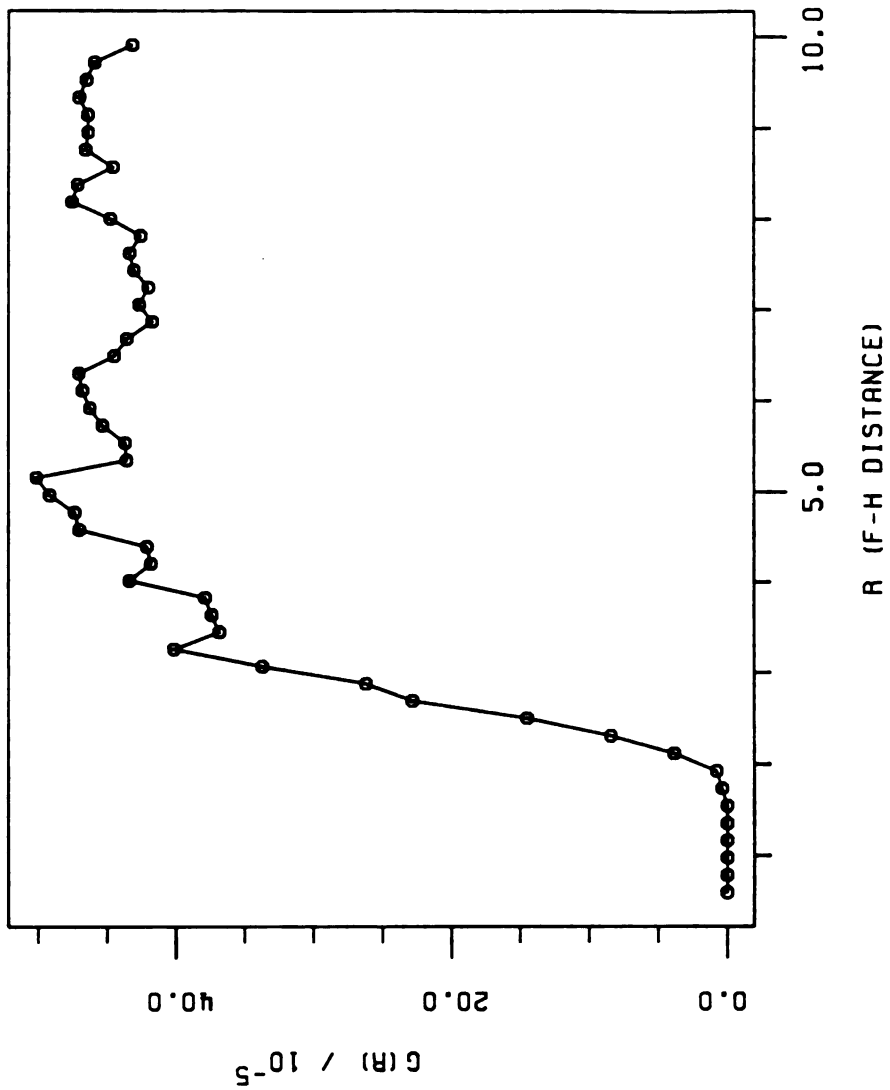


FIGURE 5.18

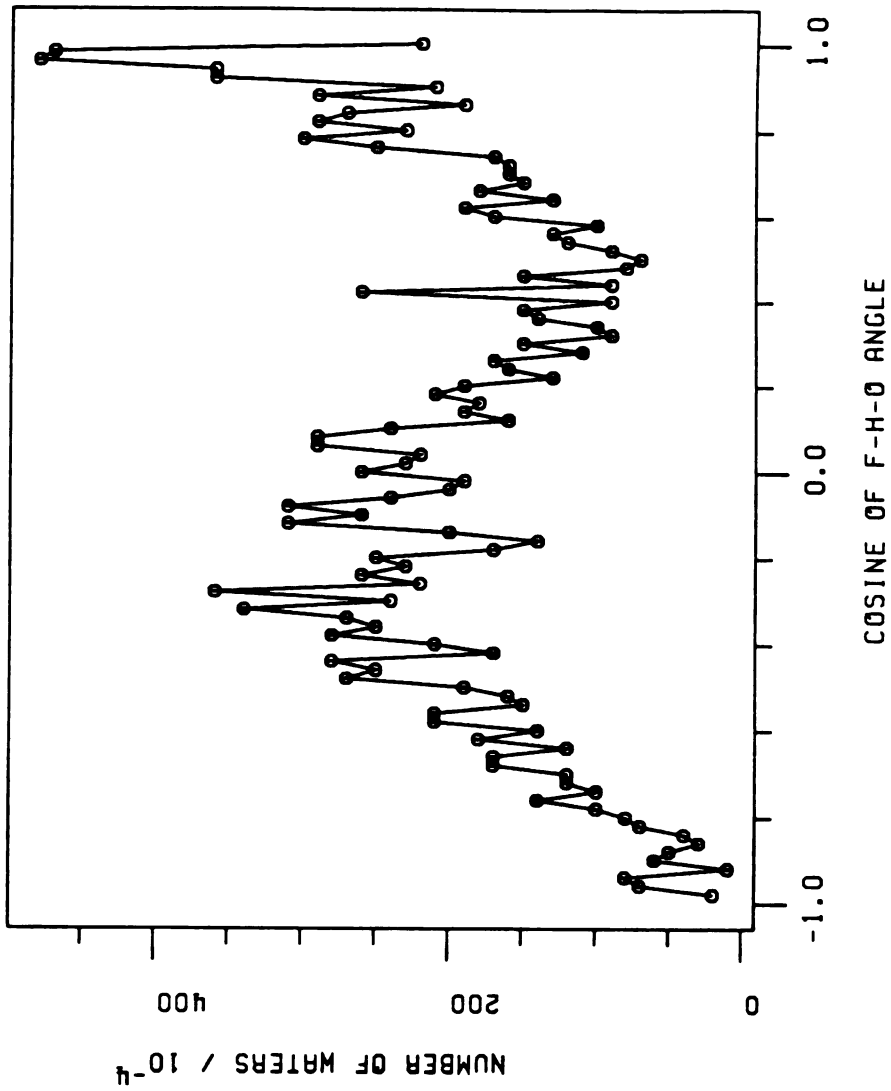


FIGURE 5.19

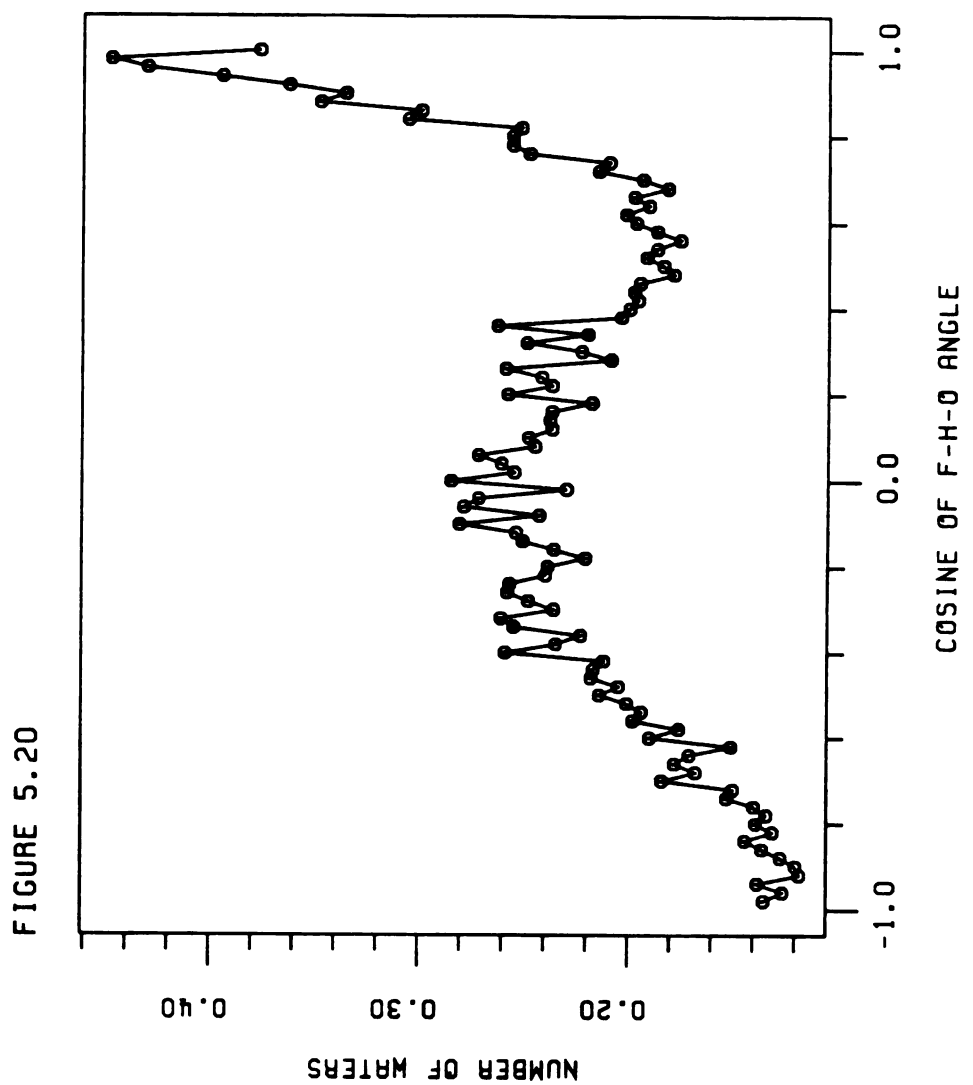


FIGURE 5.20