UCSF
UC San Francisco Electronic Theses and Dissertations

Title
Molecular modelling of peptides and proteins

Permalink
https://escholarship.org/uc/item/7d2033kh

Author
Daggett, Valerie D.

Publication Date
1990

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Molecular Modelling of Peptides and Proteins

by

Valerie Daggett

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

Copyright 1990
by
Valerie Daggett

ACKNOWLEDGEMENTS

First and foremost my thanks go to Peter Kollman and Tack Kuntz. They have been wonderful advisors. In fact, I cannot imagine better mentors and I will always be indebted to them.

I also thank Peter Kahn, Phyllis Kosen and my parents. Without their early help, encouragement and criticism this thesis would not exist.

Many people in Peter Kollman's lab have helped me along the way. I thank my office-mates Jim Caldwell and David Pearlman, they know why. I also thank my co-workers, Frank Brown (Chapter 5) and Stefan Schroeder (Chapter 7).

Thanks also go to the people in the computer graphics lab; their help and the use of the facilities were indispensable for this work.

Although many people have contributed to making U.C.S.F. a stimulating environment, I am especially appreciative of discussions with, and/or help along the way from, Darwin Alonso, Fernando Bazan, Erin Bradley, John Cashman, Gia DePillis, Ken Dill, Lydia Gregoret, and Wolfgang Sadee.

And, my special thanks to the person who put up with me through all of this, Patrik.

Molecular Modelling of Peptides and Proteins

by

Valerie Daggett

ABSTRACT

This thesis describes theoretical studies of peptides and proteins. The aim of this work was to elucidate the forces that stabilize small peptides and to explore the structure-function relationship in proteins. A variety of methods were employed: molecular dynamics, free energy perturbation calculations and semi-empirical molecular orbital calculations. Three separate studies focussing on small peptides are presented. The first entails free energy perturbation calculations of charge interactions with the helix dipole. The results are consistent with the helix dipole model, namely the introduction of charges that can interact favorably with the helix dipole stabilize the helix. The calculated free energies were too large to be realistic; however, this finding led to the investigation of different dielectric models: linear distance dependent dielectric function, inclusion of explicit solvent molecules, and a function that is sigmoidally dependent on distance. We found that the latter model reproduced a number of experimental properties of a small, helical peptide, with the advantage that it is much less computer-intensive than solvated systems. The results of a molecular dynamics simulation of polyalanine to address equilibrium motion and helix-coil transitions is also presented. This is the first simulation to show unwinding and refolding of an helix without imposing constraints on the system. The unfolded state remains very compact, however, and does not extensively sample conformational space.

Three projects involving proteins are also presented. The first involves a new method for estimating the free energy of interaction for particular portions of a molecule with its surroundings. From these studies we were able to rationalize the drop in catalytic activity when the catalytic base, Glu 165, of triosephosphate isomerase (TIM) is replaced by Asp. This method should also be of use in drug design by allowing a quick

evaluation of the effect of charged substituents. These studies of TIM were extended to include molecular dynamics simulations of wild-type TIM and five experimentally characterized mutant isomerases. We employed both noncovalent enzyme-substrate complexes and covalent complexes, meant to mimic the transition state, in order to examine the effects of the mutation on both substrate binding and catalysis. Our simulations reproduce the highly cooperative nature of the interactions in the active site and suggest that this approach may be useful for identifying particularly promising sites for mutation. Lastly, we describe molecular orbital calculations of a reduced representation of the trypsin active site to address amide and ester bond hydrolysis. We found a novel sequence of events: the lowest energy path for formation of the tetrahedral intermediate involved approach of the serine oxygen to the substrate followed by the simultaneous proton transfer to the histidine and bond formation between the serine and substrate.

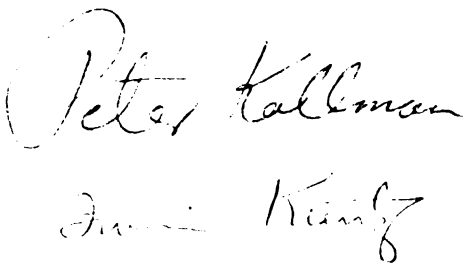


TABLE OF CONTENTS

CHAPTER ONE: Introduction

A knowledge of the mobility of proteins is essential to understanding their properties. For globular proteins internal motion has been implicated in enzyme catalysis, ligand binding, macromolecular recognition and, of course, protein folding. Simulations of proteins can provide detailed information about these processes that often cannot be obtained experimentally.

Generally the method of molecular dynamics is used to simulate protein motion. With this method one must have a well-defined starting structure. Another requirement is a potential energy function whose parameters have been derived to reproduce structures and energy trends in various model systems. Using molecular dynamics, the atoms move due to the force of their own kinetic energy and the forces exerted upon them by all other atoms. So, a trajectory of a constantly changing molecule moving through phase space is generated.

The work described in this thesis makes extensive use of molecular dynamics to study protein motion, the behavior of small peptides, and to calculate free energy changes. This thesis is divided into two sections. The first section describes studies of small peptides to explore the interactions that stabilize secondary structure, the structural transitions that occur, and different electrostatic models for use in force fields. The second section describes studies of proteins to explore the relationship between structure and function. Each chapter is designed to stand alone and contains an introduction, description of the methods used and the results obtained, and finally a discussion section. There is a common thread running through these seemingly disparate themes. The aims of the protein work are straightforward---to understand how motion and amino acid replacements, both leading to structural changes, affect function. The peptide studies were undertaken to investigate how charge and motion alter structure, working from the premise that the study of small peptides is relevant to the behavior of proteins and will eventually aid in protein design and studies of protein folding.

CHAPTER 2: Free Energy Perturbation Calculations of Charge Interactions with the Helix Dipole

The α-helix has an electric macrodipole arising from the alignment of peptide dipoles parallel to the helix axis (Wada, 1976). It has been suggested that the helix dipole plays a role in substrate binding (Hol et al., 1978; Wierenga et al., 1985; Warwicker and Watson, 1982), catalysis (Hol et al., 1978; Van Duijnen et al., 1979; Hol, 1985), ion binding (Quiocho et al., 1987), and protein folding (Hol et al., 1981). Peptide bonds are generally assigned a dipole moment of 3.5 Debyes(D). Confirmation for this value of the dipole moment of peptide bonds comes from microwave data on formamide gas, which has a dipole moment of 3.71 D (Kurland and Wilson, 1957). When polarization due to hydrogen bonding and the effect of a polar solvent are considered, the peptide dipole moment may be as great as 5.0 D (Wada, 1976). A continuous line dipole of 3.5 D per repeat unit (1.5 Å) is a reasonable approximation of helix dipoles in protein interiors where the dielectric constant is low (Hol et al., 1978). This representation results in 1/2 a positive charge at the amino-terminus of the helix and 1/2 a negative charge at the carboxyl-terminus; these resultant charges are essentially independent of length.

This description of the helix macrodipole with charges centered at the ends of the helix is an approximation. A more realistic model has the charges residing primarily on the four mainchain amide hydrogens at the amino-terminus that do not form hydrogen bonds and the four lone carbonyl oxygens at the carboxyl-terminus. Thus, each pole of the helix dipole spans four residues (Figure 2.1). Therefore, it may be better to think of electrostatic interactions with the helix dipole in terms of specific peptide dipole moments, which are free to interact with nearby charges since they do not form hydrogen bonds with residues from the adjacent turns in the helix.

In any case, the alignment of dipoles within secondary structure appears to be an important determinant of the tertiary structure of proteins. Hol and co-workers (1981) have calculated the electrostatic interactions between α-helices and β-sheets for a number

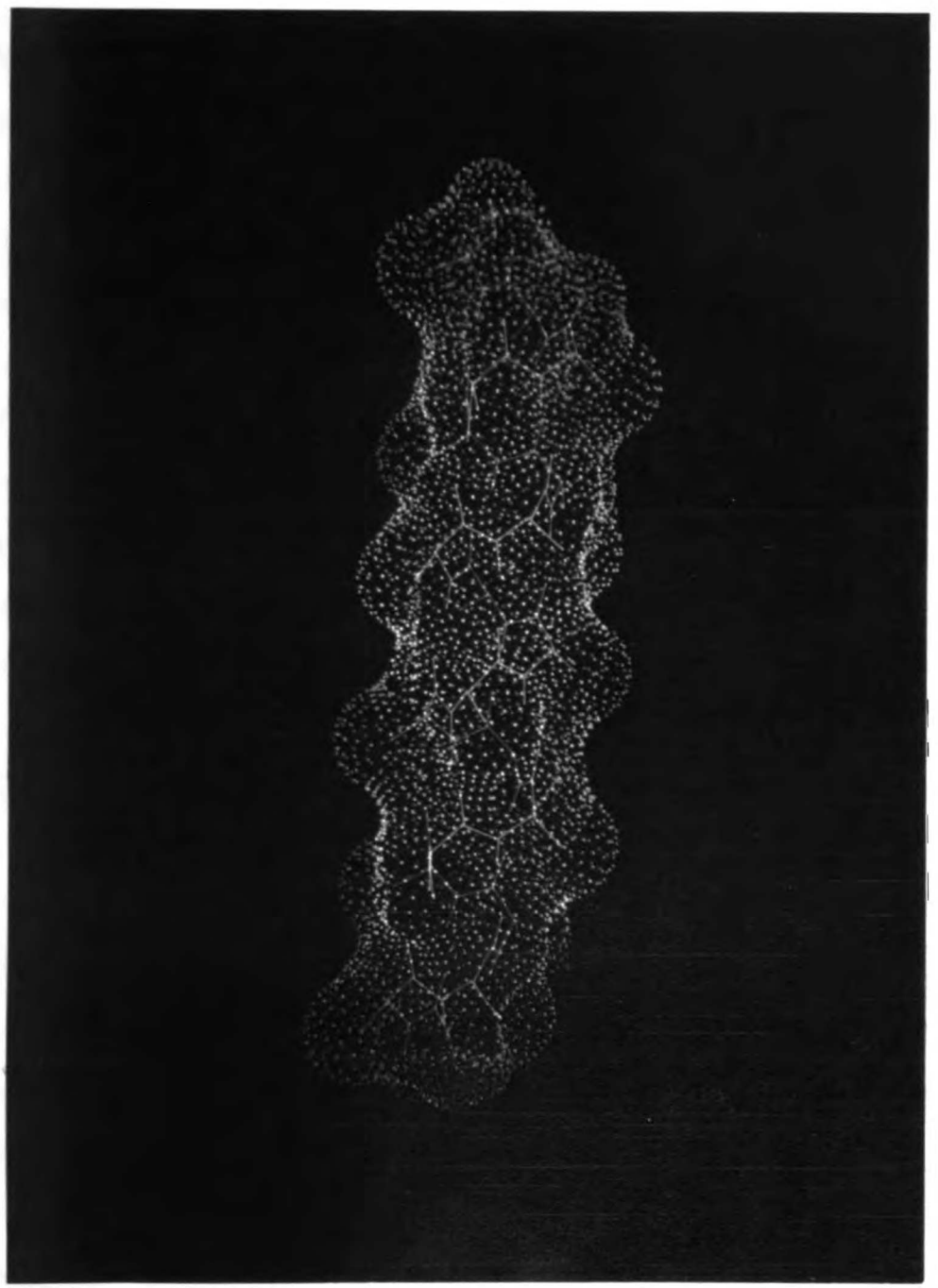


Figure 2.1. Electrostatic Surface Potential of Unblocked Polyalanine in the α-Helical Conformation. The amino-terminus is on the left. The colouring scheme is from yellow (positive) to blue (negative).

of proteins and suggest that secondary structures pack in orientations maximizing favourable electrostatic interactions. Further, these authors claim that the interactions are strong enough to play a critical role in determining the tertiary structure of proteins. Sheridan and co-workers (1982) have examined the electrostatic interaction energy of helical bundles composed of four α-helices. Their results indicate that an antiparallel arrangement of the helices is favoured over a parallel orientation by 20 kcal/mole. Parallel arrangements are sometimes found in proteins though their interhelical angles are usually larger than those of their antiparallel counterparts (Janin and Chothia, 1980; Richardson, 1981). A larger tilt angle presumably minimizes charge repulsion between the poles of the helices. Further, Sheridan *et al.* (1982) contend, in accord with Hol and co-workers (1981), that the helix dipole is important in the folding process, as they show that energies for the most direct pairwise interactions of the helices *en route* to the folded state show successively greater degrees of electrostatic stabilization. However, given that they examined an arbitrary folding pathway, their results may not be general.

Other investigators disagree with Hol, Sheridan and their co-workers, concluding instead that the role of the α-helix dipole in stabilizing and determining tertiary structure is only marginally significant. Rogers and Sternberg (1984) have examined the role of α-helix dipoles in stabilizing the tertiary structure of proteins using three dielectric models: the uniform dielectric model, the distance dependent dielectric model, and the cavity model. They find that electrostatic stabilization is very sensitive to the dielectric model employed. For the first two models they calculate interaction energies for α-helix/α-helix packing of the same magnitude as those calculated by Hol *et al.* (1981) and Sheridan *et al.* (1982) The cavity model, however, yields interaction energies that are up to an order of magnitude lower than those calculated using the other two models. Rogers and Sternberg argue that the cavity model is the most reasonable one for macromolecules since it describes the different dielectric responses of the solvent and the protein. Hence, they conclude that the role of the α-helix dipole in stabilizing tertiary structure is not sub-

stantial. All three models are, however, gross simplications of the actual system, which can be separated into at least four distinct regions: bulk water with a high dielectric constant; the first hydration sphere surrounding the protein with unknown dielectric properties; bulk protein with a heterogeneous dielectric environment; and the outer surface of the protein with unknown dielectric properties.

Recent experimental work suggests that interactions with the helix dipole plays a role in stabilizing small, isolated helical peptides. Small peptides (< 20 residues) are not predicted to be helical in water (Zimm and Bragg, 1959; Sueki et al., 1984). Yet, the C-peptide of RNase A is partially helical under certain conditions (Bierzynski et al., 1982). In exploring the factors responsible for the unpredicted stability, Baldwin and co-workers found that charge interactions with the helix dipole are important in stabilizing the helical structure (Shoemaker et al., 1985, 1987). Scheraga's group recently reported an extension of the Zimm-Bragg theory, incorporating the effect of specific peptide charge-dipole interactions on helix stability, that predicts reasonably accurately the overall helix probabilities of various C-peptide derivatives (Vasquez et al., 1987).

In this paper we explore charge interactions with the helix dipole. To this end, we have employed the free energy perturbation method, which uses molecular dynamics to evaluate a statistical mechanically derived formulation of the free energy (Singh et al., 1987). We chose to use the free energy perturbation method rather than to evaluate static structures with Coulomb's law because there are much smaller statistical errors associated with calculated free energy changes than with changes in internal energies or enthalpies. We also feel that it is important to allow structures to sample thermally allowed motions, and, in fact, we found a number of cases where ΔG differed significantly from what one calculates using the ΔE from idealized static structures or energy minimized structures.

The free energy perturbation method has been shown to be very effective in calculating solvation free energies (Bash et al., 1987a), binding free energies (Bash et al.,

1987b), and catalytic free energies (Rao et al., 1987) with good agreement with experiment. These studies involved perturbing one amino acid into another and calculating a single number to compare with experimental results. Recently a slightly different application of the free energy perturbation method was described that involves perturbing the charges of various groups of atoms and determining the free energy of interaction between these groups and their environment (described in Chapter 5), thereby providing qualitative insight into the roles of specific charge interactions in substrate binding, catalysis, or structural stabilization. We also emphasize a qualitative point of view in this paper because: (1) No experimental free energies exist for these charge-dipole interactions; and (2) There are technical problems in introducing charged groups into molecular solvent models. This second problem is similar to that encountered by Bash *et al.* (1987a). They find good agreement between the calculated and experimental solvation free energies for neutral molecules but calculate large free energy changes for altering the net charge of the system and point out problems with correcting for the reaction field for perturbations that involve the creation of a charge.

The results of calculations on a model α-helix (polyalanine, 20 residues) are presented below. We chose this model in order to examine inherent helical properties without the complications involved with varied amino acid sequences. We calculated changes in free energies for perturbing the charge of a methyl group as a function of position along the helix and distance from the helix axis. In this way we examined charge interactions with the helix dipole and salt bridge formation. Similar calculations were performed on β-strands as controls, because Hol and co-workers (1981) suggest that there is little or no resultant dipole moment for an antiparallel β-strand and that the dipole moment of a parallel β-strand is roughly a factor of five weaker than the helix dipole.

METHODS

Calculation of Free Energy Changes

All calculations were performed using AMBER version 3.0 (Singh et al., 1986).

We calculated Gibbs (G) free energy changes using equation 1, where ΔH is the difference in the Hamiltonian between two states, ΔG is the free energy difference between these states, R is the gas constant, T is the absolute temperature, and the symbol $< >_{ref}$ indicates that an ensemble average is taken with respect to some reference state. (See Singh et al., 1987 for a more thorough discussion of the method.)

$$\Delta G = - RT \ln < \exp (-\Delta H / RT) >_{ref} \qquad \text{Eq. 1}$$

In the cases discussed here, H represented the interaction energy of the perturbed group with its surroundings. These interaction energies were calculated at intermediate points along the conversion pathway using an empirical force field with standard united atom parameters (hydrogens on carbon atoms are incorporated into the van der Waals radius of the carbon) (Weiner et al, 1984). Molecular dynamics at 300 K was used to generate the ensemble of structures. Using the free energy perturbation method, the free energy due to changes in intra-group interactions (that is, within the perturbing group) and inter-group interactions (that is, the group interacting with its environment) can be calculated separately. In AMBER 3.0 only inter-group interactions are included in the free energy determination.

The perturbations reported here involved changing the charge of C_β (the methyl group of alanine) from its normal value of 0.031 to, in most cases, a full positive or negative charge in specific residues along the sequence. We defined a perturbed group as an entire residue, or residues, hence all of the atoms for a particular residue were part of the perturbed group even though C_β was the only atom changed. The force field parameters of the unaltered atoms were held constant so that the free energy change for the "perturbation" of these atoms was zero by equation 1. (The control calculation in which only the charge of C_β was defined as the perturbed group confirmed that the dominant interaction was between the C_β charge and its environment and not other atoms in the perturbed residue.) Thus, the calculated free energy changes represent the interactions between the perturbed residue and its environment but do not include intra-group interactions. This

distinction is particularly important in the case of the salt bridge calculations where the charges of two residues were perturbed simultaneously. In analogy to the example above, the two residues were defined as part of the perturbing group and intra-perturbed residue interactions were not calculated. As a result, the salt bridge calculations involved interactions between the perturbed residues and the peptide but did not include the inter-residue charge-charge interaction.

Computational Details

We employed the windowing method of perturbation, which involves breaking up the perturbation into discrete steps (windows), as described by Singh *et al.* (1987). Except where noted below, progression from the unperturbed to the perturbed structure, for the various simulations described, was carried out using 5-21 windows. 200 equilibration steps and 400 steps of data collection were performed at each window, with a step size of 1 femtosecond. This corresponded to a time course of 3.0-12.6 psec for each perturbation. To test the dependence of the calculated free energy changes on the computational parameters and to ensure that we had sufficient sampling during the course of perturbation, we also varied the number of windows, the number of steps of equilibration and data collection per window, and the total time for one particular perturbation (C_β of res 1 $\rightarrow$ -1, C_β of res 20 $\rightarrow$ +1). All free energies were calculated at 300 K. A distance dependent dielectric constant ($\varepsilon=r_{i,j}$, where r is the intercharge separation between atoms i and j) and a 10 Å nonbonded cutoff were used for the calculations. (Calculations were also performed using a 50 Å nonbonded cutoff, yielding similar free energy changes to those presented here.) The reported free energies represent the average of at least two independent simulations.

Generation of Structures

We performed calculations on various conformations of polyalanine (20 residues), both with and without terminal blocking groups (Figure 2.2). The initial α-helix and β-strand structures were generated with ideal phi and psi angles using a routine from

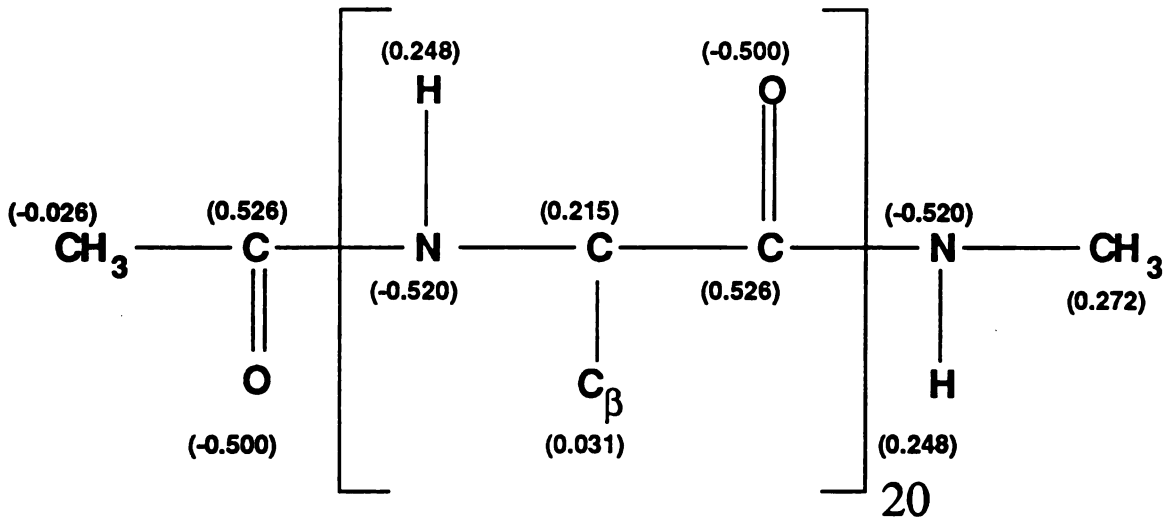


Figure 2.2. Starting Structure for Perturbation Calculations. Partial charges are given in parentheses. Models without blocking groups only contain the residues in brackets and are terminated as shown.

ECEPP (Momany et al., 1975) (α-helix: ϕ = -57, ψ = -47, antiparallel β-strand: ϕ = -140, ψ = 135, parallel β-strand: ϕ = -119, ψ = 113, extended β-strand: ϕ = 180, ψ = 180). It should be emphasized that each β-strand calculation involved perturbation of an isolated strand, not a strand within a sheet. The partial charges in Figure 2.2 gave rise to a dipole moment of approximately 3.8 Debyes per residue. Explicit counterions and solvent molecules were not present.

All structures were minimized briefly to remove bad contacts and then equilibrated at 300 K for at least 10 picoseconds (psec), using molecular dynamics, prior to free energy perturbation calculations. In some cases the dihedral angles of the helix were restrained, by applying a force constant of 100 kcal/mole-rad^2, to minimize conformationally-induced contributions to the calculated free energies. All of the salt bridge calculations were performed with constrained helical structures to avoid conformational changes accompanying salt bridge formation. The β-strand structures were constrained for all calculations, as without restraints they became drastically distorted.

Distance Dependence of Calculated Free Energies

To examine the effect of the distance between the perturbed charge and the helix axis on the calculated free energies, we calculated the free energy change for a particular perturbation (residue 1, C_β charge $\rightarrow$ -1; residue 20, C_β charge $\rightarrow$ +1) as a function of the distance between C_α and C_β. Equilibrated (25 psec) polyalanine, without terminal blocking groups, was used as the starting structure for the simulation with all C_α and C_β carbons separated by 2.0 Å bonds. The new bond length replaced the normal bond length (1.526 Å) in the parameter list. The structures achieved the new bond length within 0.2 psec of molecular dynamics and were stable during the perturbation simulation. Similarly, the starting structure for each subsequent perturbation was the final structure from the previous simulation (with 0.5 Å shorter bond lengths). The ϕ,ψ angles were restrained and the total time for each perturbation was 5.4 psec.

Dielectric Dependence of Calculated Free Energies

For a few select perturbations we calculated free energies using a dielectric constant of 80. All structures in this set (α-helix and the three β orientations) required dihedral restraining forces of 150 kcal/mole-rad^2. The large force constant was necessary to maintain the structures because of the diminution of the strength of the hydrogen bonds due to the higher dielectric constant.

Charge Distribution Dependence of Calculated Free Energies

Model peptides were then constructed, with actual amino acids at the ends of the helix, to test the effect of having more realistic charge distributions in these positions compared to the simple model of increasing C_α-C_β bond lengths. We constructed the six possible combinations of X_1-$(Ala)_{18}$-X_2, where X_1= Glu, Asp and X_2= His, Lys, Arg. The resulting six structures, with X_1 and X_2 in their ionized states, were minimized briefly and then equilibrated at 300 K for 10 psec. The free energy change for neutralizing the charges of each structure was calculated. The pertinent original and all perturbed charges are given in Figure 2.3. The ϕ, ψ angles were restrained and the total time for each perturbation was 5.4 psec.

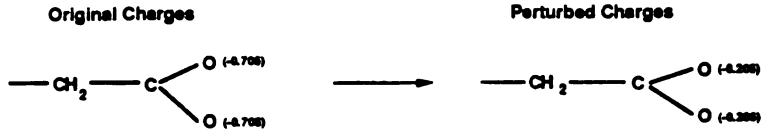
RESULTS

Overview of Charge Dipole Interactions

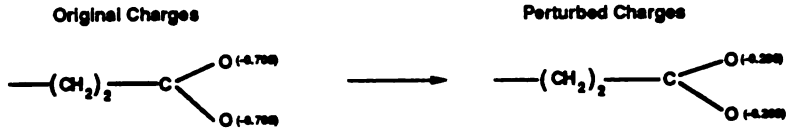
α-Helix

We present results for perturbing the charge of a methyl group at different positions along the helical axis in Table 2.1. For every case in which a charge was introduced at either end of the helix, the free energy change was consistent with the helix dipole model, regardless of whether the molecule was blocked at the termini. All charges expected to interact favourably with the macrodipole yielded negative free energies. Likewise, perturbations that involve the introduction of charges with the same sign as the poles of the dipole resulted in positive free energy changes.

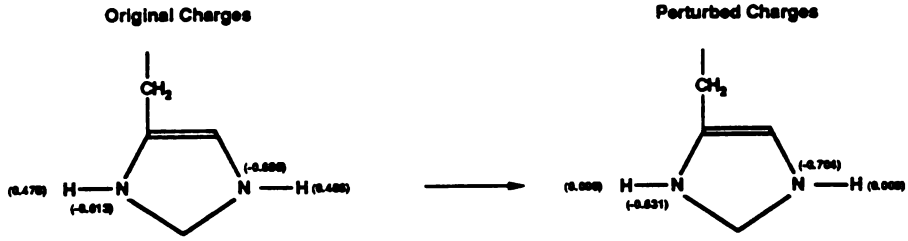
Aspartate



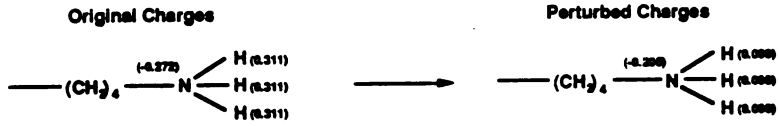
Glutamate



Histidine



Lysine



Arginine

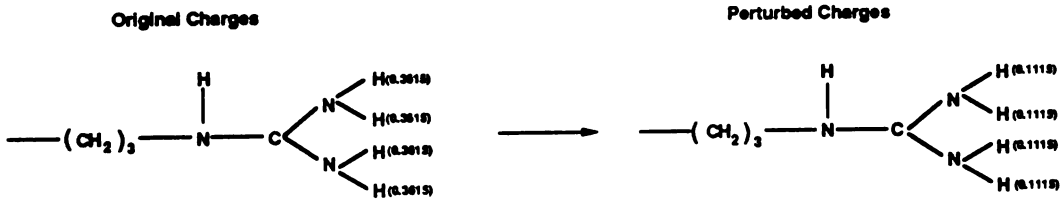


Figure 2.3. Original and Perturbed Charges for X_1-$(Ala)_{18}$-X_2 where X_1 = Glu, Asp and X_2 = His, Lys, Arg.

We also observed a differential effect of the interaction of the perturbed charge with the two ends of the helix. Comparison of the free energies for single charge perturbations at the ends of the helix reveals that the effect of introducing a charge near the carboxyl-terminus was greater than at the amino-terminus. For example, the introduction of 1/2 a positive charge at the C-terminus was 3 kcal/mole more favorable than introducing 1/2 a negative charge at the N-terminus in the unblocked α-helix structure (-8.6 vs. -5.6 kcal/mole, second column of Table 2.1). When the charges were reversed in the same structure there was a 4 kcal/mole difference between the two ends (4.5 vs. 8.5 kcal/mole, Table 2.1). This suggests either that the magnitude of the charge due to the dipole is greater at the C- than the N-terminus or that a specific interaction can occur. In the case of our model this effect can be rationalized due to the difference in the magnitude of the partial charges on the carbonyl oxygens and the amide hydrogens; the oxygens are more negative (-0.500 charge units) than the hydrogens are positive (0.248 charge units).

It is also noteworthy that the blocking groups interacted with the poles of the helix and influenced the outcome of the calculated free energy changes for perturbing charges in these positions. (Compare the last two columns of Table 2.1.) In fact, for two of the perturbations the free energy change doubled when the blocking groups were removed. The differences between the blocked and unblocked helices were most striking at the C-terminus. The proximity of the terminal methyl-amino blocking group made the introduction of a positive charge at position 20 less favourable than in the absence of the blocking group. Likewise, it was 4.3 kcal/mole more favourable to introduce half a negative charge next to the N-methyl group than in the same position in the absence of the blocking group. These results can be rationalized in terms of interactions between the negatively charged C_β of residue 20 and the methyl-amino hydrogen. (The distance between the two is 2.93 Å.) In the case of introducing a negative charge the process is more favorable with blocking groups than without because the negatively charged C_β interacts with the methyl-amino hydrogen. Conversely, introducing a positive charge is

TABLE 2.1
Free Energy Changes For Perturbing Side Chain Charges of Polyalanine (kcal/mole)

α-Helix			
Perturbed Residue No.	Perturbed C_β Charge	ΔG^a	ΔG^b
1,20	-1,+1	-23.6 ± 0.5	-29.9 ± 0.4
1,20	-0.5,+0.5	-11 ± 2	-13.9 ± 0.3
1,20	+1,-1	12 ± 2	24.0 ± 0.1
1	-0.5	-4.3 ± 0.2	-5.6 ± 0.1
20	+0.5	-7 ± 2	-8.6 ± 0.2
1	+0.5	3.2 ± 0.2	4.5 ± 0.1
20	-0.5	4.2 ± 0.2	8.5 ± 0.1
10	+0.5	-4.0 ± 0.2	-4.0 ± 0.2
10	-0.5	3.0 ± 0.2	3.2 ± 0.1
6	+1	-4 ± 1	-7 ± 1
6	-1	-2.9 ± 0.9	3.0 ± 0.1
Antiparallel β-Strand			
Perturbed Residue No.	Perturbed C_β Charge	ΔG^a	ΔG^b
1,20	-1,+1	1.7 ± 0.3	-7.2 ± 0.1
1,20	-0.5,+0.5	1.2 ± 0.1	-3.4 ± 0.1
1,20	+1,-1	-4.3 ± 0.2	5.6 ± 0.1
1	-0.5	1.4 ± 0.1	-1.4 ± 0.1
20	+0.5	-0.2 ± 0.1	-2.0 ± 0.1
10	+0.5	-0.4 ± 0.2	-0.4 ± 0.1
10	-0.5	0.1 ± 0.2	0.2 ± 0.1
Parallel β-Strand			
Perturbed Residue No.	Perturbed C_β Charge	ΔG^a	ΔG^b
1,20	-1,+1	-0.3 ± 0.2	-6.9 ± 0.1
1,20	+1,-1	-7.4 ± 0.7	5.7 ± 0.1
1	-0.5	0.3 ± 0.1	-1.6 ± 0.1
20	+0.5	-0.1 ± 0.1	-1.7 ± 0.1
10	+0.5	-0.3 ± 0.1	-0.2 ± 0.1
10	-0.5	0.0 ± 0.1	-0.1 ± 0.1

[a]Neutral blocking groups on ends of structures, dihedral angles of β-strands restrained.
[b]Dihedral angles restrained, no blocking groups.

more difficult with the blocked structure due to the proximity of the positive charge on the methyl-amino hydrogen of the blocking group. Thus, these neutral blocking groups may influence helix properties more than one might have expected.

The results for introducing single charges near the center of the helix are also listed in Table 2.1. The free energies for perturbing the charge of residue 10 were independent of the blocking groups; different free energies were obtained depending on whether blocking groups were present for perturbation of residue 6. There was a clear preference for a positive charge at both of these positions in the structures both with and without blocking groups, although it was also favourable to introduce a negative charge at residue 6 in the blocked helix. The addition of a negative charge at residue 6 was presumably favoured because of the proximity to the positive pole of the helix. The preference for a positive charge in these positions can be rationalized in terms of the orientation of the methyl groups. The methyl groups are on the outside of the helix and are inclined towards the N-terminus. This orientation places the methyl groups closer to the main-chain carbonyl oxygens on the preceding residues in the turn of the helix than the neighboring amide hydrogens. For example, the distance between C_β of residue 10 and the carbonyl oxygen of residue 6 is 4.40 Å; while only 3.55 Å separates the C_β of residue 10 and the oxygen of residue 7. However, the distance between C_β of residue 10 and the amide proton of residue 14 is 5.06 Å. Irrespective of the sign of the introduced charge, all of the calculated free energies for introducing a charge near the center of the helix were almost an order of magnitude lower than for charge interactions at the poles of the helix.

β-Strands

Calculations similar to those described above for charge interactions in the α-helix were performed on antiparallel, parallel, and extended β-strands; the results of these calculations are presented in Figure 2.4 and in Table 2.1. Comparison of free energies for various perturbations with and without blocking groups, shows that there were substan-

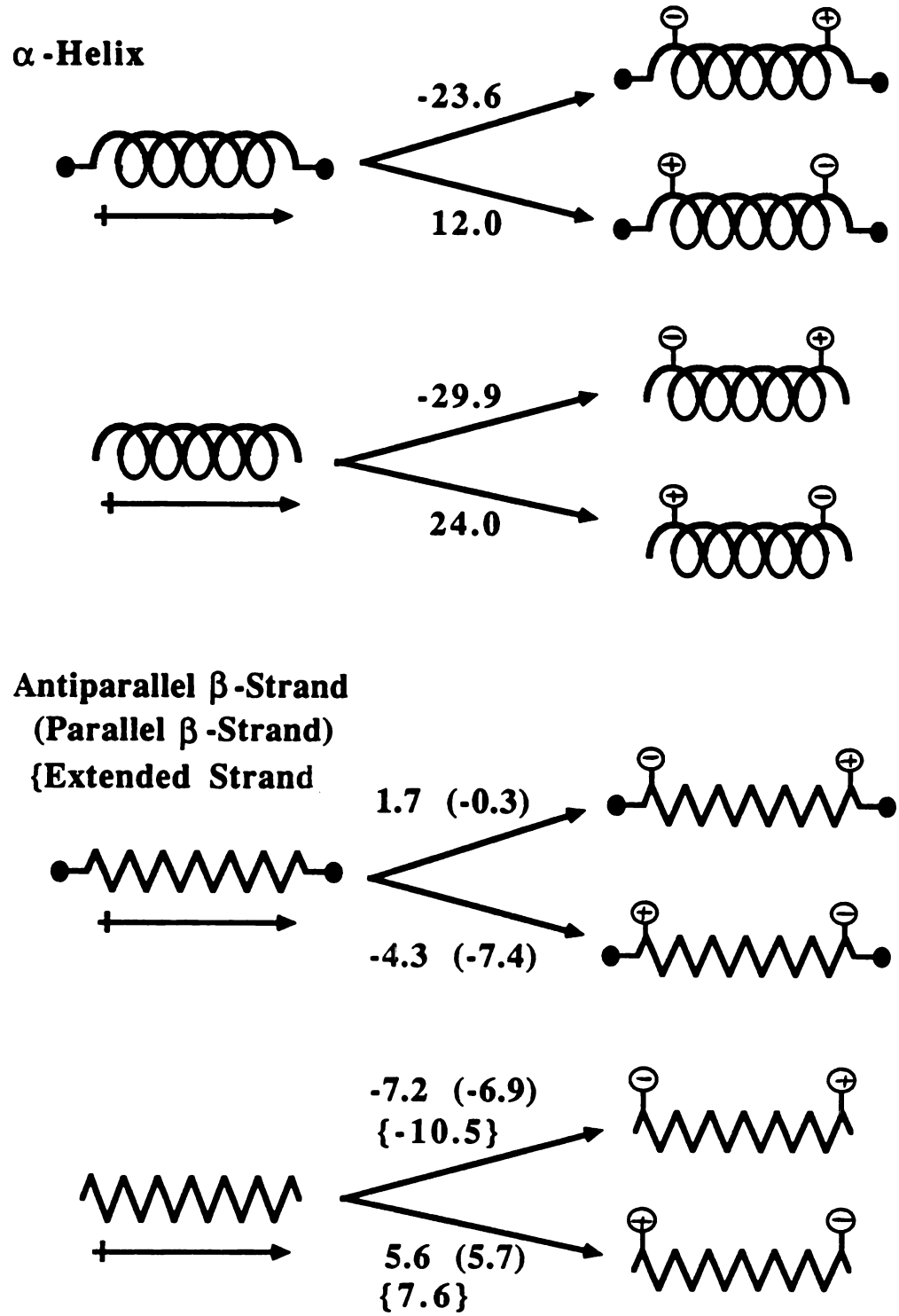


Figure 2.4. Free Energy Changes for Perturbing Side Chain Charges (kcal/mole). The free energy changes are for the introduction of full charges in the positions indicated. Solid circles represent blocking groups.

tial interactions between the blocking groups and the charges at the ends of the strands. In fact, a number of the free energies change sign upon removal of the blocking groups. We found considerable free energy changes for charges interacting with the ends of the unblocked strands (Table 2.1); like the α-helix, the β-structures preferred the introduction of negative charges at the amino-terminus and positive charges at the carboxyl-terminus. Charge perturbations near the center of the β-strands (residue 10) yielded low free energies (< 0.5 kcal/mole) and, unlike the same perturbation in the helix, did not show charge discrimination.

Comparison of α-Helix and β-Strands

A comparison of the results for any particular perturbation of both the unblocked α-helix and β-strand structures indicates that the introduction of negative charges near the N-terminus of the peptide and positive charges near the C-terminus stabilized the helix to a greater extent than the β structures. (See Table 2.1 and Figure 2.4.) Furthermore, the reversal of these charges resulted in positive free energies in the unblocked structures. Again, the effect was greater for the α-helix than for the β-strands. These results are consistent with the helix dipole model, although the magnitudes of the calculated free energies are too large to be realistic.

Sampling Dependence

Table 2.2 lists the results for varying computational parameters for a particular perturbation (Residue 1 C_β charge $\rightarrow$ -1; Residue 20 C_β charge $\rightarrow$ +1) of both a blocked and an unblocked α-helix. The free energies varied somewhat with the length of the simulation time and the time for data collection. The free energies for perturbation of the blocked α-helix converged faster than the unblocked structure. The favorable interactions between the charges introduced at the ends of the helix and the blocking groups protect the ends of the helix against fraying. The unblocked α-helix was more mobile than the blocked structure and the final structures after perturbation can differ quite a bit; the RMS deviation between the final unblocked perturbed structures of the 49 and 97

TABLE 2.2
Calculated Free Energies as a Function of Computational Parameters[a]

Unconstrained Helix with Blocking Groups				
Number of Windows	Steps of Equilibration	Steps of Data Collection	Total Time (psec)	ΔG (kcal/mole)
5	200	400	3.00	-24.7 ± 2.0
11	350	800	12.65	-26.0 ± 1.0
21	200	400	12.60	-28.1 ± 1.0
41	400	800	49.20	-28.9 ± 0.4
81	400	800	97.20	-28.2 ± 0.2
Unconstrained Helix without Blocking Groups				
Number of Windows	Steps of Equilibration	Steps of Data Collection	Total Time (psec)	ΔG (kcal/mole)
5	200	400	3.00	-31.3 ± 1.0
21	200	400	12.60	-33.0 ± 0.9
41	400	800	49.20	-34.6 ± 0.5
81	400	800	97.20	-37.6 ± 0.1

[a] Equilibrated polyalanine models (25 psec), with and without blocking groups, are used as the starting structures for perturbation. The same perturbation is performed for each combination of computational parameters: the C_β charge of residue 1 → -1 and the C_β charge of residue 20 → +1.

psec simulations was 1.29 Å (matching all atoms). Hence, with systems that can undergo conformational changes the free energies as a function of simulation time may not converge because they do not converge on the same structure. Another point worth mentioning is that the total time of the perturbation and the degree of sampling at a particular window were less important in approaching the consensus ΔG than the number of windows used for the perturbation. (Compare the 11 versus 21 window runs for the unconstrained helix.)

Despite the difference in the calculated free energy changes in Table 2.2, the results are similar enough, and the trends do not change, to have confidence in the results from short simulations. The free energies in Table 2.2 differ from those presented in Table 2.1 at least in part because the dihedral angles of the structures used for the calculations in Table 2.2 were not restrained. The use of unconstrained structures and longer simulation times led to slightly different final structures for the various simulations.

Magnitude of Calculated Free Energy Changes

Although our results are qualitatively consistent with the helix dipole model, the magnitude of the calculated free energies is too large when compared with experiment. While accurate experimental free energies are not yet available, the free energy changes are expected to be approximately 0.5-2.0 kcal/mole, since the apparent helical content can be altered with small changes in temperature, pH or salt concentration (Shoemaker et al., 1985, 1987). That we calculated free energies of much greater magnitude is almost entirely due to two factors: the distance between the charge and the helix dipole and the use of a low effective dielectric constant. Both of these points are addressed below.

Distance Dependence

The large calculated free energy changes were partially due to the initial placement of the charge(s). We perturbed the charges of particular alanine methyl groups, therefore the introduced charges were very near the helix axis. To simulate more realistic distances of the charge from the axis, we increased the C_α-C_β bond lengths. A plot of the

calculated free energy as a function of C_α-C_β bond length appears in Figure 2.5. The free energy changes decreased in magnitude as the side chain charge got farther from the helix axis. Given these results, short side chains, or longer mobile side chains that allow close approach of the charged functional group to the helix dipole, should be effective in stabilizing or destabilizing helices.

Dielectric Dependence

Free energy changes were calculated using a dielectric constant of 80 for one particular perturbation, the charge of residue 1 was perturbed to -1 and that of residue 20 to +1, in the α-helix and the β-structures. The free energy change for this perturbation in the α-helix was -1.0 kcal/mole. When this same perturbation was carried out on the parallel and antiparallel β-strands, a ΔG of -0.4 kcal/mole was obtained. These values are much closer to what one would expect experimentally, although they probably represent lower bounds for the free energies.

Charge Distribution Dependence

Having shown that the free energy changes decrease in magnitude as the "pseudo bond" distance between the C_α and C_β carbons increases, we examined a more realistic model based on the six possible charged combinations of an helix of the form X_1-$(Ala)_{18}$-X_2---where X_1=Asp, Glu and X_2= His, Lys, Arg. The free energies for neutralizing the charges of these amino acids were calculated. The results are presented in Table 2.3. All of the free energies were positive since we began with favourable charge-helix dipole interactions and removed the charges on the functional group interacting with the helix dipole. (See Figure 2.3 for the original and perturbed charges used.)

In comparing the various combinations of X_1 and X_2, the most important factor in determining the magnitude of the free energy was the distance of the charged portion of the residue from the ends of the helix. It is important to note that the ϕ, ψ angles of the helix were restrained but the side chains were unhindered conformationally. The free energies for perturbing the charges of the Glu-$(Ala)_{18}$-His and Asp-$(Ala)_{18}$-His peptides

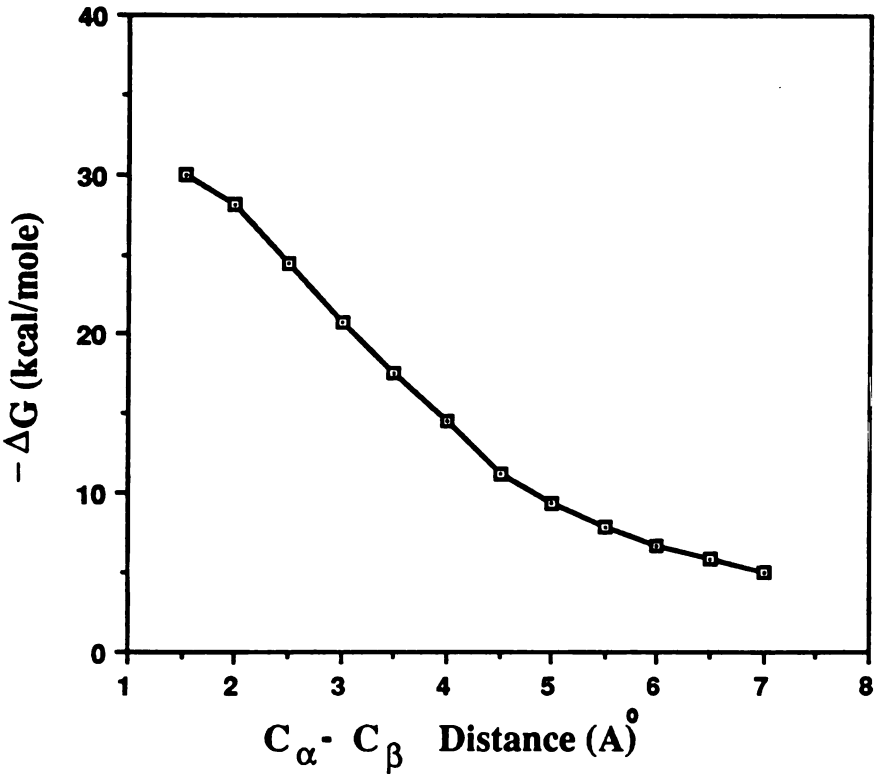


Figure 2.5. Free Energy Changes as a Function of C_α-C_β Bond Length. The free energy changes are calculated for the same perturbation (residue 1 $C_\beta \rightarrow$ -1, residue 20 $C_\beta \rightarrow$ +1) in polyalanine models with various C_α, C_β bond lengths.

were essentially the same (Table 2.3). Although the glutamate is in a position to interact more strongly with the positive pole of the helix dipole than the aspartate, the histidine is closer to the negative pole of the helix in the Asp, His peptide than in the Glu, His peptide. Similar compensation can be seen when comparing the other combinations. The glutamate of the Glu-$(Ala)_{18}$-Lys peptide is 0.4 Å closer to the helix than the aspartate of the Asp-$(Ala)_{18}$-Lys helix. The Asp-$(Ala)_{18}$-Lys peptide was stabilized by these charge-dipole interactions by almost 2 kcal/mole more than the Glu, Lys combination. The reason for this appears to be that the lysine in the Asp-$(Ala)_{18}$-Lys peptide is 2.3 Å closer to the negative pole of the helix than in the Glu, Lys case. A similar argument can explain the stabilization of the Glu-$(Ala)_{18}$-Arg helix compared to the Asp-$(Ala)_{18}$-Arg peptide, although in this case the arginine is roughly the same distance from the negative end of the helix dipole in both peptides while the glutamate is 1.2 Å closer to the N-terminus than is the aspartate. Nevertheless, perturbation of the charges of actual amino acid residues gives qualitatively similar results to those obtained for increasing the C_α , C_β bond lengths. The distance between the charge and the poles of the helix dipole is critical in explaining the stabilization conferred to a structure by virtue of charge-dipole interactions. It is difficult to anticipate, however, the positions of conformationally mobile residues.

Salt Bridge Calculations

The free energy changes for salt bridge formation along an α-helix were calculated by simultaneously perturbing the C_β charge of two residues, either three or four residues apart. It is important to note that the calculated free energy changes represent interactions between the salt bridge and the helix but do not include interactions between the two perturbed groups (e.g. charge-charge interactions). The charge-charge interactions can be estimated using the following relationship:

$$\Delta G = 332\, q_1\, q_2 / r_{1,2}\, \varepsilon \qquad \text{Eq. 2}$$

The average C_β-C_β distance for the *i*→*i+4* salt bridges is 6.00 Å, which yields a free

TABLE 2.3
Free Energies for Perturbing Charge Distribution of the End Residues of X_1-$(Ala)_{18}$-X_2^a

X_1	X_2	ΔG	Dist. to Helix[b]	
			X_1	X_2
Glu	His	10.4 ± 0.7	2.5	4.1
Glu	Lys	6.6 ± 0.3	3.0	4.9
Glu	Arg	8.0 ± 0.2	2.5	3.4
Asp	His	11.0 ± 0.7	4.3	3.4
Asp	Lys	8.5 ± 0.1	3.4	2.6
Asp	Arg	6.3 ± 0.5	3.7	3.6

[a] The calculated free energy change is for the simultaneous perturbation of X_1 and X_2 as shown in Figure 2.6.

[b] This is the closest distance (in Å) between perturbed atoms of X_1 and X_2 and the nonhydrogen bonded amide hydrogens (residues 1-4) and carbonyl oxygens (residues 17-20) in the final structures following the perturbation simulations.

energy change of -9.2 kcal/mole for a distance dependent dielectric constant. The average C_β-C_β distance of 5.75 Å for $i \rightarrow i+3$ salt bridges yields a free energy change of -10.0 kcal/mole. These free energy changes, like those presented above for single charge perturbations, are only useful for qualitative comparisons since the values are roughly -0.7 kcal/mole if a dielectric constant of 80 is used. So, the total change in free energy for introducing a salt bridge into a structure is the sum of the charge-charge interaction and the calculated free energies in Figure 2.6 for the interactions between the salt bridge and the helix.

Figure 2.6 shows the calculated free energy changes for forming $i \rightarrow i+4$ and $i \rightarrow i+3$ salt bridges along an α-helix. Virtually all of the free energy changes for forming $i \rightarrow i+4$ salt bridges depended on the polarity of the perturbed charges, e.g. salt bridges at the ends of the helix were favoured when the salt bridge charges closest to the poles of the dipole were complementary to the dipolar charges. Salt bridges at the C-terminus of the blocked structure were an exception. These results can be rationalized in terms of electrostatic interactions between the perturbed groups and the methyl-amino blocking group. The negatively charged C_β of residue 20 can interact, as discussed above, with the positively charged hydrogen of the methyl-amino group (2.93 Å separates these atoms), making a seemingly unfavourable perturbation favourable. The same salt bridges behaved as one would predict in the absence of terminal blocking groups. That there is a preference for a (+,-) salt bridge in the interior of the helix (residues 9 and 13) over the (-,+) orientation points out the importance of local interactions, since one would expect the (-,+) salt bridge to be preferred on the basis of favourable interactions with the helix moment.

In our calculations the free energy contribution to helix stability from $i \rightarrow i+3$ salt bridges was not consistently dependent on directionality of the salt bridge. In fact, almost all of the free energies for formation of the salt bridges in models with blocking groups were favourable, regardless of whether the salt bridge was aligned with or against the charges of the helix dipole. In the absence of blocking groups, there was some

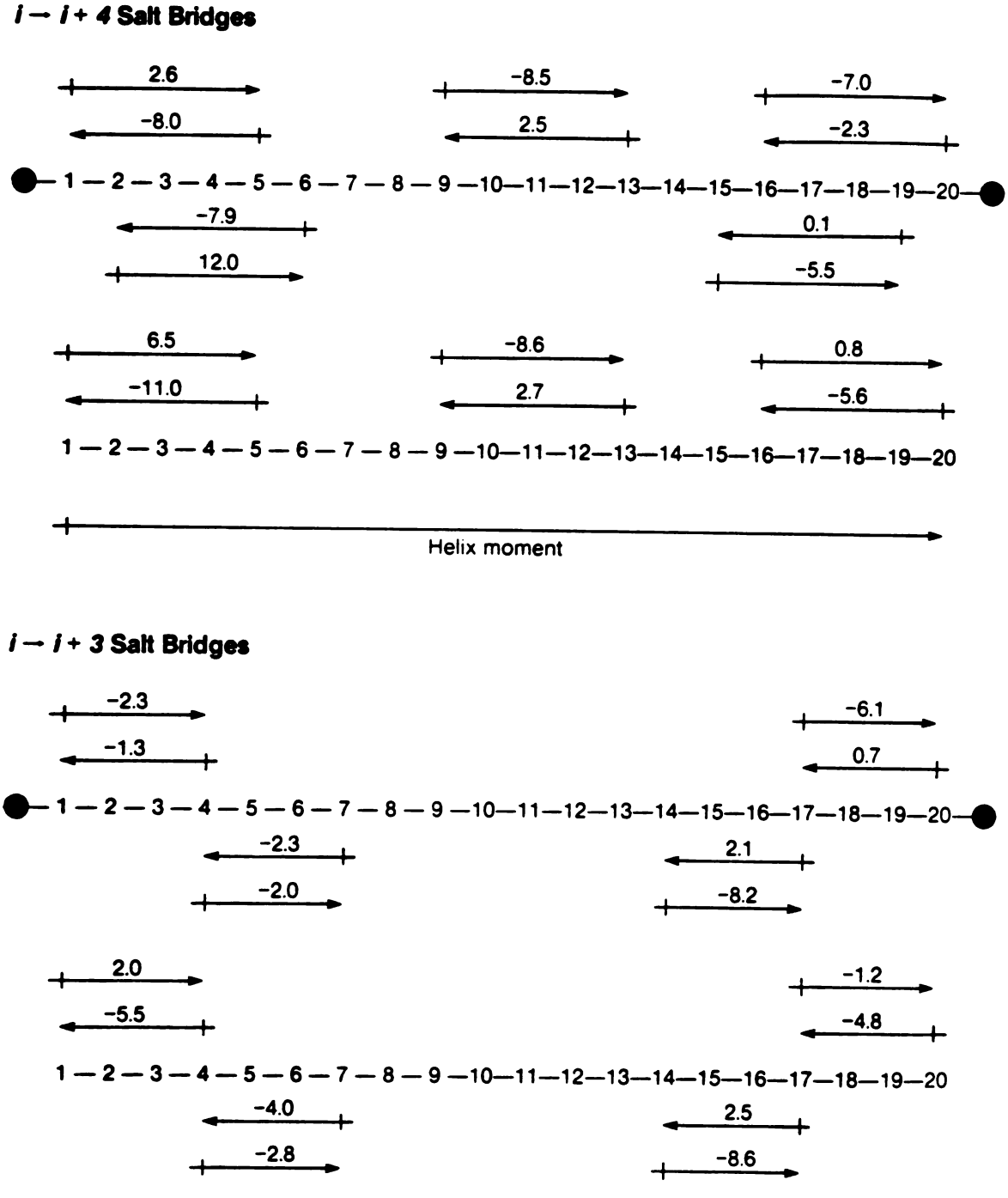


Figure 2.6. Free energy changes for forming salt bridges in an α-helix (kcal/mole). The free energy changes are for the introduction of full charges in the positions illustrated, where the arrowhead indicates the position of the negative charge. Solid circles represent blocking groups.

dependence on polarity.

With few exceptions, the formation of *i→i+4* salt bridges was more favourable than *i→i+3* salt bridges. The *i→i+4* salt bridges are essentially colinear with the macrodipole. The *i→i+3* salt bridges extend across the helical axis at angles of approximately 30 degrees. The difference in the magnitudes of the free energies between *i→i+4* and *i→i+3* salt bridges may be due to more direct interactions of the *i→i+4* salt bridges with the helix dipole. This effect is probably less important with longer side chains where the charged functional group is farther from the helix axis. Even in our simple model, though, the dominant term in determining the difference between *i→i+3* and *i→i+4* salt bridges at the ends of the helix is probably repulsive in nature. As can be seen in Figure 2.1, the region of positive electrostatic potential near the N-terminus spans four residues. So, when residue 1, for example, carries a favourable negative charge, formation of a salt bridge would favour placement of the corresponding positive charge at residue 5 over residue 4. Likewise, for the opposite, unfavourable orientation of the salt bridge (+,-), *i→i+3* salt bridges are less unfavourable than the *i→i+4* spacing, since the negative charge is still within the region of positive electrostatic potential in the fourth position.

DISCUSSION

Recent work by Baldwin and co-workers on small peptides in water, suggests that charge interactions with the helix dipole can stabilize helical structures (Shoemaker et al., 1985, 1987; Marqusee and Baldwin, 1987). They synthesized various analogs of the C-peptide of RNase A and some *de novo* peptides to examine the dependence of helix stability on the position of the charge along the helix axis as well as the stability conferred to the structure by salt bridges (Table 2.4).

The experimentally observed helix stabilities of the C-peptide homologs as a function of amino acid composition are qualitatively consistent with the calculated free energy changes presented in this paper. For example, when the charged residues interacting favourably with the helix dipole are replaced by neutral residues, helix con-

tent decreases. In fact, for the first two cases in Table 2.4 Baldwin and his co-workers find negligible amounts of helix when the N-terminal Glu is replaced by Ala and the C-terminal His is replaced by Ala (Shoemaker et al, 1985). A similar effect for an His at the C-terminus of an helix has also recently been reported by Sali *et al.* (1988). Further, Shoemaker *et al.* (1987) find that stability is dependent on the N-terminal charge (cases 4, 5 and 6 of Table 2.4), as would be predicted by our results on varying the charge of the terminal residue without blocking groups (Table 2.1).

We found that salt bridges were secondary to favourable single charge interactions with the ends of the helix, in terms of direct interactions with the dipole. For example, the introduction of a negative charge at the N-terminus and a positive charge at the C-terminus of the unblocked peptide, stabilized the structure by 29.9 kcal/mole (Table 2.1); while, the formation of an $i \rightarrow i+3$ salt bridge at each end of the helix, aligned to interact favourably with the dipole, stabilized the structure by 10.3 kcal/mole (Figure 2.6). A similar effect has been seen experimentally by Shoemaker and co-workers (1985) (case 3 of Table 2.4). When Glu 9 is replaced by leucine, helix stability increases even though Glu 9 is in a position to form a salt bridge with His 12. Bradley *et al.* (1989) report a similar role for an His residue in an helical peptide. Hence, experimental results also suggest that single charge interactions with the poles of the helix dipole stabilize helices more than multiple charges near the poles of the helix, even when these charged groups can form salt bridges.

Marqusee and Baldwin (1987) studied the role of $i \rightarrow i+3$ and $i \rightarrow i+4$ salt bridges in stabilizing alanine-based peptides. Essentially, they find that $i \rightarrow i+4$ salt bridges stabilize helices to a greater extent than do $i \rightarrow i+3$ salt bridges, and for both spacings the effect is greater when the charges are in positions to interact favourably with the helix dipole. Our calculated free energy changes for formation of salt bridges with analogous orientations and spacings (Figure 2.6) are consistent with these experimental results. Marqusee and Baldwin (1987) suggest that the difference in the ease of making hydrogen bonds

TABLE 2.4
Effects of Substitutions on C-Peptide Derivatives

Peptide[a]	Substitution	Helix Content
1. KETAAAKFERAHA[b]	H(12)→A	no helix
2.	E(2)→A	no helix
3.	E(9)→L	↑
4. Ac-AETAAAKFLAAHA[c]	Ac-A→succinyl-A	↑
5.	Ac-A→A	↓
6.	Ac-A→K	↓

[a] The C-terminus of each of the peptides is amidated. Ac represents an acetyl group.

[b] Shoemaker et al., 1985.

[c] Shoemaker et al., 1987.

between preferred rotamer conformations of glutamate and lysine may explain why an $i \rightarrow i+4$ salt bridge is more helix stabilizing than an $i \rightarrow i+3$ orientation. Our results on salt bridge calculations in polyalanine suggest that the spacing of the salt bridge itself is important in determining the stability conferred to the helix, since we do not have different side chain conformations to consider.

As mentioned in the results section, we found asymmetry of the helix---the magnitude due to the charge near the C-terminus was greater that at N-terminus. This may be an inherent property of the helix dipole and not just a property of our peptide model. Both experimental (Wolfenden, 1978) and theoretical (Bash et al., 1987a) studies on the solvation of acetamide and N-methylated derivatives suggest that the N-H hydrogen bonding potential of the amide group is relatively unimportant in determining the solvation properties of these compounds and that the interaction of water with the carbonyl group is the major factor in determining the very negative solvation free energy of the amide group.

We found that charge interactions with both the antiparallel and the parallel β-strands yield much lower, but significant, free energies than the analogous perturbations of the α-helix. Wada (1976) points out that one-third of the residue moment lies along the axis of both antiparallel and parallel β-strands. We found no evidence, however, for a macrodipole in the β-structures. Perturbation of an extended β-strand yielded free energy changes of the same magnitude as the antiparallel and parallel β-strands. There should only be minimal alignment of the dipoles in the extended orientation. That we calculated similar free energy changes for all three of the β structures suggests that we may only be observing an end effect, providing further support for the importance of local interactions. To further complicate the matter, the single strand β-structure geometries were not stable; even with constraints the strands twist and bend a great deal.

A comparison between the calculated free energies and what one would expect given static structures gives some idea of the deviations from ideality. For example, the

interaction between an ideal α-helix (represented as a line dipole) and favourable charges 1.5 Å from the poles of the dipole yield interaction energies of -42.4 kcal/mole, neglecting screening and assuming complete alignment of the dipoles. In contrast, we calculated free energy changes of -29.9 kcal/mole for the same interaction using the perturbation method. The difference between the two values can easily be reconciled given the use of $\varepsilon=1$ for the interaction of charges with the static structure and a distance dependent dielectric constant using the perturbation method. Another factor contributing to this discrepancy is the local perturbation in structures that occurs during dynamics. Using Wada's (1976) values for the extent of axial residue dipole moment alignment (μ_{ax} / μ_{res} = 0.376 for parallel β-strands, 0.312 for antiparallel strands, and 0.127 for extended strands), results in interaction energies of -3.0, -2.5, and -1.0 kcal/mole for parallel, antiparallel, and extended β-strands, respectively. These values obviously differ markedly from the free energy changes given in Table 2.1 and Figure 2.4. Hence, the β-structures are probably too mobile, even when constrained, to detect such subtle differences in the alignment of the dipoles and instead specific local interactions dominate the calculated free energies.

In proteins there is a clear preference for negatively charged residues to occur near the N-terminus of an α-helix and positively charged residues near the C-terminus. The results of two commonly referenced studies, cataloguing these tendencies, agree that there is a greater probability for glutamic acid to be found at the N-terminus (first 3 residues) than aspartic acid (Chou and Fasman, 1978; Crawford et al., 1973). These studies differ in the probabilities of occurrence of positively charged residues at the C-terminus, though. Chou and Fasman (1978) find that lysine is favoured over histidine and arginine follows, although there are intervening residues; in contrast, Crawford and co-workers (1973) have in decreasing order of probability His > Lys > Arg. On the basis of the results for calculated free energies as a function of the distance between C_α and C_β (Figure 2.4), we would expect asparatate to be favoured over glutamate at the N-terminus and

His > Lys > Arg at the C-terminus. However, in our simulations with actual charged amino acids at the ends of the helix, glutamate is able to more closely approach the poles of the helix than is aspartate. Glutamate may be favoured over aspartate for this reason. The differences between the ordering of residues at the C-terminus between the two sources may be due to differences in their definitions of turns and helices and insufficient sampling [e.g. Chou and Fasman (1978) examined 29 proteins and Crawford *et al.* (1973) used a basis set of 11 proteins]. Our results with histidine, lysine and arginine at the C-terminus of an α-helix do not allow discrimination between the two sources. Instead, we find that the stability conferred to a structure by charge-dipole interactions is determined by how closely the charged functional group can approach the poles of the helix. Further and more extensive studies are necessary, both in terms of simulations and in cataloging the tendencies for charged residues at the ends of helices with a larger basis set of proteins, to alleviate ambiguities.

Essentially all of our results are consistent with the helix dipole model and qualitatively consistent with the available experimental data. While the magnitudes for the calculated free energies using the dielectric model ε=r and formal charges on the β carbons are too large to be experimentally reasonable, they drop to more moderate values as the distance between the charge and dipole increases. To model these interactions more rigorously, counterions on the charges and solvent should be included. There are difficulties in doing this, however. The calculated free energies for perturbing the charges of an helix in water include the interaction of the charge with the peptide and a much larger term due to reorientation of water molecules that accompanies "growing" a charge in water. The huge interaction energy between the charge and the water molecules overwhelms the relatively small interaction between the charge and the peptide. Therefore we need a control case to compensate for the reaction field term. The β structures could perform this control function but we would be left with the problem of looking for small differences between large numbers. The simulations are not computa-

tionally feasible, either, because the free energies for introducing charges into water converge very slowly. Weighing these factors the simpler case of $\varepsilon=r$ allows us to qualitatively evaluate different interactions and test limiting cases by the use of different dielectric models.

Why not just use Coulomb's law to calculate electrostatic energies? With idealized geometries we cannot realistically evaluate any subtle local geometry changes that might accompany the introduction of charged residues. If we use molecular mechanical minimization and compare internal energies, the uncertainty in these energies is significantly greater than in free energies (Bash et al., 1987a,b). Thus, it makes sense to use the free energy approach in a semi-quantitative application such as this. We emphasize that the computer time required for these simulations is quite modest, a typical mutation in Table 2.1 requires less than 30 minutes on a FPS 264 array processor.

Also, as mentioned above, to model these electrostatic interactions more realistically, in addition to solvent, counterions should be included in the system. However, the addition of counterions leads to further complications; besides screening the charge on the methyl groups of the peptide, the counterions interact with the structure. On another level, Matthew (1985) points out that it is not rigorously correct to uniformly assign the same set of partial charges to peptide dipoles throughout the structure, as the microenvironment around particular dipoles is probably different and should be taken into account to obtain accurate results. Given the simple system examined here with all residues exposed, the use of a single set of partial charges is reasonable.

The use of a distance dependent dielectric constant ($\varepsilon=r_{i,j}$), which is expected to partially compensate for the absence of solvent, leads to an overestimation of the effect of changing a full charge and provides another explanation for the large calculated free energies. The commonly accepted value for the dielectric constant of the protein interior is between 1 and 5 (Pethig, 1979). However, recent experimental work (Rees, 1980; Russell and Fersht, 1987; Russell et al., 1987) suggests that the effective dielectric con-

stant of the protein interior might be between 40 and 50, although this is probably due to interactions through the solvent. So, using a distance dependent dielectric constant, which yields an average dielectric constant of approximately 5 for our system, significantly underestimates the influence of solvent. The calculations using a larger dielectric constant of 80 yield free energies that are more reasonable.

Although more reasonable free energies are obtained by using ε=80, there are problems with using such a high dielectric constant in this simple manner. Under these conditions the α-helix will fall apart during dynamics (with the force field within AMBER) due to the diminution of the strength of the hydrogen bonds. For example, the average electrostatic component to the hydrogen bond in the α-helix is worth 4.6 kcal/mole with ε=r but drops to 0.17 kcal/mole when ε=80. Another problem with a high uniform dielectric constant is that interactions through the peptide and through the water immediately surrounding the peptide are treated like bulk water, leading to an underestimation of the free energy change for a process.

Recently, continuum dielectric models have been successful at modeling shifts in pK_a of a particular residue as a result of site-specific mutations (Gilson and Honig, 1987; Sternberg et al., 1987). With continuum, or cavity, models the protein is considered a cavity of some given shape with a low dielectric constant (usually ε=2-4) in a solvent with an high dielectric constant (ε=80). It is conceivable that such an approach would provide more reasonable absolute free energy changes than the method that we employed but it would be at the expense of molecular detail. There are also a number of other drawbacks to using continuum models. Calculations using such models generally require very large amounts of computer time; while, the free energy perturbation method allows us to quickly calculate free energies for our system. Also, most charges are at the interface between the high dielectric solvent and the lower value of the protein in cavity models. This interface, or dielectric boundary, is where the fields are most sensitively dependent upon position, and hence where continuum models are at their weakest (Rogers, 1986).

For an isolated helix it is precisely this region that is of interest. Many other peculiarities of cavity dielectric models have been discussed by Rogers (1986). So, although cavity models might give more reasonable free energies for charge dipole interactions, it is at the expense of molecular detail. It is molecular detail that we are interested in and we showed a number of cases where specific local interactions are important. Also, as our model becomes more sophisticated, by using varied amino sequences, other contributions to the free energy, such as van der Waals forces, may become important; continuum dielectric models do not account for other than electrostatic interactions.

Despite the shortcomings mentioned here, further refinement of our model should address the problem of the magnitude of the free energies, but the sign of the energies would not be expected to change.

CHAPTER 3: Molecular Dynamics Simulations of Small Peptides: Dependence on Dielectric Model and pH

The α-helix is the most abundant form of secondary structure found in globular proteins (Creighton, 1983) and may also be important in directing protein folding. The framework model of protein folding proposes that secondary structure is formed early in folding and that these preformed, marginally stable units of secondary structure coalesce to form tertiary structure (Ptitsyn and Rashin, 1975; Karplus and Weaver, 1976; Kim and Baldwin, 1982). This hypothesis implies that secondary structure should be present under conditions where folding occurs spontaneously, but until recently most attempts to detect secondary structure within small peptides in water were unsuccessful (Epand and Scheraga, 1968). In 1971 Brown and Klee were able to detect α-helical structure in the 13 residue C-peptide from ribonuclease A, which was contrary to the prevailing dogma. Since that time much work has been carried out on the C-peptide to determine the interactions responsible for the unexpected helical stability (Shoemaker et al., 1985, 1987). A recent study by Bradley and co-workers (1990) expands upon this idea and describes peptides of *de novo* design that are helical in aqueous solution. We performed simulations of one of these peptides, under a variety of conditions, in an attempt to reproduce both the qualitative and semi-quantitative experimental results and elucidate interactions important for stabilizing the helical conformation.

We chose to study the peptide with the highest helix content observed by Bradley et al. (1990). The peptide has blocked ends (acetylated amino-terminus and amidated carboxy-terminus) with the $i \rightarrow i+4$ spacing of potential salt bridges, with the exception of one possible $i \rightarrow i+3$ interaction near the C-terminus (Glu 13...His 16). Figure 3.1 shows the sequence of this peptide along with the experimentally observed connectivities determined using two-dimensional nuclear Overhauser effect spectroscopy (NOESY). The origin of the nuclear Overhauser effect is dipolar cross-relaxation between protons, thereby measuring through space interactions. This peptide shows its greatest helical

content at 5 degrees Celsius and low pH (between 2 and 5 pH units). For this reason, Bradley and co-workers rule out salt bridges, which are expected to be maximal near neutral pH, as the dominant mode of stabilization of the helix. Instead, they suggest that the pH dependence arises from interactions between His 16 and the negative pole of the α-helix macrodipole.

We performed molecular dynamics (MD) simulations of the peptide in Figure 3.1, using an empirical force field (Weiner et al., 1984). The electrostatic interactions in this force field are described by Coulomb's law ($E=Q_iQ_j / \varepsilon R_{i,j}$, where Q is the charge and R is the distance between atoms i and j). Often, the electrostatics are modelled by setting $\varepsilon=1$ and including explicit water molecules. In another approach, the macroscopic dielectric constant, ε, is treated as an adjustable function, ε(R), such as $\varepsilon=R_{i,j}$, $\varepsilon=4R_{i,j}$, or $\varepsilon=4$ to compensate for the lack of water. Other workers have used slightly more sophisticated functions: [$\varepsilon = (r-1)^2 + 2$] (Warshel, 1979) and [$\varepsilon = 4 \exp(0.1r)$] (Srinivasan and Olson, 1980). All of these distance-dependent functions have the aesthetic disadvantage that they do not reach a limiting value at the bulk dielectric constant of water. Inclusion of water molecules provides the most realistic system for studying biologically relevant molecules, but there are often situations where it is computationally impractical to include water explicitly.

In an earlier study aimed at calculating free energies for charge-helix dipole interactions, we found that $\varepsilon=R$ severely overestimated electrostatic interactions in small peptides (Chapter 2). Since small peptides are entirely exposed to solvent, the dielectric constant should be quite high and approach the value for bulk water ($\varepsilon=80$) at relatively short distances from the charge. The function $\varepsilon=R$ does not screen electrostatic interactions enough to mimic this effect for small peptides in solution, nor is it effective for surface regions of proteins. Setting $\varepsilon=4R$ screens long-range electrostatic interactions more effectively than $\varepsilon=R$ but it overscreens short-range interactions. In our force field, hydrogen bonds are predominantly electrostatic in nature. Both $\varepsilon=R$ and $\varepsilon=1$ reproduce hydro-

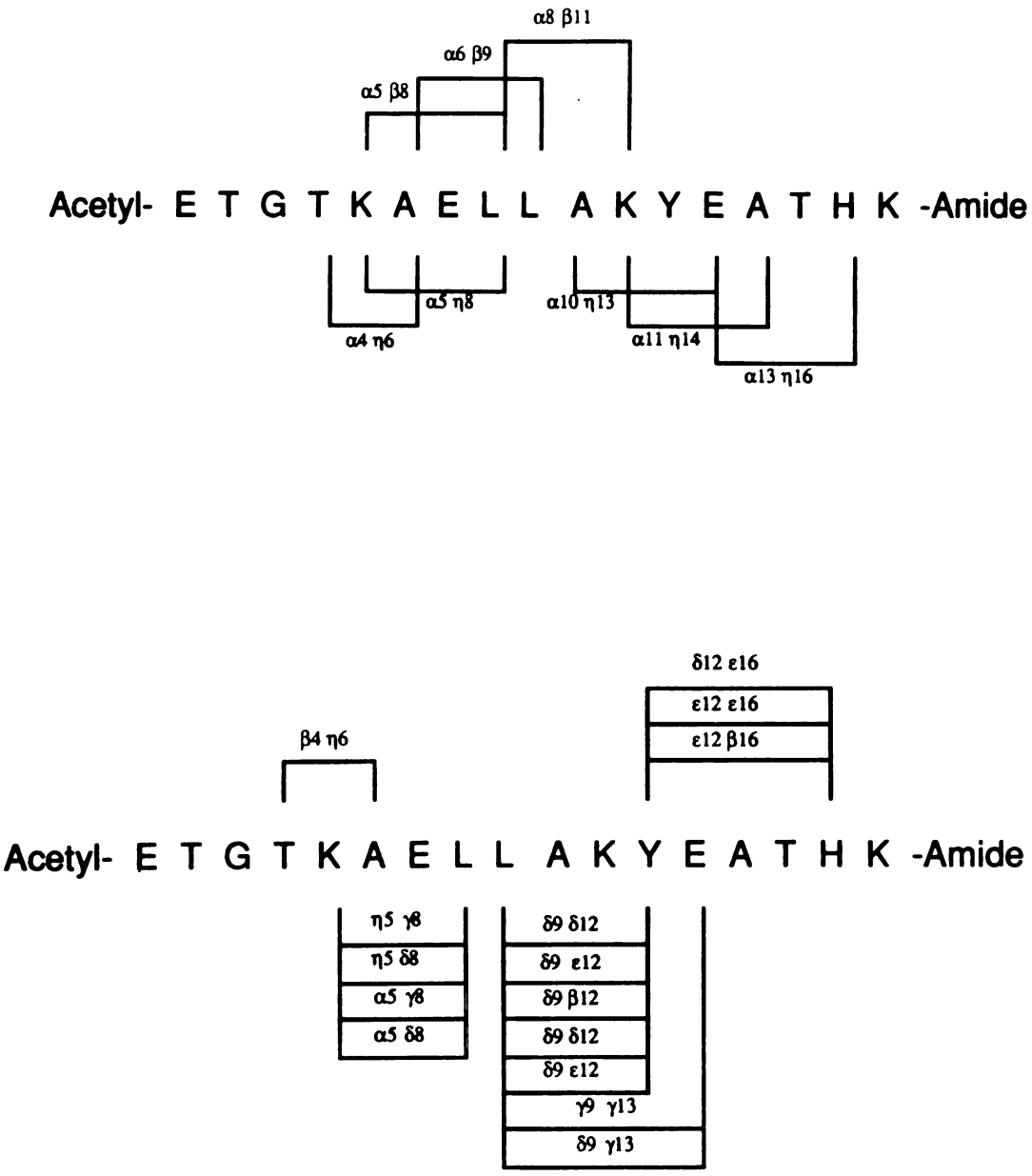


Figure 3.1. The primary sequence of the 17-residue peptide and 2-dimensional NMR data at 5° C and pH 2. (See Bradley, 1990 for experimental details.) A) Mainchain NOESY connectivities. B) Side chain NOESY connectivities.

gen bond lengths in a variety of systems while $\varepsilon=4R$ does not (Weiner et al., 1984). Use of a uniform dielectric constant of 80 leads to a loss of hydrogen bonds (Chapter 2). We wanted a function that could account for both the short- and long-range interactions in a physically meaningful way to give $\varepsilon=1$-2 at short distances and $\varepsilon=80$ above 10-15 Å. No one of the functions mentioned above fulfills these criteria. For this reason we implemented the distance-dependent dielectric function of Ramstein and Lavery (1988), which is a modified form of the function derived by Hingerty and co-workers (1985).

The function developed by Hingerty and co-workers is based on Debye's theory of ionic saturation (Debye, 1929). Debye's formulations for the distance dependence of the effective dielectric constant were derived by solving Maxwell's equations for an element of volume and an ion separated by a solvent of isotropic polarizability. This solvent approximation is not valid for water, but it appears to be insensitive to the explicit form of the dielectric function as long as the function increases to 80 quickly (Hingerty et al., 1985). The relationship derived by Debye is only valid for distances greater than 3 Å. Hingerty et al. (1985) extended this work to include distances below 3 Å by using the method of image charges (Perutz, 1978) for a charge in a cavity immersed in water. They then combined the two regions graphically to give a single expression for all distances.

The general form of the function derived by Hingerty and co-workers (1985) is shown in Figure 3.2. The sigmoidal distance dependence of this dielectric function can be rationalized by considering the structure of water molecules near charges. The spatial ordering of liquid water can be determined using X-ray diffraction. Using this technique, Norten and Levy (1969) have shown that the radial distribution function of liquid water is quite variable at short distances and it approaches one around 8-10 Å from the probe molecule. This suggests that water molecules show nonrandom arrangements for 2 to 3 molecules from the probe. (These ideas and others related to water in the vicinity of macromolecules have been reviewed by Cooke and Kuntz (1974).) In any case, since the dominant contribution to the dielectric constant under normal conditions is due to

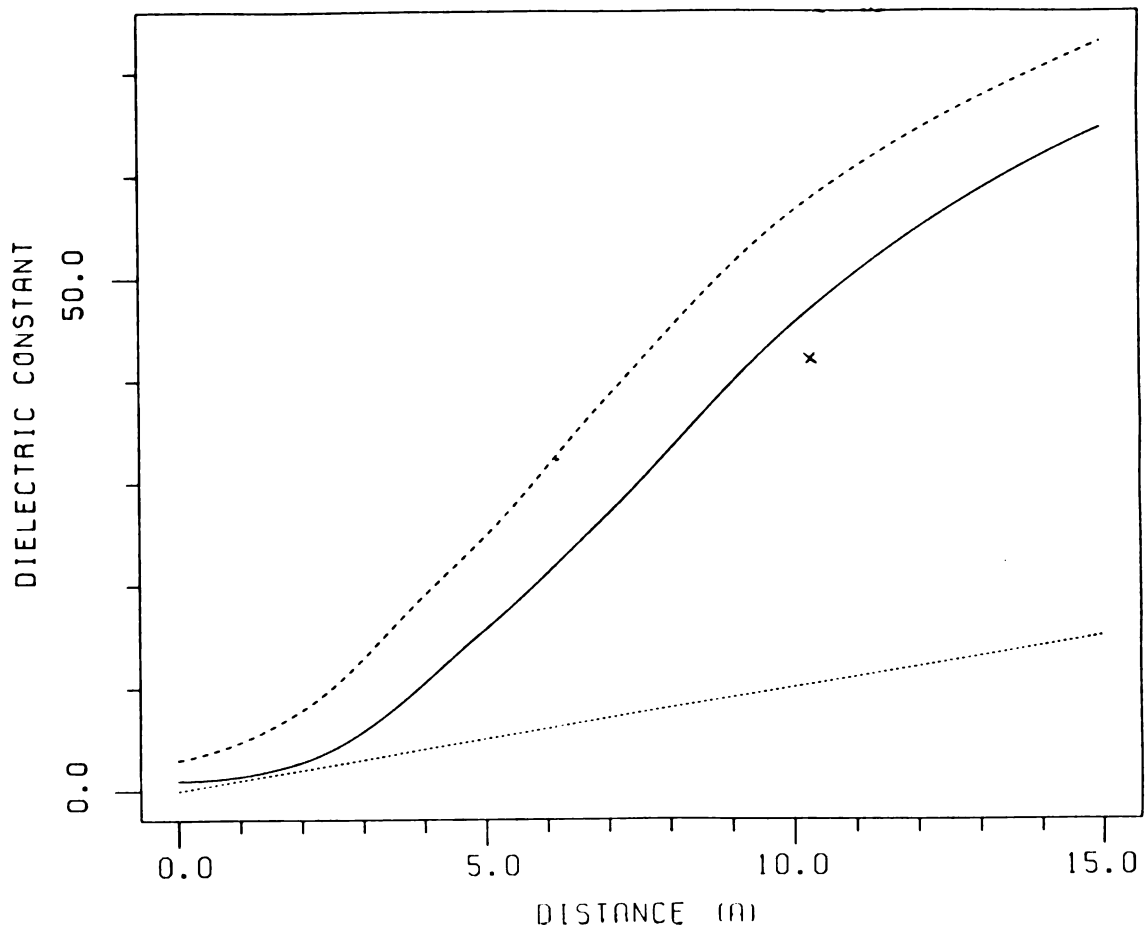


Figure 3.2. The dielectric constant as a function of distance from a charge. The solid line (——) represents the sigmoidal distance dependent dielectric function given in equation 1, with D=78 and S=0.3. (— — —) represents an experimentally derived curve (Conway, 1951) and (- - - -) denotes the linear dielectric function ($\varepsilon=r$). The X represents the average distance and average effective dielectric contstant as estimated by Rees (1980) (only the values for lysines less than 15 Å from the heme in cytochrome c were used).

reorientation of dipoles, the dielectric constant will be low where the dipoles are restricted, e.g. closest to the charge. The dielectric constant increases moving away from the charge as the water molecules become less restricted rotationally and quickly approaches the bulk value with loss of structure.

Hingerty and co-workers (1985) have studied metal ion binding to RNA using their sigmoidal dielectric function, yielding results in good agreement with experiment. Ramstein and Lavery (1988) modified this function to easily accomodate other solvents and to include a parameter to control how quickly the dielectric constant changes with distance. They have applied their function to study conformational changes in DNA, again with good agreement with experiment. To our knowledge, neither function has been applied to peptides or proteins, however there is some precedent for using this type of sigmoidal dielectric function for calculating electrostatic interactions in proteins (Mehler and Eichele, 1984).

We have three main goals for the molecular dynamics calculations: to determine how well the sigmoidal dielectric function compensates for lack of solvent; to gain insight into the forces that stabilize the helical conformation of this peptide; and to see how well molecular dynamics simulations sample experimentally relevant regions of conformational space. Simulations of this small peptide provide a nice system for testing this dielectric function because we can carry out fairly long simulations, and we have experimental 2-dimensional NMR data at low pH that allow us to check the molecular dynamics simulations semi-quantitatively. The primary reason that our comparison to experiment is only semi-quantitative is that actual distances between the atoms giving rise to the NOESY crosspeaks have not been determined. This is due to the presence of at least two distinct conformational populations for this peptide: the major fraction is helical, the other appears to be extended (Bradley et al., 1990). We expect time-averaged distances less than 3 Å, though, for the atom pairs indicated in Figure 3.1 because the intensities of the crosspeaks are strong. Also, ratios of mainchain crosspeaks and the cou-

pling constants derived from NMR experiments reflect the regions of conformational space in which this peptide resides. We have estimates of helix content from circular dichroism (CD) measurements as a function of pH to compare to simulations with different ionization states of the Glu and His residues. NMR studies are not possible at the higher pH, as the peptide is not soluble enough for the high concentrations required for NMR measurements (Bradley, personal communication).

We performed parallel molecular dynamics simulations of the peptide in Figure 3.1 at low pH (Glu and His residues protonated) starting in an α-helical conformation using the linear dielectric function (ε=R), the sigmoidal distance dependent dielectric function of Hingerty and co-workers, and ε=1 with the peptide immersed in a bath of water molecules. We also did a simulation starting with the peptide in a completely extended conformation at low pH, using the sigmoidal dielectric function. The simulations at low pH were compared to the NMR data collected at pH 2. Simulations starting in the helical conformation corresponding to moderate pH (Glu negatively charged and His protonated) and high pH [Glu negatively charged and His neutral, proton on N_ε which is generally the most highly populated tautomer (Tanokura, 1983)] were also performed using the sigmoidal dielectric function to compare to the pH dependence of helix content measured by CD.

Use of the sigmoidal dielectric function at low pH resulted in distances (averaged over the trajectory for particular interactions) that were consistent with the NMR data. This function surpassed the theoretical ideal of simulating a peptide with solvating water molecules. It had additional advantage of being much less computer-intensive. The interactions that stabilized the helix differed depending on the dielectric model employed. The simulations at low, moderate and high pH using the sigmoidal dielectric function showed comparable helix contents, instead of decreasing with increasing pH as observed experimentally (Bradley et al., 1990).

METHODS

Dielectric Models

Three dielectric models were used for the molecular dynamics simulations: $\varepsilon=R$ (where R is the distance between the charged groups); $\varepsilon=1$ with water molecules solvating the peptide; and a sigmoidal distance dependent dielectric function [equation 1 is essentially the function given by Ramstein and Lavery (1988) with the exception that they used D/2 instead (D-1)/2]. The sigmoidal dielectric function is given below, where R is the interatomic distance, and S and D are constants (Equation 1).

$$\varepsilon(R) = D - [(D-1)/2]\, e^{-RS}\, [(RS)^2 + 2RS + 2] \qquad \text{Equation 1}$$

The slope of the function can be calibrated by the choice of S and the plateau of the curve is determined by the value of D, the bulk solvent dielectric constant. D was set to 78 to represent an aqueous system. We used S=0.3 to give a curve similar to that derived experimentally for the dielectric constant as a function of distance from an ion (Conway et al., 1951) and consistent with experimental results for the dielectric constant of a protein (Rees, 1980; Russell and Fersht, 1987; Russell et al., 1987). We expect the distance dependence of the dielectric function for proteins and peptides to rise less steeply than that determined for an ion based on accessibility of the charges and their effect on the solvent waters; this idea is supported by the experimental data. This function is plotted in Figure 3.2 (with S=0.3), along with the experimental values for a monovalent ion in water (taken from Conway et al., 1951), the average of the experimental results for the dielectric constant in proteins (Rees, 1980), and the linear function with $\varepsilon=R$.

Generation of Structures

The initial model of the peptide in Figure 3.1 was built as an ideal α-helix with extended side chains. The ionization states of the amino acids were chosen to represent the peptide in three pH regions: low pH (e.g. both the Glu and His residues were protonated, corresponding to < pH 4), moderate pH (Glu negatively charged and His protonated, pH 5-7), and high pH (Glu negatively charged and His neutral with the proton on N_ε, corresponding to > pH 8). The peptide had 17 residues plus the two blocking groups

and contained 173 atoms at low pH, 170 atoms at moderate pH, and 169 atoms at high pH. The initial low pH structure was minimized and then used as the starting structure for all other molecular dynamics simulations beginning in the helical conformation. (See Figure 3.3.) A simulation of the peptide starting in a completely extended conformation at low pH (ϕ=180 and ψ=180) was also performed. The protocol described above was used to prepare the extended strand for molecular dynamics.

For the simulation of the peptide in water the minimized helix was solvated by placing the peptide in a box of Monte Carlo equilibrated water molecules (Jorgensen et al., 1983). Any water molecule over 8 Å from any peptide atom was discarded. This procedure resulted in the peptide immersed in a rectangular bath with the following dimensions: 47.30 Å x 28.30 Å x 26.66 Å. 990 water molecules were needed to fill this volume; the system contained a total of 3143 atoms. The entire system was minimized briefly prior to MD.

Computational Details of Molecular Dynamics Simulations

Minimization and molecular dynamics simulations were accomplished using AMBER version 3.0 (Singh et al., 1986). Standard united atom parameters [nonpolar hydrogens are not represented explicitly, the van der Waals radius on the atom to which they are connected is increased accordingly (Weiner et al., 1984)] were used, except for the glutamic acid residues in their protonated state (the standard parameters from Weiner et al., 1984 are for glutamate) and the blocking group at the C-terminus (referred to as residue type NHH). The parameters for these residues were chosen to be as consistent as possible with similar atom types and are given in footnote 1. The united-atom representation is presented here because MD of all-atom models resulted in nonhelical, collapsed structures of higher energy with both of the distance-dependent dielectric models (appendix 2).

The MD simulations were 1200 picoseconds (ps) in duration for the models without explicit solvent molecules and 100 ps for the peptide in water (Figure 3.3). The

PROTOCOL

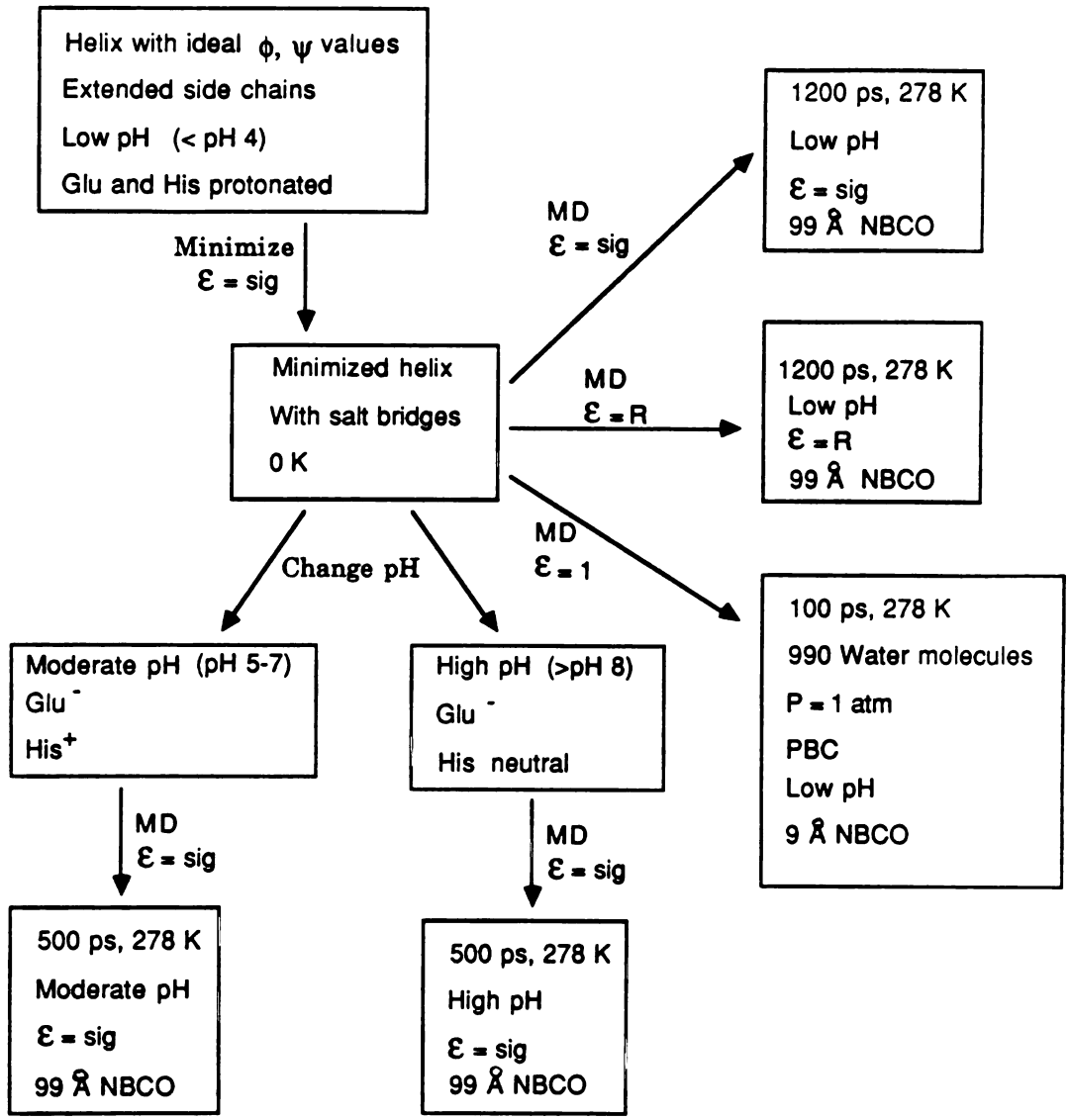


Figure 3.3. Protocol for Generation of Structures and MD Simulations.

simulation using ε=R took approximately 19 hours on a FPS 264 array processor (Floating Point Systems, Portland, OR), with the sigmoidal dielectric function the time increased to about 43 hours on the same machine. The water simulation took approximately 35 hours to complete on a Cray X-MP (equivalent to ~105 hours on the FPS). Therefore, the water simulation took 30 times more computer time than the sigmoidal function, allowing for different simulation times and different machines.

SHAKE (Ryckaert et al., 1977) was used on all bonds so that 2 femtosecond time steps could be used. The temperature was maintained at 278 K by coupling to an external bath (Berendsen et al., 1984). For the water bath simulation the pressure was constant at 1 atmosphere and periodic boundary conditions were employed. Intermediate structures generated during MD were saved every picosecond for analysis. A 99 Å non-bonded cut-off was used for the simulations with the dielectric functions described above, and a 9 Å cut-off was used for the simulation in a bath of water molecules. The cut-off is related to the box size; a 9 Å cut-off was chosen so that the solute molecule would not see its periodic image [e.g. cut-off < 1/2 (shortest box dimension) - 2 Å].

RESULTS

Structural Description of Molecular Dynamics Simulations

End-to-End Distances

The overall dynamics of the peptides, using different dielectric models and in different pH regions, can be visualized most easily by comparing the end-to-end distance of the structures. (Unless the extended strand structure is specifically mentioned, the discussion that follows refers to MD starting in the helical conformation.) Table 3.1 contains the starting end-to-end distance (measured from N of residue 1 to O of residue 17) *in the* minimized helix before MD and how this distance changes during MD. Plots of end-to-end distance as a function of time, for each model, are shown in Figure 3.4. The simulations using the sigmoidal dielectric function at low and high pH and the water model maintained the end-to-end distance of the minimized helix during MD (Table 3.1).

TABLE 3.1
End-to-End Distance (in Å) for Helical Structures using Different Dielectric Models[a]

Dielectric Model[b]	pH[c]	Time of MD (ps)	<d>	$<\Delta d^2>^{1/2}$	$\lvert d-<d>\rvert_{max}$
ε=r	low	100	19.1	3.4	8.3
ε=sig	low	100	26.7	1.6	5.5
ε=1,Water	low	100	25.0	1.4	3.3
ε=r	low	500	18.6	1.8	8.8
ε=sig	low	500	25.8	2.1	9.5
ε=sig	moderate	500	23.6	2.4	12.1
ε=sig	high	500	24.8	1.7	11.9

[a] End-to-end distance is measured from N of residue 1 to O of residue 17. The initial distance for this peptide was 25.2 Å.

[b] ε=sig refers to the sigmoidal distance dependent dielectric function.

[c] Values as defined in text: low pH < pH 4; moderate pH = pH 5-7; high pH > pH 8.

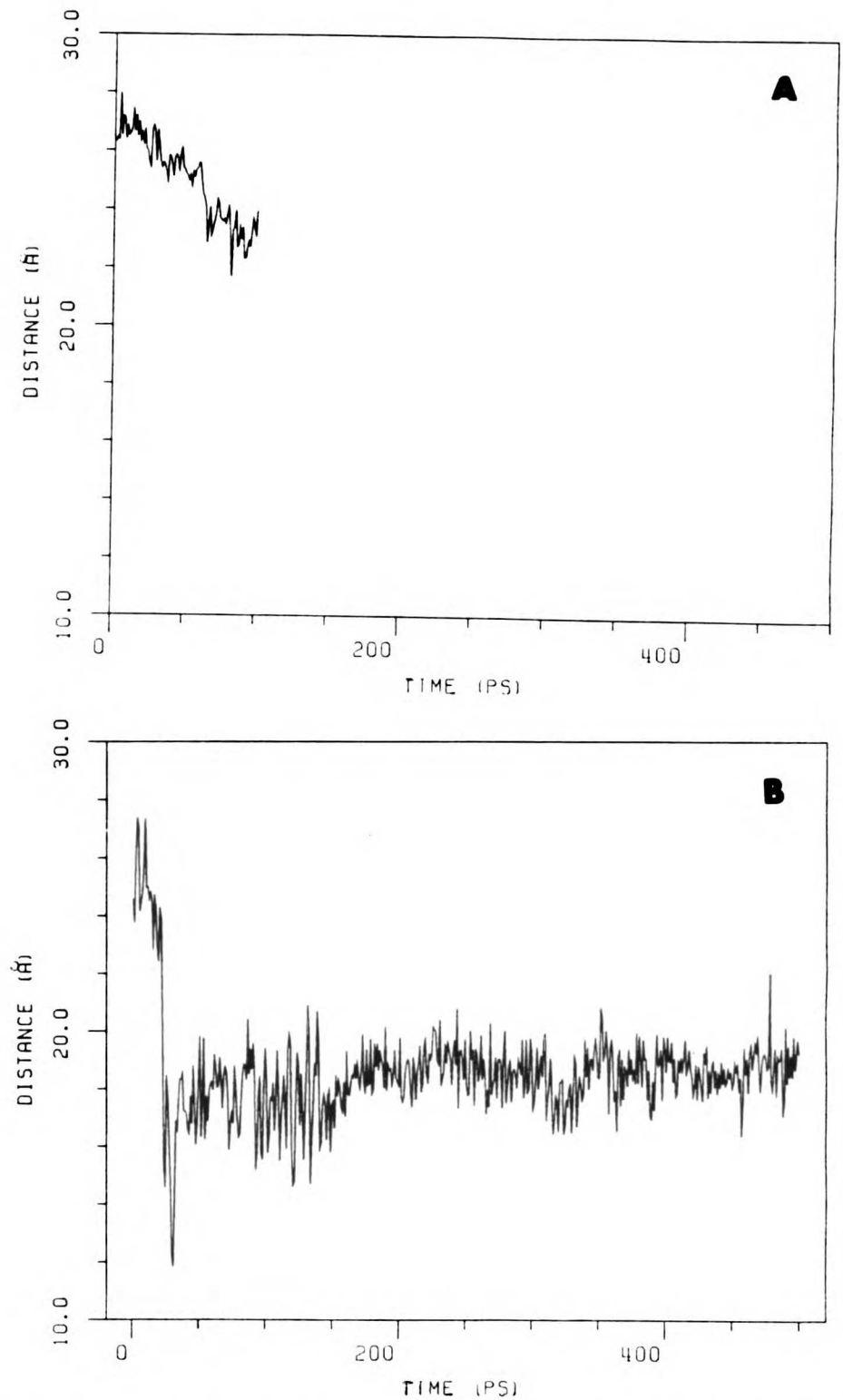


Figure 3.4. End-to-end distance of the helical models as a function of simulation time: A) ε=1 with water molecules at low pH; B) ε=R, low pH; C) ε=sig, low pH; D) ε=sig, moderate pH; and E) ε=sig, high pH.

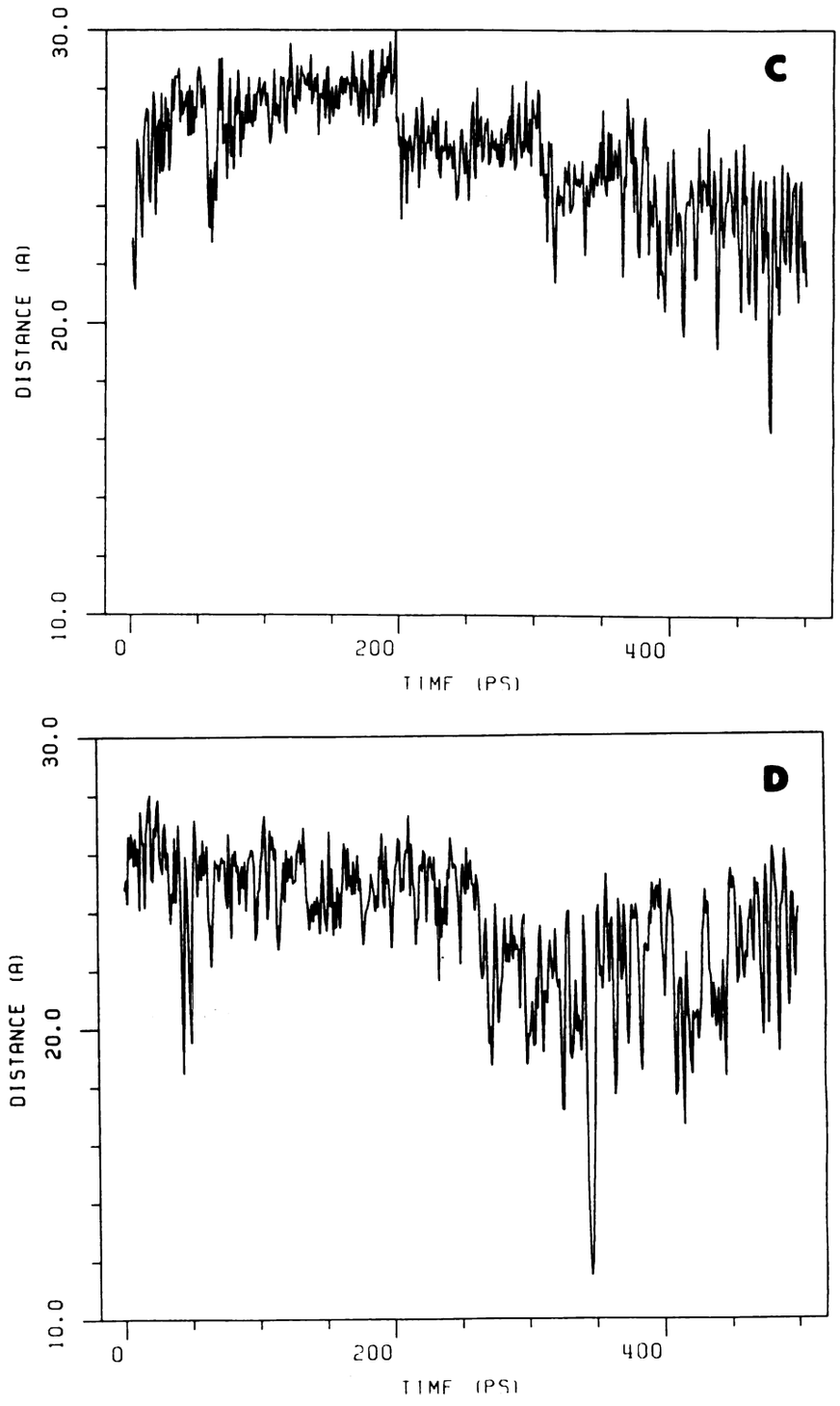


Figure 3.4: continued

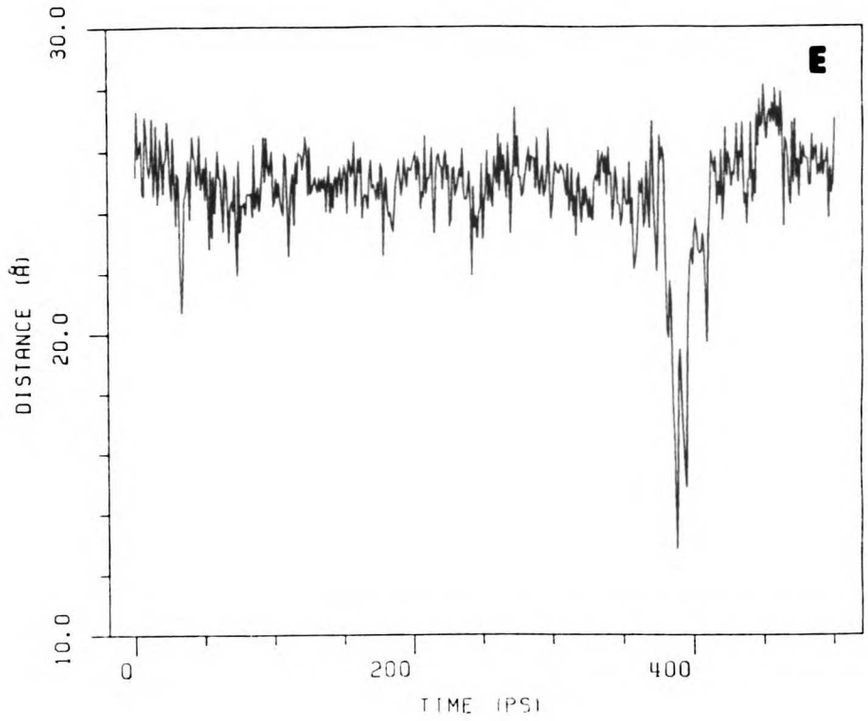


Figure 3.4: continued

The average distances for these simulations were close to the starting distances, although there were fairly large excursions from the mean distance (Figures 3.4C, 3.4D and 3.4E). Although the average end-to-end distances at low pH using the sigmoidal dielectric function and water were similar, the peptide in water sampled less conformational space: both the root mean square length displacement and the maximum deviation from the mean distance were smaller in water than the values using the sigmoidal function (Table 3.1 and Figures 3.4A and 3.4C, comparing the water simulation with the first 100 ps of the simulation with the sigmoidal function). Also, it is clear that the simulation with water has not equilibrated even after 100 ps.

The fluctuations for the dielectric functions shown in Figure 3.4 correspond to real motion and not transient features at the start of the simulation, for they persisted and gave a similar profile for the simulation at low pH with ε=sig when continued to 1.2 nanoseconds (1200 ps). We have seen similar fluctuations in a different small peptide system in which MD was carried out for 4 ns and the fluctuations persisted throughout the entire simulation (Chapter 4). The displacement and deviation values for the linear dielectric function were close to the values using the sigmoidal function but the average distance was much shorter. The peptide became more compact within approximately 40 ps using the linear function and remained compact for the rest of the simulation (Figure 3.4B).

Dihedral Angles

To compare sampling of conformational space with the different dielectric models the angular variance of ϕ and ψ for each residue in the simulations at low pH is shown in Figure 3.5. All of the models exhibited more motion at the ends of the structure than in the center of the helix, which is indicative of fraying. The water model always showed the lowest fluctuations. The linear and sigmoidal functions had similar profiles after 500 ps with the exception that the sigmoidal function showed larger fluctuations near the N-terminus and the linear function exhibited larger fluctuations near the C-terminus (data

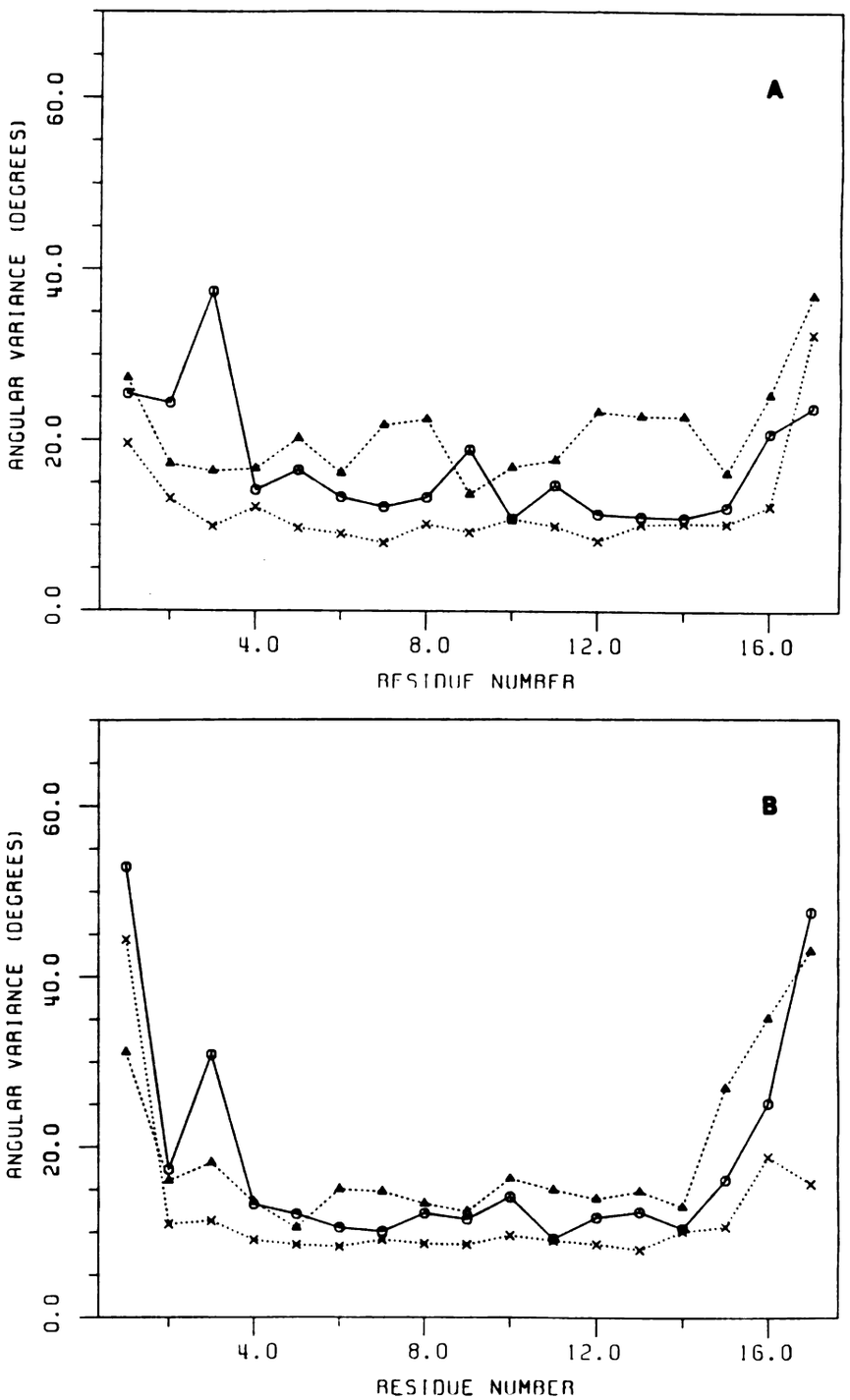


Figure 3.5. Angular variance in ϕ and ψ during 100 ps of MD as a function of residue number for the three dielectric models: explicit water molecules with ε=1 (x); ε=R ($\triangle$); and the sigmoidal distance dependent dielectric function (o): A) ϕ and B) ψ.

not presented). The average angular variance in ϕ and ψ values, respectively, for the different models at low pH during 100 ps of MD were 17 and 19 for the sigmoidal function; 21 and 19 for the linear function; and 12 and 12 for water. Based on the fluctuations in the dihedral angles, both of the simulations using the dielectric functions showed better conformational sampling than the water model.

Positional Fluctuations

The root-mean-square positional fluctuations during molecular dynamics are given as a function of atom number for the three dielectric models at low pH in Figure 3.6. The α-carbons of the structures were matched to correct for rotational motion during MD. For the water simulation, Lys 17 exhibited the greatest movement (Figure 3.6A). For the most part, the profile is uniform; the root-mean-square deviation for the side chain motion is only slightly larger than the values for the mainchain atoms.

The simulation using the sigmoidal dielectric function displayed large movement at the ends of the structure, which again is indicative of fraying (Figure 3.6C). Tyr 12 showed the largest movement during the simulation (approximately 5.5 Å). Interestingly, the peaks of the r.m.s. fluctuations fell into two groups: polar and nonpolar. The polar groups showed approximately 1.5 Å higher average fluctuations than the nonpolar residues, with the exception of Glu 13.

The greatest motion using the linear dielectric function was also exhibited by Tyr 12 (Figure 3.6B). For ε=R, the charged residues showed r.m.s. fluctuations of approximately 3-3.5 Å, with the exception of the neutral, protonated Glu 13. The simulations with ε=R gave a more uniform profile than did the sigmoidal function.

Measurements of Helix Content

Based on the number of intermediate structures generated with MD had ϕ and ψ within $\pm$ 15 degrees of the ideal α-helical angles (-57,-47), we determined the percentage of time the peptide was in the helical region of conformational space for all dielectric models and at different pH values. The overall average percentage of helix (over all

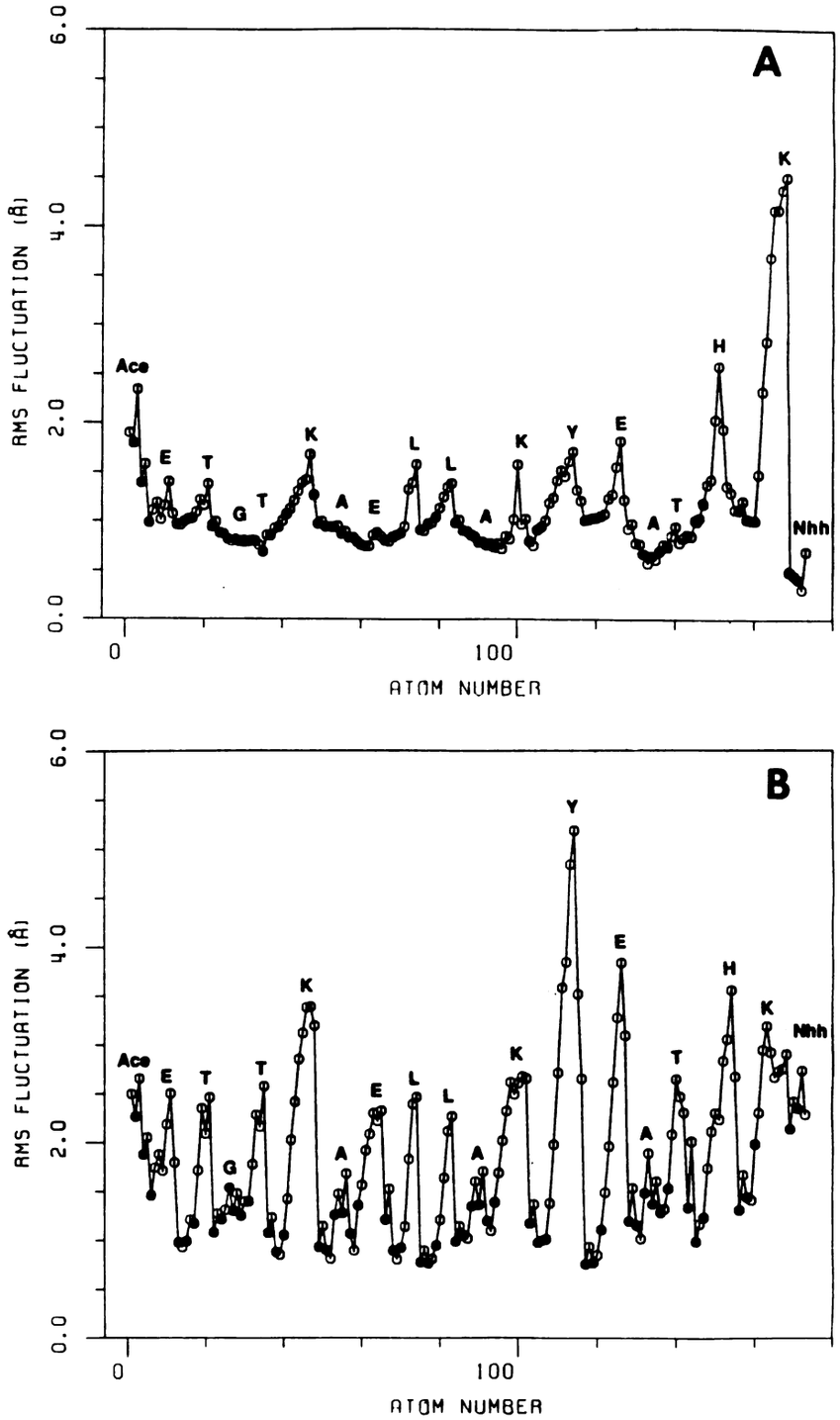


Figure 3.6. Root mean square atomic fluctuations during MD for different dielectric models. Amino acids marked by 1 letter code and blocking groups are labelled Ace and Nhh. Darkened circles represent backbone atoms (C,C_α, N). A) $\varepsilon=1$ with explicit water molecules; B) linear distance dependent dielectric function; and C) sigmoidal distance dependent dielectric function.

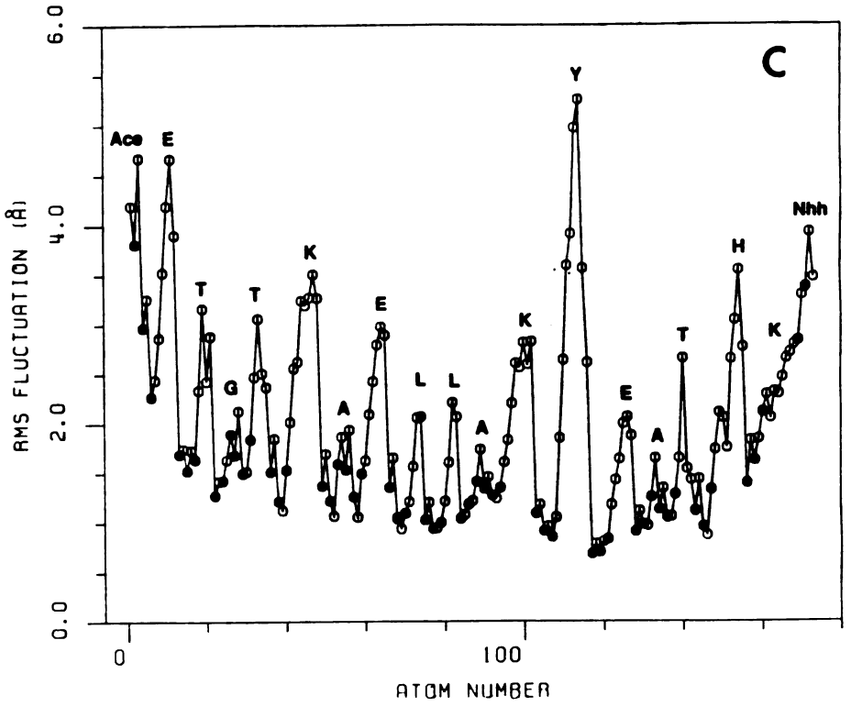


Figure 3.6: continued

residues and 100 ps at low pH and 500 ps at higher pH) for the different models is shown in Figure 3.7. The helix content was the greatest using the water model. Using the sigmoidal dielectric function, there was no significant variation in the helix content with pH.

We then considered another method of estimating helical content based on ϕ and ψ values that takes repeating structure into account. A region was considered helical if both ϕ and ψ were with 20 degrees of their ideal values for at least 3 residues in succession. Various properties determined using this method are given in Table 3.2. The choice of ϕ, ψ values to define the helical region was arbitrary. Therefore, these measurements of helicity are most appropriate for internal comparison of the models rather than direct comparison with experiment.

All of the models started the MD simulations in the helical conformation. But, the fraction of helix as a function of time differed for the various models. The fraction of helix is the number of helical residues as defined above (not necessarily contiguous) divided by the total number of residues, 17. The cut-off to be considered helical was 3 of 17 residues. The fraction of helix using ε=R dropped to very low levels immediately and remained low (Figure 3.8B). The simulation with the sigmoidal function showed high fractions of helix for the first 100 ps of MD but dropped to zero periodically throughout the simualtion, while maintaining an high overall helix content (Figure 3.8C). During this same period of time, the fractional helicity of the peptide in water was high and rarely dropped below 50% (Figure 3.8A).

One can use the definition of helical structure defined above to compute a number of structural properties. The average helix length using the sigmoidal function dropped from 5.3 residues during the first 100 ps to 4.6 over 500 ps, which is approximately the level of the other pH regions (Table 3.2). The simulation at low pH using the sigmoidal function gave the highest average helix length of the dielectric functions. The value using ε=R was 25% lower (3.6 residues, averaged over 500 ps). The average helix length was overwhelmingly the greatest in the water simulation. Using this approach to identify

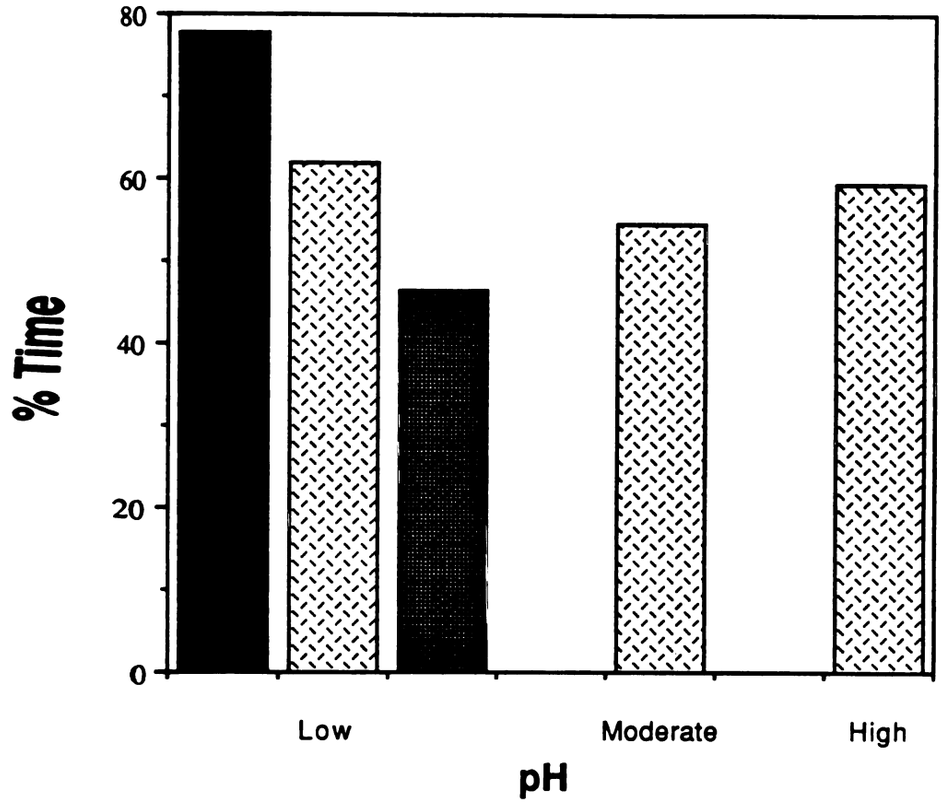


Figure 3.7. Bar graph showing the percentage of time ϕ and ψ were within 15° of **ideal** α-helical angles (-57, -47) for different dielectric models and in different pH **regions**. The low pH averages are over 100 ps and the higher pH averages are **over** 500 ps. The water model is represented by a solid black bar ■, ■ is for **the** linear dielectric function; and ▨ represents the sigmoidal dielectric **function.**

TABLE 3.2
Model Dependent Properties of Peptide Conformations
Starting Molecular Dynamics in the Helical Conformation

Property[a]	Low pH			Moderate pH	High pH
	Water 100 ps	ε=r 100 ps	ε=sig 100 ps	ε=sig 500 ps	ε=sig 500 ps
<# of residues/helix>	8.8	4.6	5.3	4.3	4.3
Maximum # of residues/helix	16	13	15	11	15
<# of helices/peptide>	1.4	0.6	1.5	1.1	1.3
Maximum # of helices/peptide	3	2	3	3	3
% of time peptide has at least 1 helix	100	52	98	80	85
% of time in β region	0.0	2.8	0.4	0.1	0.0
% of time in α-helix region	79.5	39.1	59.1	47.3	53.9
% of time in other regions[b]	20.5	58.2	40.5	52.6	46.1

[a] Conformational distinctions based on ϕ and ψ being within ± 20 degrees of ideal values (α-helix region was $-77 < \phi < -37$, $-67 < \psi < -27$; β region was $-159 < \phi < -99$, $93 < \psi < 155$ to span both antiparallel and parallel β structures) and for the α-helix ϕ and ψ for at least 3 residues in succession had to fulfill this criterion.

[b] Most of these conformations had kinks and turns that disrupted the helical structure.

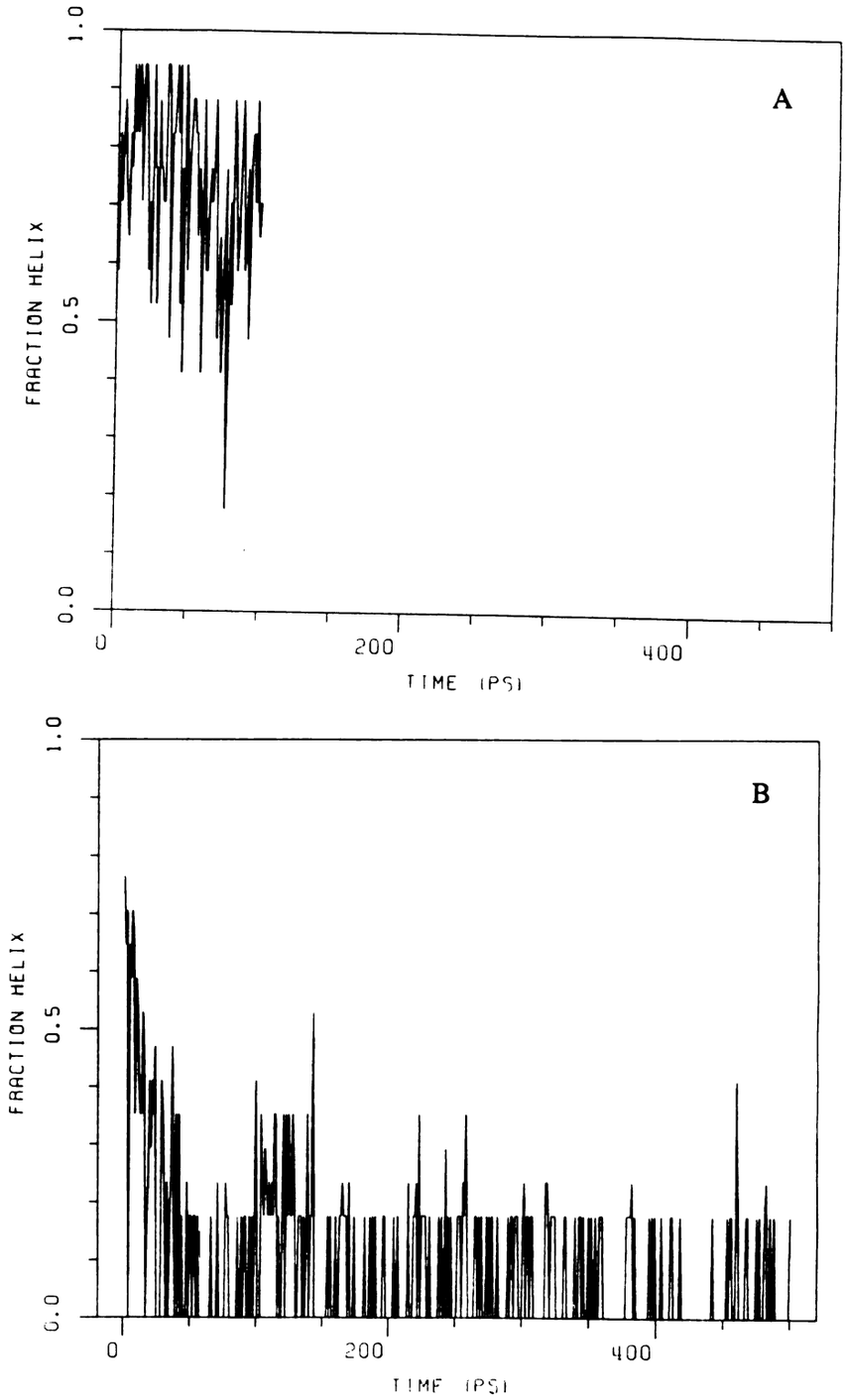


Figure 3.8. The fraction of helix as a function of time during MD for the three dielectric models: A) explicit water molecules with ε=1; B) linear distance dependent dielectric function; and C) sigmoidal distance dependent dielectric function.

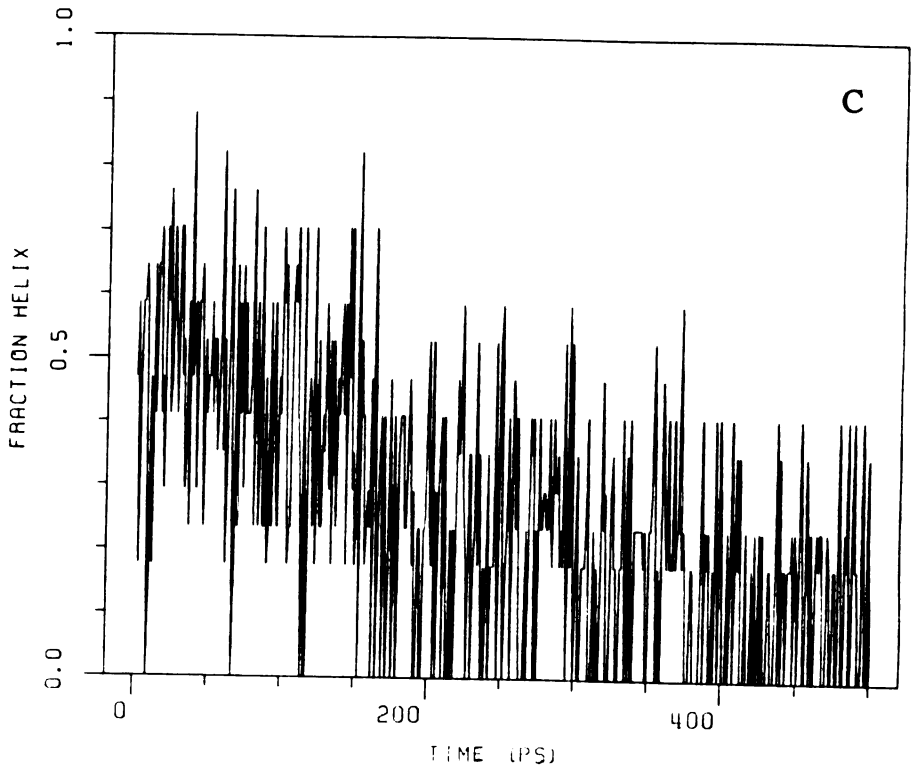


Figure 3.8: continued

helical regions a peptide can have more than one helix. In fact, all of the simulations using the sigmoidal dielectric function and the model with solvating waters had 3 regions with helical ϕ and ψ values per peptide at some point during the simulation. The simulation with the linear dielectric function, on the other hand, had a maximum of two helical regions per peptide. The average number of helices per peptide at low pH was much lower for the linear dielectric (0.6) than the sigmoidal function (1.5) and water model (1.4).

Comparison to 2-D NMR Data

Figures 3.1A and 3.1B show the experimentally observed NOESY connectivities. We monitored distances during our simulations corresponding to these crosspeaks. Because of the r^{-6} distance dependence of the nuclear Overhauser effect (dipole-dipole interaction), we define an effective NOE interatomic distance $r_{eff}(i,j)$ between i and j as

$$r_{eff}(i,j) = (\langle r_{i,j}^{-3} \rangle^2)^{-1/6} = \langle r_{i,j}^{-3} \rangle^{-1/3}$$

where $r_{i,j}^{-3}$ is calculated for each intermediate structure generated during MD and subsequently averaged over the trajectory. The effective distances for particular non-bonded interactions from the simulations corresponding to the experimentally observed NOESY crosspeaks are given in Tables 3.3 and 3.4. Since the calculations used united atom models, it was necessary to correct the distances prior to weighting the distances. One or two C-H bond lengths (1.09 Å) were subtracted from the actual distances, depending on whether the interaction was between two united atoms (e.g. C-H...H-C interactions, 2 bond lengths subtracted) or between an united atom and an amide proton (1 bond length subtracted because amide protons were present explicitly). For many of the observed NOESY crosspeaks we had more than one possible interaction to consider because our atoms are unique while the experimental hydrogens have not always been assigned stereospecifically. For this reason, the shorter distance for each pair is considered in the discussion that follows, although both distances are given in Table 3.4.

TABLE 3.3
Weighted Distances[a] (in Å) from Molecular Dynamics Trajectory Corresponding to Observed NOESY Mainchain Crosspeaks with Different Models[b]

Observed NOE	Helix			Strand
	Water 100 ps	ε=r 100 ps	ε=sig 100 ps	ε=sig 500 ps
$\alpha_4\eta_6$	2.8	3.2	2.9	4.6
$\alpha_5\eta_8$	2.6	3.0	2.3	6.6
$\alpha_5\beta_8$	2.2	2.9	1.9	7.4
$\alpha_6\beta_9$	2.0	3.0	1.9	5.6
$\alpha_8\beta_{11}$	1.9	2.9	2.4	6.6
$\alpha_{10}\eta_{13}$	2.4	3.3	2.6	4.6
$\alpha_{11}\eta_{14}$	2.4	3.5	2.4	6.7
$\alpha_{13}\eta_{16}$	2.5	2.9	2.5	5.5

[a] See text for how distances are calculated.

[b] See Figure 3.1A.

TABLE 3.4
Weighted Distances (in Å) from Molecular Dynamics Trajectory Corresponding to Observed NOESY Side Chain Crosspeaks with Different Models[a]

Observed NOE	Helix			Strand
	ε=1,Water 100 ps	ε=r 100 ps	ε=sig 100 ps	ε=sig 500 ps
$\beta_4\eta_6$	4.1	4.2	4.1	5.3
$\eta_5\gamma_8$	4.5	4.9	4.6	9.5
$\alpha_5\delta_8$	1.9/3.6	2.8/4.6	1.8/3.2	7.1/8.7
$\eta_5\delta_8$	4.3/5.7	4.8/6.1	4.5/5.8	9.4/10.4
$\alpha_5\gamma_8$	2.2	2.9	1.9	7.6
$\delta_9'\beta_{12}'$	3.5/5.4	5.2/5.2	2.9/3.4	8.6/9.1
$\delta_9'\gamma_{13}$	5.1/5.7	2.2/2.5	4.0/4.1	9.8/10.2
$\delta_9\varepsilon_{12}$	4.0/6.2	6.8/7.1	3.3/4.6	8.5/9.4
$\delta_9'\varepsilon_{12}$	1.6/3.9	5.5/6.6	4.2/4.8	9.5/10.0
$\delta_9\delta_{12}$	3.9/6.2	5.8/6.4	2.2/3.8	8.1/9.2
$\delta_9'\delta_{12}$	1.7/4.0	4.9/5.8	3.5/4.0	9.0/9.7
$\gamma_9\gamma_{13}$	5.0	3.0	3.6	9.8
$\delta_{12}\varepsilon_{16}$	3.8/4.9	3.0/3.6	1.7/1.8	8.8/9.1
$\varepsilon_{12}\beta_{16}$	7.4/8.1	3.2/4.6	3.8/4.1	5.0/5.2
$\varepsilon_{12}\varepsilon_{16}$	5.0/5.9	3.1/3.7	1.4/1.4	7.6/7.9

[a] See text for how distances are calculated.

[a] See Figure 3.1B.

Only the first 100 ps of the simulations at low pH (the experiments were performed at pH 2) are considered in comparing to the experimental NMR data because of the short simulation time in water. All of the dielectric models gave short effective NOE distances for the mainchain interactions (Figure 3.1A and Table 3.3). Where the models deviated from one other was in the distances for the side chain interactions (Figure 3.1B and Table 3.4). The simplest way to score the dielectric models, given the uncertainties in the atom assignment and the united-atom approximation, is to assume that any distance less than 5 Å is consistent with the NMR data. Using this assumption, the sigmoidal function gave acceptable distances for all 15 NOEs, the water model and the linear model were satisfactory 11 times. Another way of comparing these data is to use a tighter cut-off of 3 Å (since the observed crosspeaks are strong) and determine the cumulative deviation from this threshold (deviation = Σ r - 3 Å, where the sum is over the 15 NOEs). Again, the sigmoidal dielectric function showed the smallest overall deviation (8.7 Å). The other models were worse: 17.8 Å (water model) and 18.5 Å (linear model). The total deviations remained approximately the same for the dielectric functions when the interactions were considered over 500 ps.

In brief, the simulation with the sigmoidal function showed the best results overall and the agreement was not much different for 100 and 500 ps calculations. In fact, the distances did not change much when we continued the simulation to 1.2 ns (data not presented). Thus, a 100-200 ps simulation with the sigmoidal function was sufficient to model the NMR data. The weighted distances for the extended strand were all too high to explain the observed experimental data.

One-fifth of the side chain crosspeaks are for Tyr-His interactions at the C-terminus of the molecule, representing two of the water model's unsatisfactory distances. Tyr 12 can interact with His 16 in two distinct edge-on modes as well as face-to-face. The simulations with the sigmoidal dielectric function showed relatively close approach of both protons of each of the three geometries, while the other models showed lower distances

for one geometry over another (e.g. the proton on one side of the ring versus the other side, Table 3.4). In the case of the water model, the difference was due mostly to the difference in the Tyr ring movement, as opposed to the His motion (Figure 3.6). The two residues showed van der Waals interactions between the rings but the distances (5-6 Å) and angles (approximately 100 degrees) were not consistent with hydrogen bonding to the aromatic ring.

Ratios of intensities of mainchain NH-C_α crosspeaks have been used to characterize the percentage of α-helix as a function of residue number for this peptide (Bradley et al., 1990). Using this approach the peptide shows a nonuniform distribution of helix (Figure 3.9). The helix is localized between residues 5-15. The percentage of helix (based on ϕ,ψ angles within ± 15° of ideal values) for our best model for reproducing the NOESY crosspeaks, the simulation with the sigmoidal function, is also given in Figure 3.9. It should be pointed out again that this criterion to determine helix content from the MD simulation is arbitrary, therefore absolute numbers are irrelevant and it is the trends that are of interest. Our model of the peptide also showed a nonuniform distribution of helix but was more helical near the N-terminus of the structure and less helical near the C-terminus, compared to the experimental profile. The $^3J_{N\alpha}$ coupling constants (Table 3.5), however, indicate that the peptide is extended at the C-terminus, beginning at residue 15, which is consistent with our model. Figure 3.10A shows a snapshot of the peptide (ε=sig) after 200 ps of MD. The extended portion of the mainchain at the C-terminus can be seen clearly. By unraveling slightly, His 16 can easily interact with mainchain carbonyl groups (discussed further below). This portion of the molecule was not extended during the entire simulation, though (Figure 3.10B).

Since experimental points are missing at the N-terminus of the peptide (Figure 3.9), the structure in this region is not clear. The only coupling constant available between residues 1 and 5 indicates that the structure is not helical. Coupling constants are available at 15° C for residues 1-3 and they are between 5 and 8 (Bradley et al., 1989). The

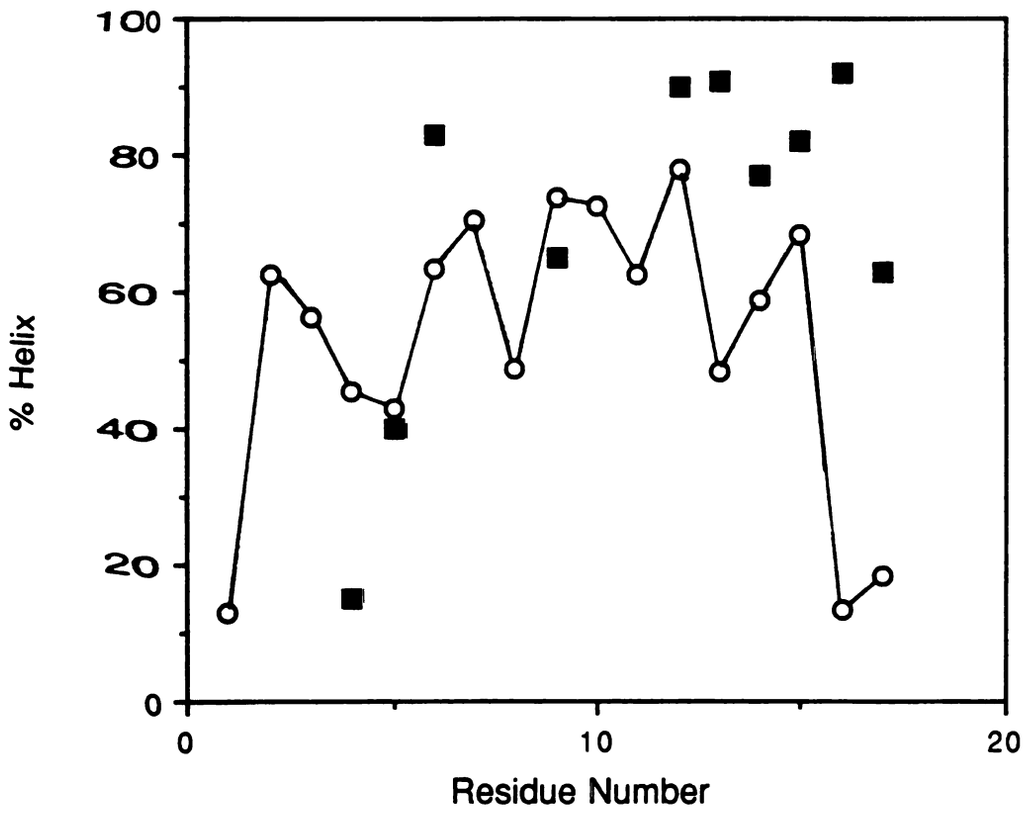


Figure 3.9. Percentage of helix as a function of residue number. Experimental values are from ratios of intensities of mainchain NOESY crosspeaks[2] and marked with solid squares. Values from the MD simulation using the sigmoidal dielectric function (open connected circles) were determined based on the percentage of the 500 ps simulation time the ϕ, ψ angles for a particular residue were within 15° of the ideal values for an α-helix.

value of 5 Hz is at the high end for an α-helix and 8 Hz is consistent with an extended structure, suggesting that most of the molecules are not helical at the N-terminus of the peptide. Our simulation with the sigmoidal dielectric model gave an higher helical content in this region than the experiment but does also sample other more extended conformations (Table 3.5).

The experimental coupling constants and those derived from <ϕ> from our simulations are given in Table 3.5. The coupling constants derived from the simulation with the peptide in water are all uniformly low. The simulation with the linear function, like the sigmoidal function, agrees with the experimental results in showing a nonuniform distribution of helix along the sequence, but the absolute coupling constants for the linear function are uniformly higher than the experimental values. In terms of reproducing the trends observed for the coupling constants along the sequence, the simulation with the sigmoidal function was the most reasonable model.

Helix Stabilizing Interactions

After determining that the structures were helical and consistent, for the most part, with the 2-D NMR data, we examined the interactions stabilizing the helical conformation. We considered salt bridges, side chains interacting with mainchain carbonyl groups or amide hydrogens, and mainchain hydrogen bonds. Each of these is discussed in turn below for the different dielectric models and in different pH ranges.

Salt Bridges

The peptide in Figure 3.1 can, potentially, make four salt bridges or charge-stabilized hydrogen bonds, depending on the pH, as an α-helix. The minimized helix had 3 charge-stabilized hydrogen bonds, the distances for the Glu 7, Lys 11 and Glu 13, Lys 17 interactions were short (1.77 and 1.75 Å, respectively), while the Glu 1, Lys 5 distance was a bit long (3.33 Å). Only one of the helical models showed distances consistent with a Glu 13, His 16 salt bridge---the simulation at moderate pH using the sigmoidal dielectric function. Table 3.6 gives the percentage of time during MD that the

TABLE 3.5
Experimental Coupling Constants ($^3J_{N\alpha}$) and Those Derived from the MD Simulations for the Helical Models at Low pH (in Hz)[a]

Residue Number	Experimental Value	ε=1,Water 100 ps	ε=r 100 ps	ε=sig 100 ps
1		7.6	9.4	9.3
2	8.1	3.2	3.4	5.9
3		3.8	5.5	2.1
4		4.0	4.1	4.0
5	3.7	3.4	6.6	6.2
6	3.7	4.5	5.8	4.3
7		4.0	7.6	3.7
8	5.2	4.0	6.3	4.4
9	4.4	4.2	4.4	4.4
10		4.3	5.5	4.2
11		4.2	7.1	5.2
12	4.4	3.7	7.4	3.9
13	3.7	4.1	6.3	4.6
14	4.4	4.1	5.4	4.2
15	8.1	4.5	6.2	4.2
16	7.4	5.1	9.6	7.9
17	6.6	7.8	9.6	8.2

[a] Experimental values from Bradley et al. (1990). Coupling constants from the simulations were calculated using the Karplus relation (Wuthrich, 1986): $^3J_{N\alpha} = 6.4\cos^2\theta - 1.4\cos\theta + 1.9$ where $\theta = |\phi - 60|$. We used the average ϕ during the MD simulation for each residue. Using this relationship an ideal α-helix (ϕ=-57) gives $^3J_{N\alpha} = 3.9$ and an antiparallel β-strand (ϕ=-140) has $^3J_{N\alpha} = 8.9$.

potential salt bridging atoms were less than 3.5 Å apart.

The simulation with the linear dielectric function showed short average distances for two of the interactions (Glu 7, Lys 11 and Glu 13, Lys 17). All of the distances for the simulation with explicit water molecules were long. The results for the simulation with the sigmoidal dielectric function at low pH depended on the simulation time. The salt bridges remained intact a good deal of time during the first 100 ps of MD but became less prevalent with time. Over the 500 ps simulation, the average distances for the Glu 7, Lys 11 and Glu 13, Lys 17 interactions were fairly short (between 3.5 and 4 Å) and the structures were within the 3.5 Å cut-off roughly 70% of the time (Table 3.6, Figure 3.10A). The root mean square fluctuations in these salt bridge distances were large for the sigmoidal function (1.5-2.0 Å) and much lower for the linear function (0.8- 0.9 Å).

The salt bridges were significantly populated at higher pH (Table 3.6). The distances for the Glu 1...Lys 5 salt bridge were never over 3.5 Å in the simulations at moderate and high pH. In the moderate pH simulation, the positively charged His 16 interacted with both of the oxygens of Glu 13, while at high pH (neutral His) the interaction did not occur. The Glu 7...Lys 11 salt bridge was somewhat less populated at high pH than at moderate pH.

The average length of time that the salt bridges, or charge-stabilized hydrogen bonds, remained intact varied with pH. The average length of time the Glu 7, Lys 11 salt bridge distances were less than 3.5 Å was about 16 psec at low pH (linear and sigmoidal functions). The average length of time increased to 249 ps at moderate pH and 70 ps at high pH. Therefore, the peptide made transient salt bridges at low pH that were not as long-lived as the same interactions at higher pH. Salt bridges (or charge-stabilized side chain-side chain hydrogen bonds) were not the dominant helix stabilizing interactions at low pH and long simulation times; however, salt bridges were very important to maintaining the helical structure for the simulations at higher pH.

TABLE 3.6
Percentage of the Time that Salt Bridges were Present During Molecular Dynamics as a Function of Model and Simulation Time[a]

Residues	Low pH				Moderate pH	High pH
	ε=1,Water	ε=r	ε=sig		ε=sig	ε=sig
	100 ps	100 ps	100 ps	500 ps	500 ps	500 ps
Glu 1 / Lys 5	0	0	68	35	100	100
Glu 7 / Lys 11	1	85	95	72	100	88
Glu 13 / Lys 17	2	83	90	51	99	95
Glu 13 / His 16	0	0	0	0	93	0

[a] This is the percentage of time that distances were less than 3.5 Å. Each Glu, Lys interaction has 3 atom pair possibilities, the shortest distance of any pair was used.

Mainchain-Side Chain Interactions

We then examined other interactions that could stabilize the helix at low pH. Some of these interactions are presented in Table 3.7, which gives the percentage of time particular polar mainchain atoms and side chain atoms were less than 3.5 Å apart.

All of the entries in Table 3.7 represent mainchain-side chain electrostatic interactions. For the most part, these interactions developed with time using the sigmoidal function (compare averages for 100 and 500 ps). Even with longer simulation times in water, this peptide did not participate in mainchain-side chain interactions (data not presented). Most of the side chain-mainchain interactions involved lysine amino groups interacting with mainchain carbonyl oxygens. The carbonyl of Glu 13, Lys 17 side chain interaction (Figure 3.10B), in particular, showed very short average distances in the simulations with the different dielectric functions in all pH regions (2.6 - 3.4 Å) but was not present in the water simulation (Table 3.7). This interaction is a charge-helix macrodipole type of interaction, and it was the only interaction that persisted at higher pH values.

The histidine was also involved in a charge-helix dipole interaction: HN_{δ} of His 16 with the carbonyl oxygen of Tyr 12 (Figure 3.10A). This interaction decreased with increasing pH. The average distance between these atoms using the sigmoidal function at low pH was 2.0 Å. This distance was under the 3.5 Å cut-off 96.0% of the simulation time; this percentage dropped to 29% at moderate pH and 0% at high pH (Table 3.7). This interaction was also strong in water (68% occupancy) and was the only mainchain-side chain interaction to occur to any extent in water.

A charge-helix dipole interaction at the N-terminus of the peptide was observed between the Glu 1 side chain and the mainchain amide proton of Thr 4 at low pH with the linear dielectric function (56.2% of the time) and the sigmoidal function (23.4%, Figure 3.10B). This interaction was less prevalent with the sigmoidal dielectric function because the Glu 1 side chain was also involved in a side chain-side chain hydrogen bond with Thr 4.

TABLE 3.7
Percentage of the Time During Molecular Dynamics that Side Chain-Mainchain Interactions were Present as a Function of Model and Simulation Time[a]

Residues[b]	Atoms[c]	Low pH				Moderate pH	High pH
		ε=1,Water	ε=r	ε=sig		ε=sig	ε=sig
		100 ps	100 ps	100 ps	500 ps	500 ps	500 ps
Ace / Lys 5	O / HN_ε	0	96	0	61	0	0
Glu 1/ Thr 4	O2 / HN	0	56	3	23	0	0
Gly 3/ Glu 7	O / HO	1	5	1	44	0	0
Ala 6/ Lys 11	O / HN_ε	0	94	0	23	0	8
Glu 7/ Lys 11	O / HN_ε	0	11	41	85	0	1
Ala 10 / Glu 13	O / HO	1	0	27	85	0	0
Tyr 12 / His 16	O / HND	68	34	100	96	29	0
Tyr 12 / Lys 17	O / HN_ε	0	0	0	2	0	0
Glu 13 / Lys 17	O / HN_ε	0	68	3	76	95	71
His 16 / Nhh	O / HN	8	71	40	12	4	42

[a] This is the percentage of time that particular, specified distances were less than 3.5 Å.

[b] Ace represents the N-terminal blocking group and Nhh represents the C-terminal blocking group.

[c] If more than one combination is possible (ex. with Lys residues) the interaction with the highest percentage is listed.

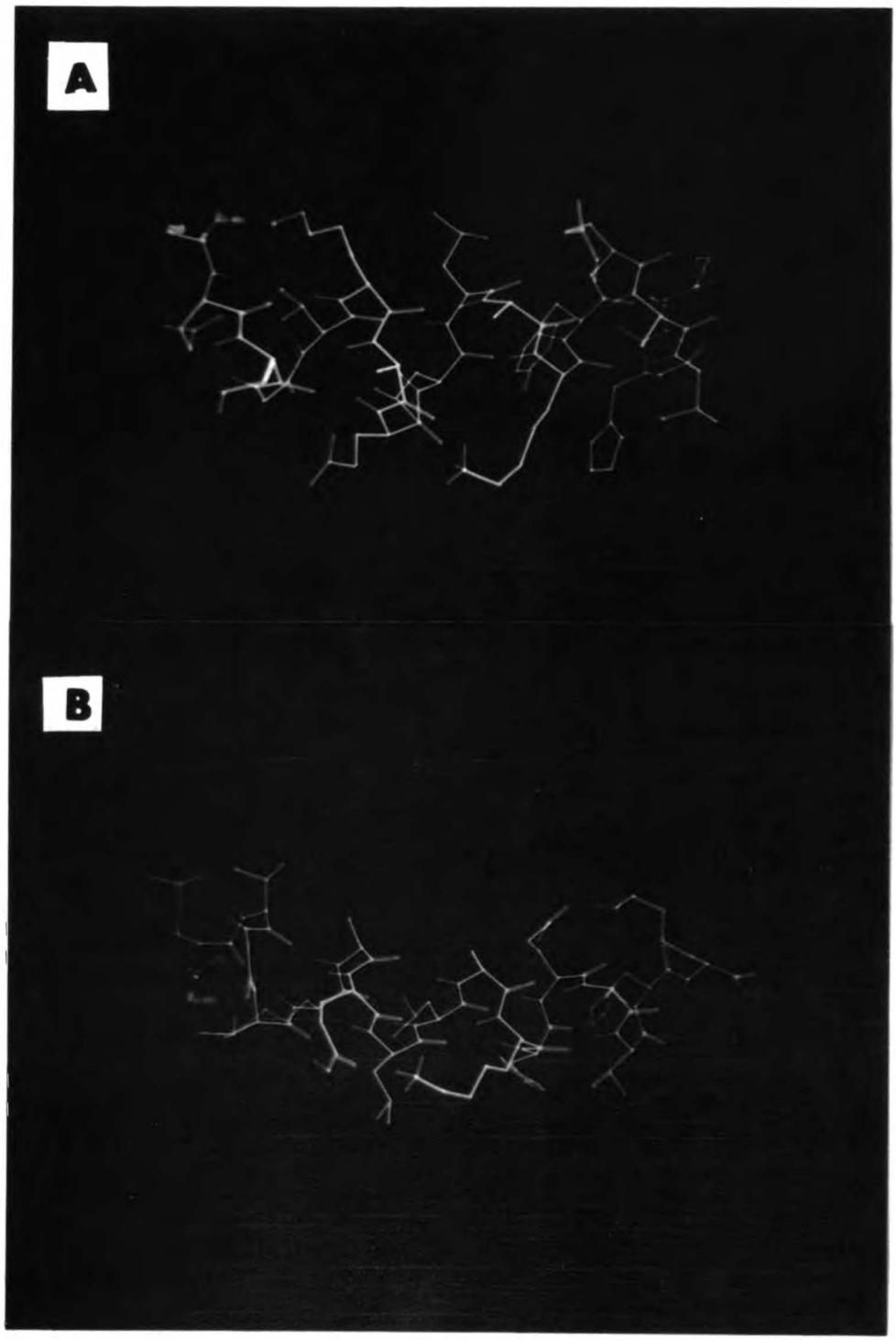


Figure 3.10. Snapshots of the peptide at low pH after MD using the sigmoidal distance dependent dielectric function. Mainchain atoms are in green, hydrogens are shown in red, and oxygens are colored blue. The N-terminus is on the left with the blocking group labelled. A) Structure after 200 ps of MD. B) Structure after 500 ps of MD.

Although the side chain-mainchain interactions given in Table 3.7 were helix stabilizing, they all caused distortions of the helix (Figure 3.10B). Most of the side chain-mainchain interactions observed at low pH were not observed at higher pH. The charged residues were involved in side chain-side chain interactions during the simulation at higher pH instead of with backbone atoms.

Mainchain Hydrogen Bonds

To determine the importance of mainchain hydrogen bonds in stabilizing the helix, we computed the average percentage of time the normal (*i*–>*i*+*4*) hydrogen bonds were intact in each model during MD. If the distance between the appropriate amide hydrogen and carbonyl oxygen was less than 3 Å, the interaction was considered to be an hydrogen bond, disregarding the angle of approach. The percentage of time the hydrogen bonds were intact, averaged over all of the i → i+4 hydrogen bonds (13 in total), for each model is given in Figure 3.11. The mainchain hydrogen bonds were quite important in maintaining the helix in water (intact 98% of the simulation time). Using the sigmoidal dielectric function the hydrogen bonds spent 82% of the time intact during the first 100 ps of the simulation but dropped to the values in the higher pH regions when averaged over 500 ps. The hydrogen bonds were intact approximately 50% of the 500 ps simulation time using the linear dielectric function.

DISCUSSION

We present results of molecular dynamics simulations of a small peptide using 3 different dielectric models (a linear distance dependent dielectric function, a sigmoidal distance dependent dielectric function, and ε=1 with explicit water molecules) at low pH and simulations at moderate and high pH with a single dielectric model (the sigmoidal function). Use of the sigmoidal dielectric function resulted in distances for particular interactions (averaged over the trajectory) that were consistent with the NMR data. In fact, this function gave the best results overall. Compared to simulations in water the sigmoidal function was much less computer-intensive with the additional advantage of

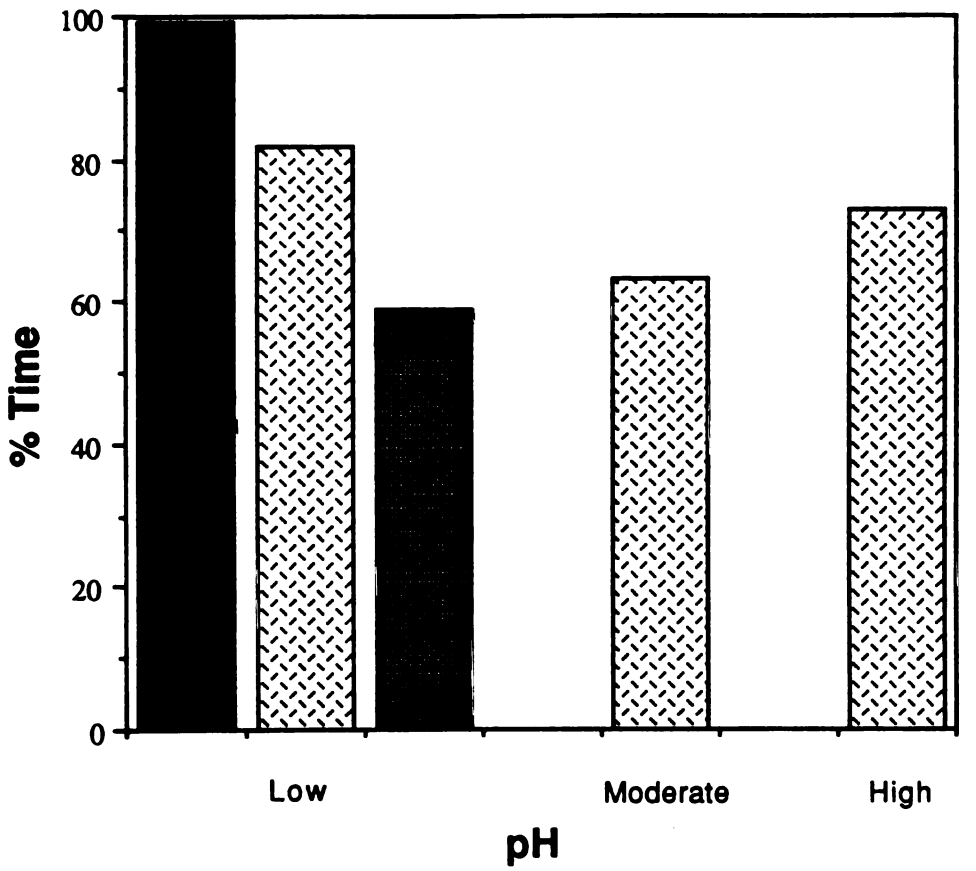


Figure 3.11. Average percentage of time the normal *i–> i+4* mainchain hydrogen bonds were intact during MD in different pH regions and using different dielectric models: ε=1 with water molecules, dark solid bar ■ ; the linear dielectric function, ■ ; and the sigmoidal dielectric function, ▨ . An hydrogen bond was considered intact if the NH...O=C distance was less than 3 Å.

sampling conformational space more effectively. The interactions that stabilized the helix differed depending on the dielectric model employed. Simulations at moderate and high pH using the sigmoidal dielectric function showed comparable helix contents, contrary to the experimental results (Bradley et al., 1990).

Our main aims for the MD simulations of this peptide were to compare different dielectric models and to identify interactions important in maintaining the helical state. This discussion focusses first on a general comparison of the dielectric models, then a comparison of the simulations with experiment, and finally on an overview of the types of interactions observed during the simulations. The comparison of the dielectric models is also implicit in the last two sections.

Comparison of Dielectric Models

Simulations making use of macroscopic dielectric functions have been questioned because these functions are merely *ad hoc* corrections to screen electrostatic interactions. Rigorous, realistic modeling of systems of biological interest should include water to yield a microscopic representation of the dielectric effects. Many of the common water models may not reproduce these effects, though. In any case, simulations of macromolecules with explicit water molecules are not always practical because of computer limitations or constraints imposed by the nature of the problem. Molecular dynamics simulations with water are most appropriate when one wants to sample thermally accessible states near the X-ray structure of a protein. (See, for example, Levitt and Sharon, 1988.) If, instead, one is interested in more large scale motions and sampling of conformational space, macroscopic dielectric approaches may be warranted. Such approximations allow the calculations to proceed much faster than when solvent is included so that slow processes (say the nanosecond time scale) may be visualized in simulations on the picosecond time scale. Brown and Kollman (1987) made use of this in MD simulations of "loop closing" in triose phosphate isomerase.

Many simulations that do not include solvating water molecules employ the linear distance dependent dielectric function $\varepsilon=R$. This function is used because it screens long-range electrostatic interactions to partially compensate for lack of solvent and because simulations using this function can be accomplished in far less time than when solvent molecules are present. It has been argued that the linear dielectric function is reasonable for calculating electrostatic interactions in proteins, where the dielectric constant is generally considered to be between 1 and 5 (Pethig, 1979). Recent experimental work suggests that the dielectric constant in a protein can be as great as 50 at moderate distances (around 10 Å) (Rees, 1980; Russell and Fersht, 1987; Russell et al., 1987), although the actual interactions are probably through solvent. [See Harvey (1989) for further discussion of dielectric effects in proteins.] Use of $\varepsilon=R$ will, therefore, overestimate these interactions in proteins. The problem of overestimating electrostatics is an even greater problem for small peptides that are completely exposed to solvent. Therefore, there is a need for physically reasonable macroscopic dielectric functions for studying peptides and proteins with force field methods. For this reason we implemented the sigmoidal distance dependent dielectric function of Hingerty and co-workers (1985) as modified by Ramstein and Lavery (1988) into our force field (Weiner et al., 1984).

There are two main ways to evaluate the validity of the sigmoidal dielectric function in our force field. The first and most obvious approach is to compare the theoretical results to experiment. The second is an internal check by comparing results using the sigmoidal function to those obtained with more established dielectric models, $\varepsilon=1$ with explicit water molecules and $\varepsilon=R$. We addressed both of these issues. We have both qualitative and semi-quantitative experimental results to compare to our simulations. The second issue regarding internal comparison was addressed by examining conformational sampling in the different dielectric models.

We observed differences in the conformational sampling characteristics of the peptide with different dielectric models. Sampling is especially important for a system like

this, since the peptide has been shown experimentally to sample both helical and nonhelical conformations (Bradley et al., 1989, 1990). To sample experimentally relevant regions of conformational space, our models must also exhibit these characteristics. The sigmoidal function gave distances corresponding to NOESY crosspeaks that were close to the values determined using the water model, but the sigmoidal dielectric function had the advantage of sampling more of conformational space. The simulation with the sigmoidal function, in all pH regions, sampled nonhelical conformations throughout the simulation yet returned repeatedly to the helical state. In contrast, the simulation with the linear dielectric function started in an helical conformation and drifted quickly from that region of conformational space. The peptide in water did not sample much of nonhelical conformational space. The estimated helix contents from populations of ϕ and ψ angles and the coupling constants also indicate that the helix was too stable in water to model the experimental behavior. Due to the motional damping effect of the water, simulations in water aimed at averaging over multiple conformations require large amounts of computer time. It is important to explore this point quantitatively when long water simulations become practicable. Thus, we conclude that the sigmoidal function is a reasonable substitute for explicitly solvating the peptide with water and, in fact, was a better representation for this particular system where sampling is very important.

Comparison with Experiment

The peptide in Figure 3.1 was used for simulation studies at low pH with the various dielectric models because the NMR data provided us with a semi-quantitative check of the MD trajectories. Unfortunately our simulations using the all-atom representation of the peptide were unsuccessful (Appendix 2). The united atom representation resulted in simulations that were more reasonable but present a problem when comparing to proton NMR data. We corrected the non-bonded distances by assuming perfect alignment of protons and subtracting full C-H bond lengths. This correction makes precise comparison to experiment impossible, but given the level of conformational sampling of the peptide,

exact distances cannot be determined experimentally.

The simulation with the sigmoidal dielectric function provided distances that were consistent with the experimentally observed NOESY crosspeaks and overall better represented the experimental data than either the linear dielectric function or the water model. The water model was preferable to the linear dielectric model, though. Interactions between Tyr 12 and His 16 were first identified in a movie of the MD trajectory for the simulation using the sigmoidal dielectric function. Three side chain NOESY crosspeaks were later assigned to interactions between these residues, which provides some support for the reasonability of our models. The Tyr, His interactions appeared to be due to van der Waals interactions and not hydrogen bonding with the aromatic ring, as has been discussed recently by Levitt and Perutz (1988). They point out that it is important to use all-atom models to reproduce these interactions, which may explain why we do not observe hydrogen bonding to the ring. The Tyr, His interactions were strong using the dielectric functions and less prevalent in water. Tyrosine ring motion was damped by the presence of waters. This effect has also been discussed by McCammon and Harvey (1987). Experimentally, the protons on different sides of the Tyr ring could not be distinguished because of overlap of peaks. Peaks may overlap for a variety of reasons. One possible explanation is dynamical averaging on the NMR time scale. The results using the sigmoidal dielectric function are consistent with this idea, as interactions on both sides of the ring gave the same distance whereas the other models did not.

After establishing that the sigmoidal dielectric function gave results consistent with the NMR results and internally consistent with other dielectric models, we looked into the interactions that stabilized the helical conformation. Bradley and co-workers (1990) suggest that His 16 interacts with the helix dipole to stabilize the structure. The peptide shows the highest helix content at low pH as measured by CD (between 2 and 5 pH units, 33%). At low pH and moderate pH, where the His is protonated, the charge-helix dipole interaction should be maximal. Between pH 5 and 7 the helix content drops to 28% and

then drops further to 18% above the pK_a of His.

We also observed a decrease in His 16 interactions with the helix dipole with increasing pH in our simulations. In determining which models exhibited charge-helix dipole interactions, we considered only specific interactions between side chain atoms and mainchain atoms at the ends of the peptide. We did not consider the longer-range interactions between a charge and the field produced by the dipole. This is reasonable given that the dielectric constant is quickly attenuated with distance in water and with the use of the sigmoidal dielectric function. It is not clear, experimentally, whether the helix dipole effect is a true charge-dipole interaction or the result of forming specific hydrogen bonds and electrostatic interactions between a charge and mainchain atoms at the ends of the helix. The only clear experimental example of charge-helix dipole interactions shows specific interactions (Quiocho et al., 1987). Although His 16 interactions with the helix dipole decreased with increasing pH, the overall helix content was comparable in all pH regions, contrary to the CD data. Hence, our results both agree and disagree with the experimental data.

There are a few possible reasons for the discrepancy between our results of helix content in different pH regions and the CD data. We may still be overestimating electrostatic interactions with the sigmoidal dielectric function; at high pH we saw the His 16-helix dipole interaction disappear but the formation of salt bridges resulted in high amounts of helix. This issue could be addressed by adding counterions or performing an high pH simulation in water, where the charged side chains are expected to be more extended based on our simulations at low pH. Another explanation for the discrepancy is that the CD data may not be reliable.

There is a discrepancy between the helix content determined using CD and NMR for this peptide (Bradley et al., 1989). By averaging the ratios of NOESY crosspeaks, determined by Bradley and co-workers, we estimate the helix content at 69.8%. (See Bradley et al., 1989 for the experimental data.) The helix content is approximately 33%

based on CD measurements (Bradley et al, 1990). The helical signal may be masked in the CD spectra by a Tyr signal, and it is not clear whether this effect would be linear with pH. The difference in helix content may be an inherent property of comparing two very different techniques; the nuclear Overhauser effect is a short-range phenomenon while more long-range order is necessary to give rise to an helical CD signal. Woody and co-workers estimate, using theoretical means, that the alignment of 7-11 residues is necessary to produce an α-helix-like spectrum (Manning et al., 1988). They also show that distortions of the helix can lead to diminished signal intensities. In particular, they looked at distortions involving the outward tilting of the mainchain carbonyl groups, 3_{10}-helices are an example of this type of distortion. Transient distortions of this type can certainly occur in a single α-helix in water and, in fact, the peptide appears to form a kink in the center of the molecule (Figure 3.9). We observed similar behavior in our models and such kinks can force carbonyl oxygens to tilt away from the helix. This peptide might also adopt some 3_{10}-helix structure but by the experimental work performed to date this conformation cannot be distinguished from α-helix. From our simulations we would anticipate differences in the helix contents compared with the CD and NMR results, as our average helix lengths were low while the shorter-range interactions were consistent with the NMR data. (The simulation with the sigmoidal dielectric function provides the best example of this.) Unfortunately, we do not have NMR data at higher pH to compare with the CD data and our simulation results.

Types of Interactions

All of the dielectric models showed appreciable amounts of helix during the MD simulations, but they exhibited different helix-stabilizing interactions (Table 3.8). The peptide, using the sigmoidal function at low pH, was stabilized mostly by side chain-mainchain interactions, and to a lesser extent by side chain-side chain interactions and mainchain hydrogen bonds. The linear dielectric function simulation showed similar behavior, although mainchain hydrogen bonds were less important. In water, the salt

TABLE 3.8
Summary of Types of Electrostatic Helix-Stabilizing Interactions Exhibited by Different Dielectric Models and Under Different Conditions[a]

Interactions at Low pH for Different Dielectric Models			
Interaction	ε=1,Water	ε=R	ε=sig
Hydrogen Bonds	+++	+	++
Side Chain-Side Chain		+	+
Mainchain-Side Chain	+++		

Interactions as a Function of pH[b]			
Interaction	Low	Moderate	High
Hydrogen Bonds	++	++	++
Side Chain-Side Chain	+	+++	++
Mainchain-Side Chain	++		

[a] Interactions were considered over 500 ps for the simulations with the dielectric functions. Interactions present less than 40% of the simulation time are left blank, 40-60% is denoted by +, 60-80% is represented by ++, and 80-100% is given +++.

[b] All simulations were performed using the sigmoidal distance dependent function. Different pH regions refer to the ionization states of the Glu and His residues (see text).

bridges that were initially present in the minimized helix were broken almost immediately and replaced by side chain-water interactions.

Mainchain hydrogen bonds were important to maintaining the structure in water and were intact throughout the simulation. One might have expected the intramolecular hydrogen bonds to be less stable in water compared to the simulations with the dielectric functions. Instead we found the intramolecular hydrogen bonds to be as strong as the water-protein hydrogen bonds. This has also been observed by Levitt and Sharon (1988) in comparing simulations of pancreatic trypsin inhibitor (PTI) with water molecules and *in vacuo*. We have observed competition by the water molecules to the mainchain hydrogen bonds in another simulation (data not presented). In that simulation the hydrogen bonds were broken on the average 4-5 times longer than with the dielectric functions because in many cases three hydrogen bonds to solvent had to be broken to reform the mainchain hydrogen-bonding pair. In contrast, Levitt and Sharon (1988) did not see enough motion in PTI during their simulation to enable water molecules to come between hydrogen-bonding groups in the protein. Competition by water does not occur in our system until 150 ps into the simulation (data not presented). All of these simulations show comparable mean hydrogen bond stabilities in water and *in vacuo*, though. Thus, we would expect differences in hydrogen bond stabilities if the simulation were continued because once the water molecules make hydrogen bonds with mainchain groups the intramolecular hydrogen bonds remain broken.

All of our helical models showed stabilization by a combination of interactions. In contrast, we found that mainchain hydrogen bonds were extremely important for stabilization of the helical conformation in a simulation of polyalanine. In that calculation, the hydrogen bonds were intact approximately 91% of the time (over 4 ns) and only remained broken for very short periods of time, on the order of 1 ps (Chapter 4). Not surprisingly, as sequences become more complicated, other interactions, besides the mainchain hydrogen bonds inherent to the structure, become more important in stabiliz-

ing the helix.

Summary

In conclusion, we have shown that simulations with the sigmoidal dielectric function provide reasonable results and can be an alternative to adding explicit water molecules where computer resources are limited. The different dielectric models exhibited differences in importance for some helix stabilizing interactions and in some cases entirely different interactions. For systems like this one where an actual structure cannot be determined for the peptide in solution, because of the presence of multiple distinct conformations, simulations can aid in identifying possible intramolecular interactions, visualizing the movement of the peptide, and interpreting ambiguous experimental results.

Footnote

(1) The nonstandard charges and atom types for glutamic acid and the amide blocking groups are as follows: Glu, OE1 type OH---charge -0.55, HO atom type HO---charge 0.442, OE2 type O2---charge -0.45; NHH, N type N---charge -0.52, H1 and H2 type H---charges 0.26. The following nonstandard angular parameters were added to the parameter list: CT-C-OH, K_θ=75.0 kcal/mole-rad^2, θ_{eq}=117.2°; OH-C-O2, K_θ=70.0 kcal/mole-rad^2, θ_{eq}=125.6°; and X-OH-C-O2, $V_n/2$=10.5 kcal/mole, γ=180°, n=2.

CHAPTER 4: A Molecular Dynamics Simulation of Polyalanine: An Analysis of Equilibrium Motions and Helix-Coil Transitions

The α-helix is an important structural element of proteins. An understanding of helix dynamics and transitions can aid in interpreting the motions of proteins. The helix-coil transition is also of interest because of its probable role in protein folding. The helix-coil transition has been the focus of many experimental and theoretical studies. However, detailed analysis of the helix-coil transition from a microscopic simulation of a peptide of approximately the same length as found in proteins is not available. One of the main reasons for this is that helix-coil transitions occur on the 10^{-6} - 10^{-9} sec time scale, whereas simulations are generally limited to the picosecond (10^{-12} sec) time scale.

A number of approaches have been taken to circumvent the problem of the difference in time scales. McCammon and co-workers used a simplified model for the polypeptide chain with only a single sphere representing each amino acid (1980). Furthermore, they only allowed the last 5 residues of a 15 residue peptide to move, but the simulation was carried out for a long period of time, 12 nanoseconds (ns). Czerminski and Elber (1989) have carefully evaluated all probable structural transitions for a tetrapeptide, including detailed analysis of the transition of the helix to extended structures. In a preliminary account, Brooks (1989) reports the use of molecular dynamics at 600 K to explore whether the α-helix is the global energy minimum for a 13-residue peptide of polyalanine. He generated 10 random structures and simulated their dynamic behavior for 2.5 ns. He found that the peptide exhibited complete unfolding and refolding. Detailed analysis of this behavior is not yet available. A number of statistical mechanical models have been described to treat various aspects of the problem (outlined in Cantor and Schimmel, 1980). Using these models, the tendency to form helices is determined by the length of the polypeptide chain, the equilibrium constant for propagation of an helix (s), and a nucleation parameter (σ) that reflects the difficulty of nucleating helices within nonhelical segments of the chain. These parameters can be determined experimen-

tally using the host-guest method in which the effect of introducing an amino acid (guest) into a long, water soluble polymer (host) with a well-characterized helix-coil transition is measured (Sueki et al., 1984). The problem with these theories is that not all of the assumptions are valid for short peptides, although they are in fairly good agreement with experiment when long polymers are considered.

These studies, with the exception of the work by Brooks, are all useful in furthering our understanding of this important process but limited in some way for describing the behavior of small helical peptides. For that reason we decided to perform molecular dynamics of a 20 residue peptide of alanine to further explore helix dynamics and the helix-coil transition. Alanine was chosen because it is the simplest polypeptide that is able to adopt the α-helical conformation. We allowed the entire molecule to move freely. We had to make some approximations, though, because of the time limitations addressed above. To increase the likelihood of transitions, we ran the simulation at an high temperature for a long period of time (4 nanoseconds), and did not include explicit solvent molecules. Using this approach we obtain a description of the overall structure and inherent flexibility of the chain as well as a structural picture of conformational changes that occur. In this way, we can address both equilibrium properties of the peptide and the dynamics of the structural transtions.

Our results correlate fairly well with the available experimental data and previous simulations aimed at addressing dynamics of an α-helix. Our results deviate from the statistical mechanical theories, as might be expected for a small peptide. The relevance of our results to protein folding is also discussed.

METHODS

Molecular dynamics simulations were performed of a 20 residue peptide of alanine starting in the α-helical conformation. The starting structure was generated using a routine from ECEPP (Momany et al., 1975) with repeating $\phi = -57°$ and $\psi = -47°$. The peptide was terminated at the peptide bond without charged or capped ends. This structure

was minimized briefly and then equilibrated at 300 K for 25 picoseconds (ps) using molecular dynamics (MD). The resulting structure was the starting point for the MD simulation and considered the time = 0 ps structure. All calculations were performed using AMBER version 3.0 (Singh et al., 1986) with standard united-atom parameters (Weiner et al., 1984).

The molecular dynamics simulation began with the time = 0 ps structure described above. The temperature was increased to 400 K and maintained by coupling to an external bath using the method of Berendsen et al. (1984). One femtosecond time steps were employed and the simulation was carried out for 4 nanoseconds (ns, 4000 ps), resulting in 4 x 10^6 total steps of MD. Structures were saved every 0.5 ps during the simulation for analysis, resulting in 8000 structures. A 10 Å non-bonded cut-off was used and the pair-list was updated every 50 steps. A linear distance dependent dielectric function was employed ($\varepsilon=r$).

RESULTS

We performed a molecular dynamics simulation of polyalanine to explore the types of equilibrium motion and structural transitions the peptide undergoes. To increase the likelihood of observing helix-coil transitions, we carried out a long trajectory (4 ns), did not use explicit solvent. and maintained an high temperature (400 K). First, we present the results pertaining to the structural and equilibrium properties of the simulation, and then we deal with the actual transitions between different structures.

Overall Dynamical Properties of the Molecular Dynamics Simulation

Figure 4.1A illustrates the amount of motion that occurred during the simulation by a plot of the end-to-end distance (between the amide hydrogen of residue 1 and the carbonyl oxygen of residue 20) as a function of time. The average distance was 28.3 Å, with root-mean-square fluctuations of 2.4 Å. An ideal α-helix of 20 residues is approximately 30 Å long. This polyalanine peptide maintained the ideal length fairly well throughout the simulation but there were some large excursions to more compact structures. Many

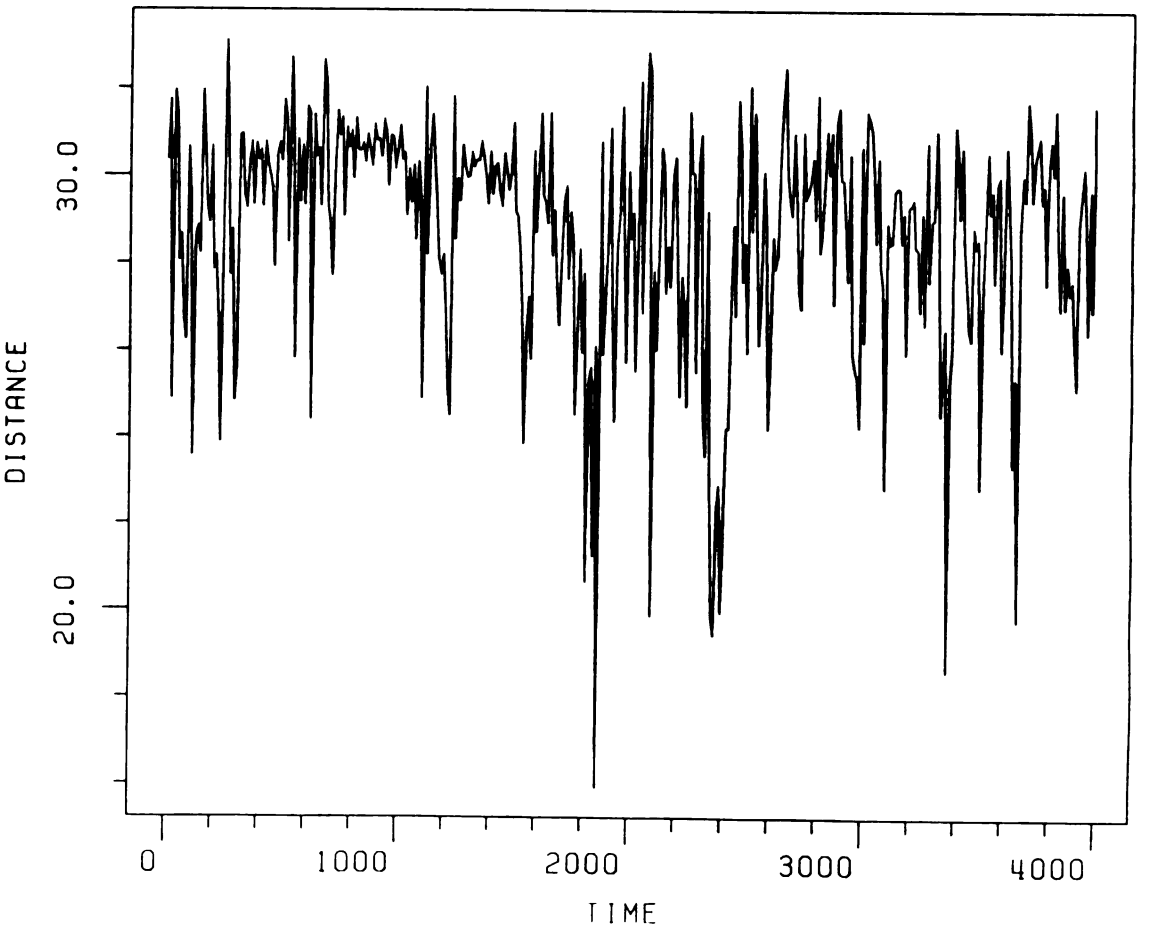


Figure 4.1A: End-to-end distance as a function time (plotted every 5 ps).

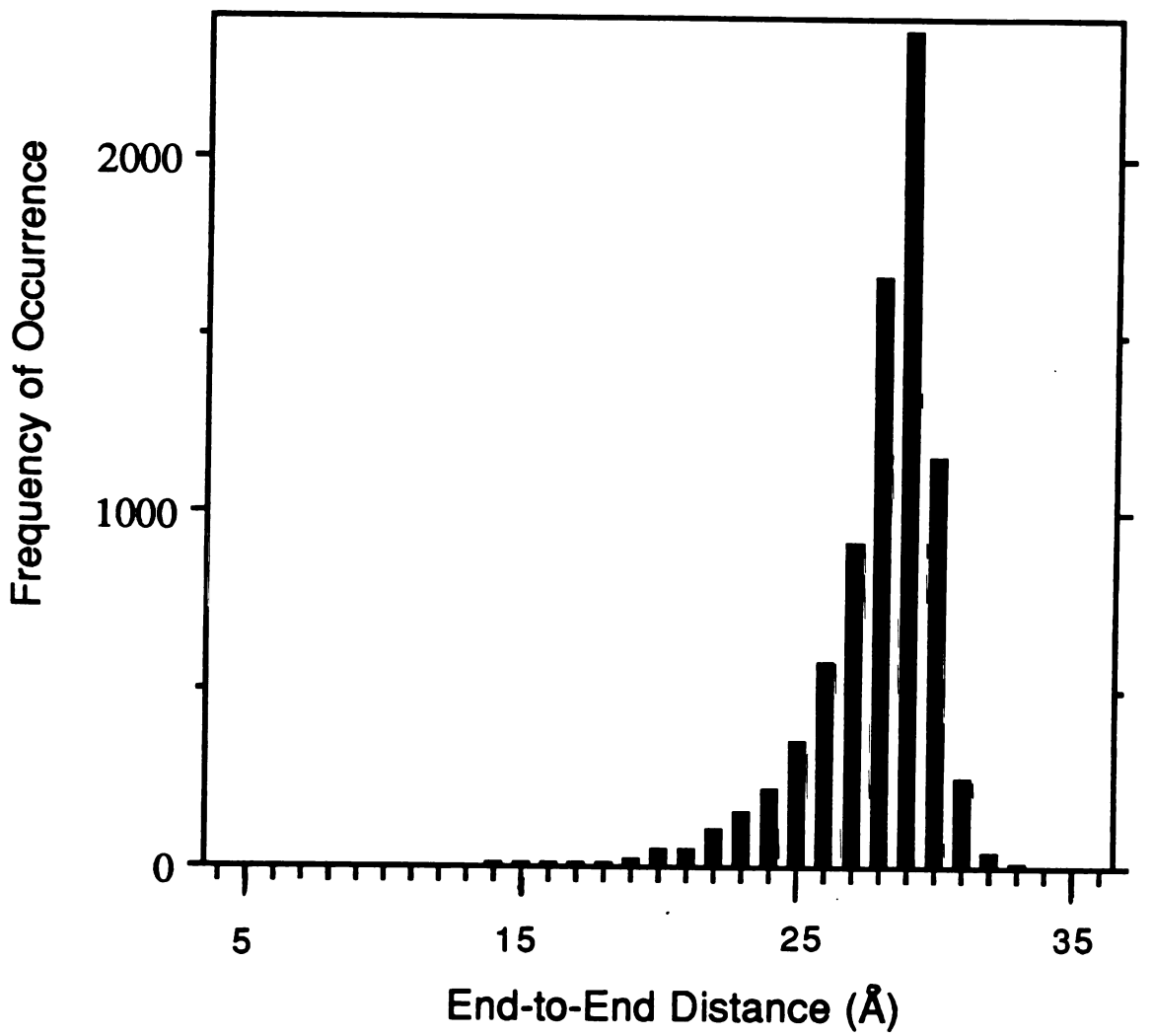


Figure 4.1B: Distribution of end-to-end distances (all 8000 structures considered).

of these displacements represent accordian-like motions and bending of the peptide. The end-to-end distance dropped below 10 Å and remained there for 3 ps with the minimum occurring at 3360 ps and a length of 5.26 Å (not shown because of the plotting interval). Also, as seen in Figure 4.1A, the system appeared to be well-equilibrated; a drift was not observed over time and the fluctuations persisted throughout the entire simulation.

All of the large deviations from the ideal length favored compaction of the structure (Figures 4.1A and 4.1B). The distribution function for the end-to-end distance is shown in Figure 4.1B. The function shows departures from Gaussian behavior; it is sharply peaked and drops off quickly with distances greater than the maximum. This distribution gives a characteristic ratio of 3 (characteristic ratio = $\langle r^2 \rangle / nl^2$, where n=number of residues -1, r is the end-to-end distance and l represents the spacing between C_α atoms, 3.8 Å). The maximum instantaneous characteristic ratio observed during the simulation was 5.

Another way to view the amount of motion occurring during the simulation is by the variance in the ϕ and ψ angle, defining the structure of the peptide. The peptide was predominantly helical during the simulation with an average ϕ, over all time and all angles, of -65.7° and $\langle \psi \rangle$ = -40.4°. Thus, ϕ increased from the starting values (~-57°) while ψ decreased (starting value ~-47°), but they were still near the helical values. Figure 4.2 shows the average angular variance of the ϕ and ψ angles as a function of position over the entire 4 ns. The values were higher at the ends than in the center of the structure for both ϕ and ψ, which is indicative of fraying of the helix. The fluctuations of ϕ were on average slightly higher than ψ ($\langle \delta\phi \rangle$ = 16°, $\langle \delta\psi \rangle$ = 15°).

Figure 4.3 shows the correlations in the fluctuations of ϕ of residue 10 with the fluctuations of ψ of all other residues. The strongest correlation was between ϕ_{10} and ψ_9. The motions were damped quickly moving away from the center of the peptide. The strong anticorrelation between ϕ_{10} and ψ_9 caused the peptide bond to move in a crankshaft-like motion that localizes any distortions by keeping the rest of the helix more

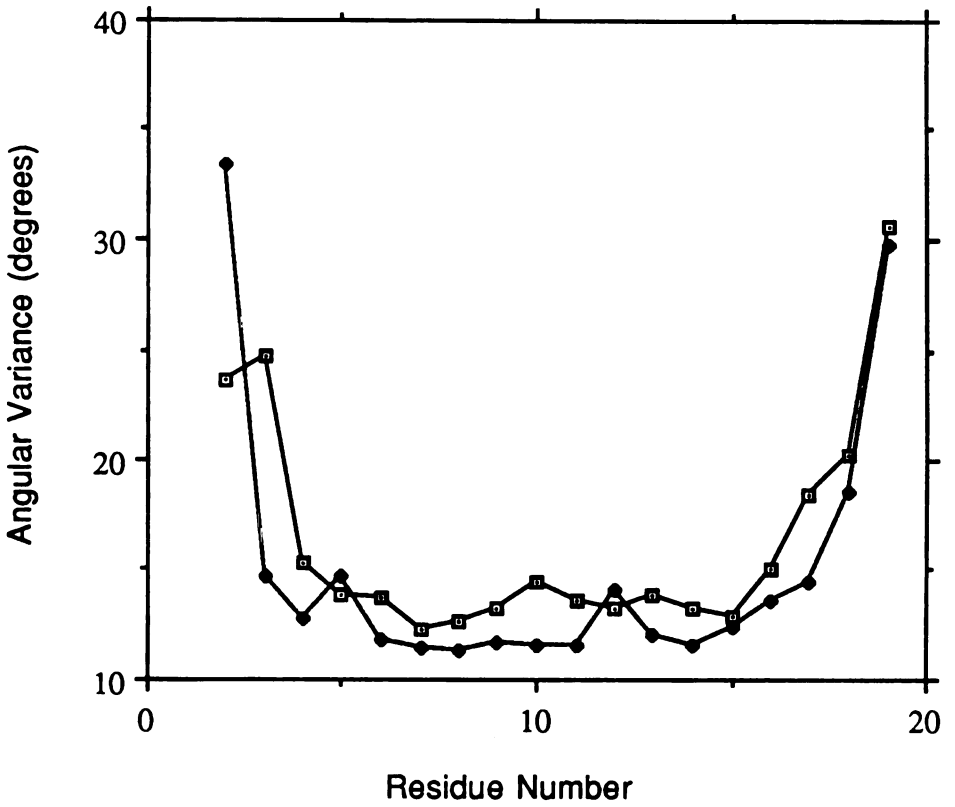


Figure 4.2: Angular variance of the dihedral angles ϕ and ψ as a function of residue number. ϕ is shown as open squares and ψ is represented by solid points.

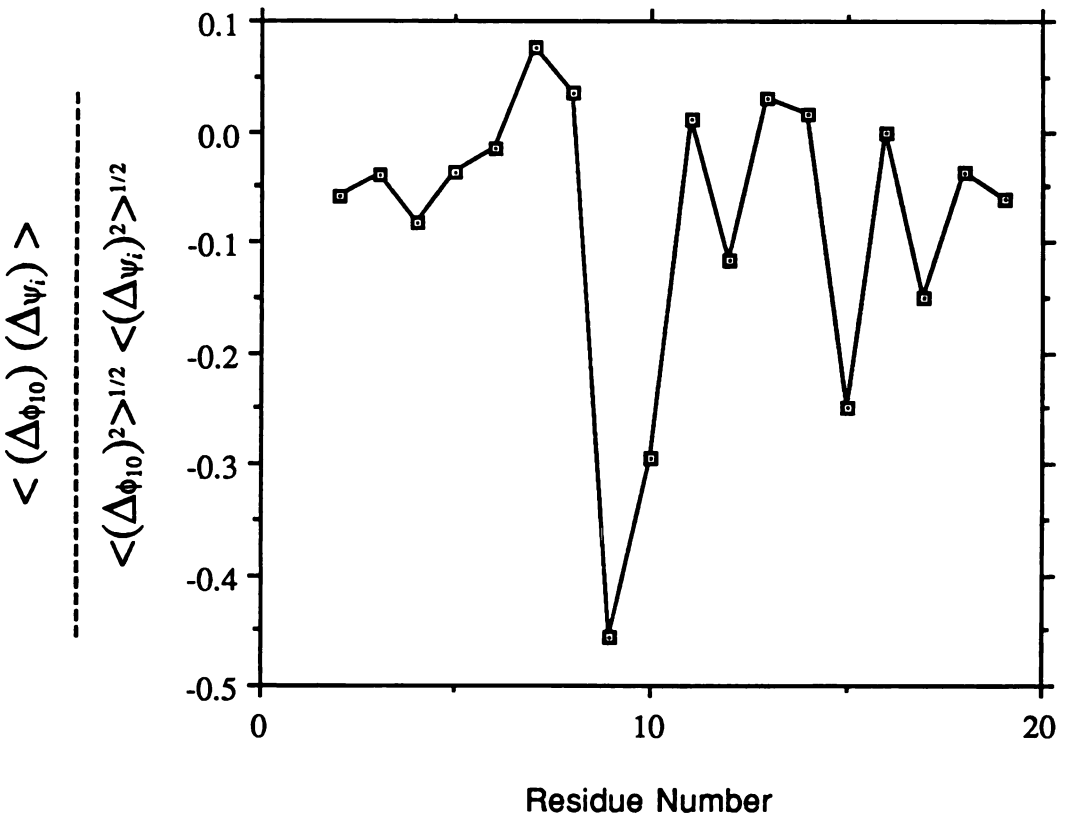


Figure 4.3: Cross correlation coefficients for fluctuations in residue 10 with fluctuations of all other residues.

or less fixed. This strong anticorrelation between ϕ_i and ψ_{i-1} was prevalent throughout the structure (data not presented). This anticorrelation was most striking for ϕ_3 and ψ_2 with a correlation coefficient of -0.72 compared to an average of -0.45 in the center of the structure. This value is predominantly due to a long-lived structure (approximately 400 ps) in which ϕ_3 decreased by roughly 70° and ψ_2 increased by 180°.

Structural Properties of the Peptide

Table 4.1 lists some of the gross structural features of the peptide at different times during the simulation. All of the properties in Table 4.1 are based on finding helical portions of secondary structure within the peptide. We defined helical regions as having at least three residues with ϕ and ψ angles within 20 degrees of ideal values (ϕ = -57°, ψ = -47°). This particular cut-off to define helical regions is arbitrary, but we wanted to use a value greater than the mean dihedral fluctuations. With this definition a particular structure can have multiple helices. This approach for determining helix content neglects the end residues, as both ϕ and ψ are necessary to assign helical regions. Therefore, 100% helix refers to a structure containing 18 residues with helical ϕ and ψ values. To determine the overall percentage of helix, all helical residues meeting the criteria above were considered for a particular structure and then averaged over all structures.

The peptide remained very helical for the first nanosecond of the simulation (81%) but dropped slightly over the next nanosecond (74%, Table 4.1). The overall helical content reached a plateau at 68% during the last 2 ns of the simulation. During the course of this simulation the peptide was completely helical 534 times, or 7% of the simulation, thereby showing that fragments of helix contributed to the overall high helix content. The peptide was completely nonhelical 106 times, or 1% of the simulation time. Therefore, a value of 50% helix does not mean that the peptide was completely helical and completely nonhelical 50 % of the time. As can be seen in Table 4.1, the completely nonhelical conformations became more prevalent with time. The average time between nonhelical structures was high during the first nanosecond but decreased with time, tapering to 24 ps

TABLE 4.1
Gross Structural Features of the Peptide
During the Molecular Dynamics Simulation

Simulation Time (ns)	Percentage Helix	<Time> (ps)[a] 0 % → 0 %	<Time> (ps)[a] 100 % → 100 %
0 - 1	81	333	2
1 - 2	74	53	62
2 - 3	68	24	125
3 - 4	68	24	500

[a]The average time between either completely helical (100%) structures or between completely nonhelical (0%) structures.

after 2 ns. Conversely, the peptide was completely helical every 2 ps on the average during the first nanosecond but increased to 0.5 ns during the last nanosecond of the simulation. Although, the average time between completely helical structures increased during the second half of the simulation, the percentage of helix remained constant, indicating the increase of multiple helical fragments within a single structure.

Using the definition of helix content described above, the fraction of helix as a function of time was determined (Figure 4.4A). As can be seen, the fraction of helix remained high throughout the simulation but there were large excursions from the mean. Given this plot, the helix was clearly stable and highly populated, such that excursions to less helical structures were quickly converted back to helical conformations. Figure 4.4B shows a blow-up of a region of Figure 4.4A to give a finer scale and closer view of these deviations. Although these transitions were relatively rare they occurred very rapidly. In fact, we are probably observing the upper time limit for this process because we only saved structures every 0.5 ps for analysis. During the simulation the peptide never passed from a completely nonhelical to a completely helical structure. Instead, structures with intermediate amounts of helix were observed.

Table 4.2 shows the average overall time and number of occurrences of particular transitions. The transitions shown are for helix contents greater than or equal to the upper limit and less than the lower limit, but do not necessarily represent transitions going directly from one state to another without intermediate states. The number of times the transitions occurred increased as the criteria for defining conformational regions became less restrictive, while the average overall time for the transitions decreased. The transition times were highly variable; all of the transitions had mean fluctuations of at least 60% of the mean value. The range of rates for the various transitions was determined by using the highest and lowest observed transition times. The range in transition time was from 10^{-11} - 10^{-8} sec. Snapshots depicting different amounts of helical content, from the trajectory, are shown in Figure 4.5: the peptide with 100% helix content, 50% helix con-

TABLE 4.2
Transitions Between Conformational States with Different Helical Contents

Transition	# of Times Transition Occurred	<Overall Time> (ps)	Time Scale of Transition (sec)[a]
100 % → 0 % → 100 %	11	668	10^{-8} - 10^{-9}
90 % → 10 % → 90 %	29	271	10^{-9} - 10^{-10}
80 % → 20 % → 80 %	108	74	10^{-9} - 10^{-11}
70 % → 30 % → 70 %	300	27	10^{-9} - 10^{-11}
60 % → 40 % → 60 %	734	11	10^{-9} - 10^{-11}
50 % → 50 % → 50 %	1140	7	10^{-9} - 10^{-11}

[a]The transitions were considered for all helix contents above teh first cut-off given and below the second. The range of times for the transitions was determined by using the high and low values of the time for each particular transition.

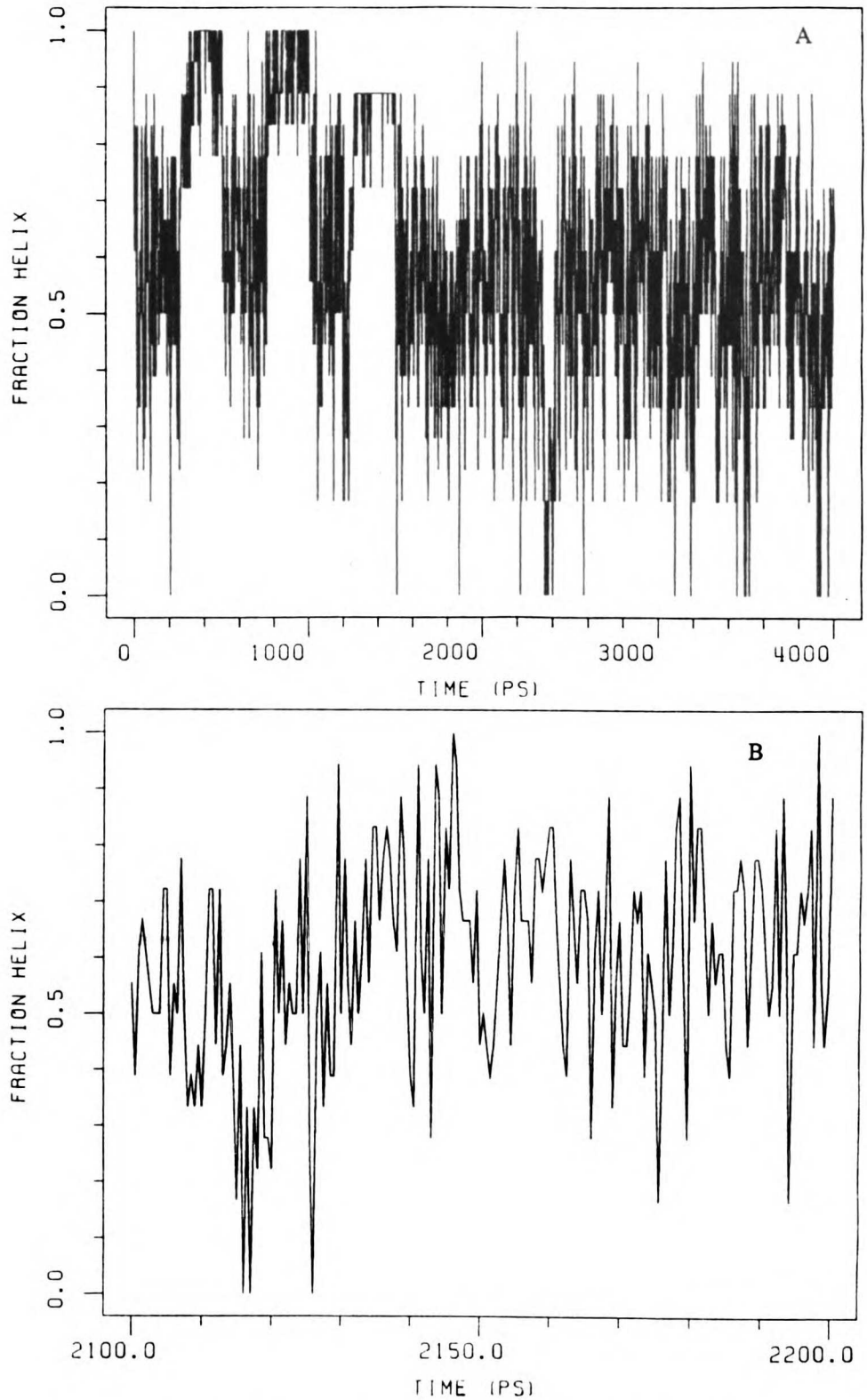


Figure 4.4: (A) Fraction helix as a function of time (plotted every 5 ps). Helix content based on dihedral angles as explained in the text. (B) Blow-up of fraction helix for an one hundred ps time period (points plotted every 0.5 ps).

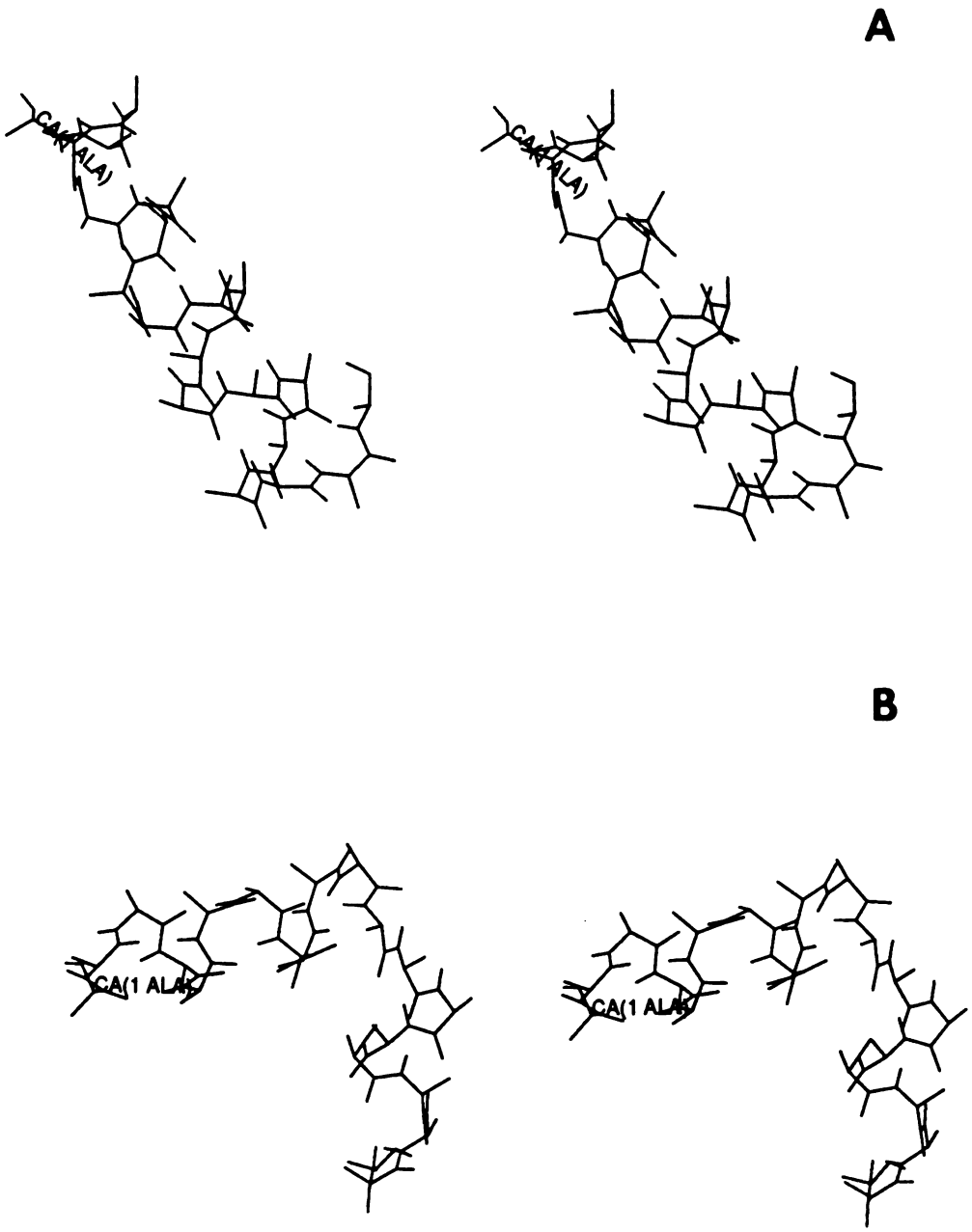
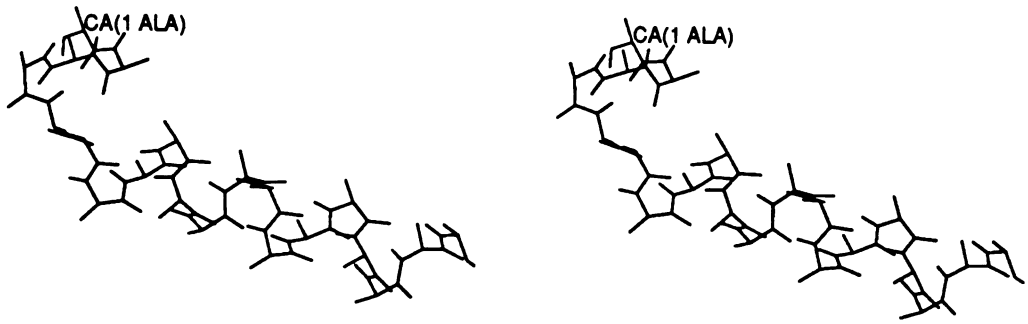


Figure 4.5: Stereoview of snapshots of polyalanine during the molecular dynamics trajectory, representing different helical contents and simulation times: (A) 2355.5 ps, 0%; (B) 3366.0 ps, 0%; (C) 1850.0 ps, 28%; (D) 2349.0 ps, 50%; (E) 3360.0 ps, 50%; and (F) 3716.0 ps, 100% helix. Each peptide is labelled at the amino-terminus.

C



D

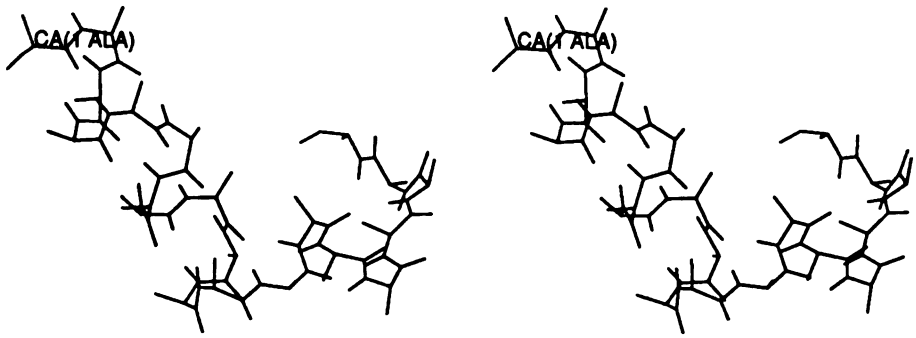
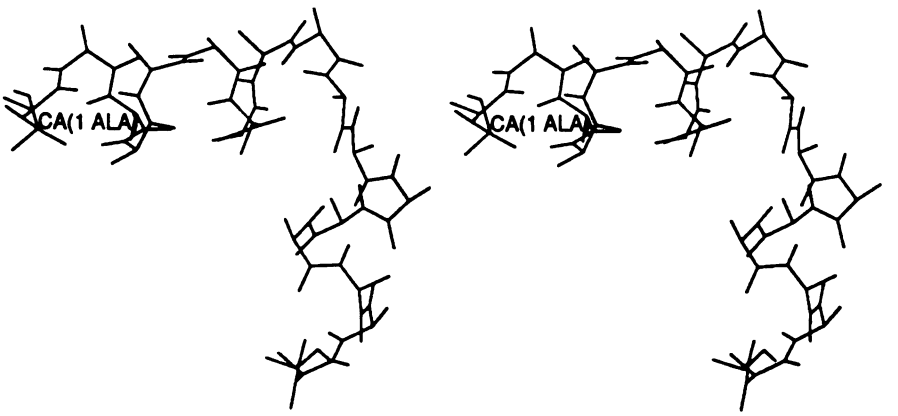


Figure 4.5: continued

E



F

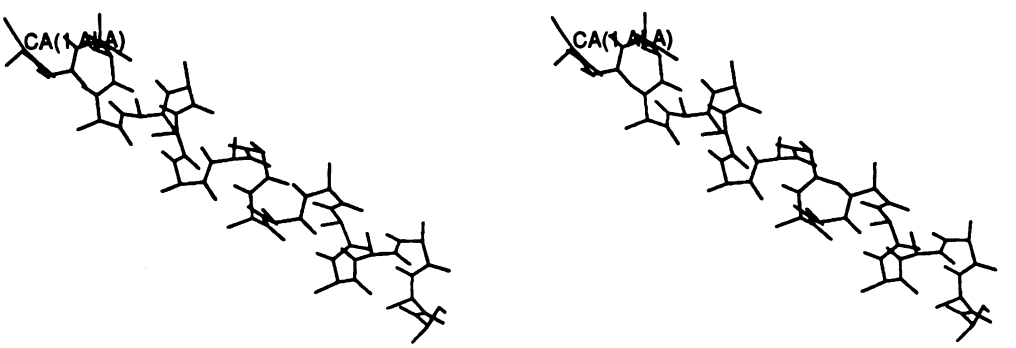


Figure 4.5: continued

tent, 28% helix, and 0% helix.

The presence of multiple helices along the sequence can be seen in Table 4.3. Early in the simulation, 40% of the structures had more than 1 helix and the percentage increased to 62% at the end of the trajectory. The average number of helices per peptide increased slightly from 1.5 to 1.7. The maximum number of helices was 4 in all time frames. The average helix length became shorter with time, dropping from 14 to 10 residues during the trajectory, with mean fluctuations of 3 residues for all time regions (the average length included all helical residues within a structure that were flanked by at least two other residues with helical ϕ and ψ values). The presence of multiple helices is also shown clearly by the distribution of helix lengths with time.

Figure 4.6 contains histograms of helix lengths during the first 500 ps of the simulation and during the last 500 ps (3.5-4.0 ns). The profiles shown in Figure 4.6A are based on the definition of helicity described above. As can be seen there are many more long helical segments (>12 residues) during the first part of the simulation than the last. Even at the start of the simulation, there were many short helices but the average helix length was 9 residues and the overall helix content was 82%. During the last 500 ps, the average helix length dropped to 6 residues but the percentage of helix only dropped to 70%.

Figure 4.6B also shows histograms of helix length during the same periods of time; however, a different definition of helix content was used. Here, helices were defined based on the fractional occupancy of the normal hydrogen bonds (i→i+4, 16 in total for this peptide) during the simulation, where an hydrogen bond was considered intact if the distance between the appropriate amide hydrogen and the carbonyl oxygen was less than 2.5 Å. The helix lengths given in Figure 4.6B represent the number of intact hydrogen bonds in a structure plus 4. The distribution has clearly shifted in favor of long peptides compared to the profiles using the other, more local definition of helicity. Here again shorter helical lengths were observed during the last 500 ps of the simulation. The average number of helical residues was 13 during the beginning of the simulation and

TABLE 4.3
Helical Properties During the Simulation

Simulation Time (ns)	<Helix Length> (residues)	<# Helices / Structure>	% of Structures with > 1 Helix
0 - 1	14	1.5	40
1 - 2	11	1.5	43
2 - 3	10	1.7	60
3 - 4	10	1.7	62

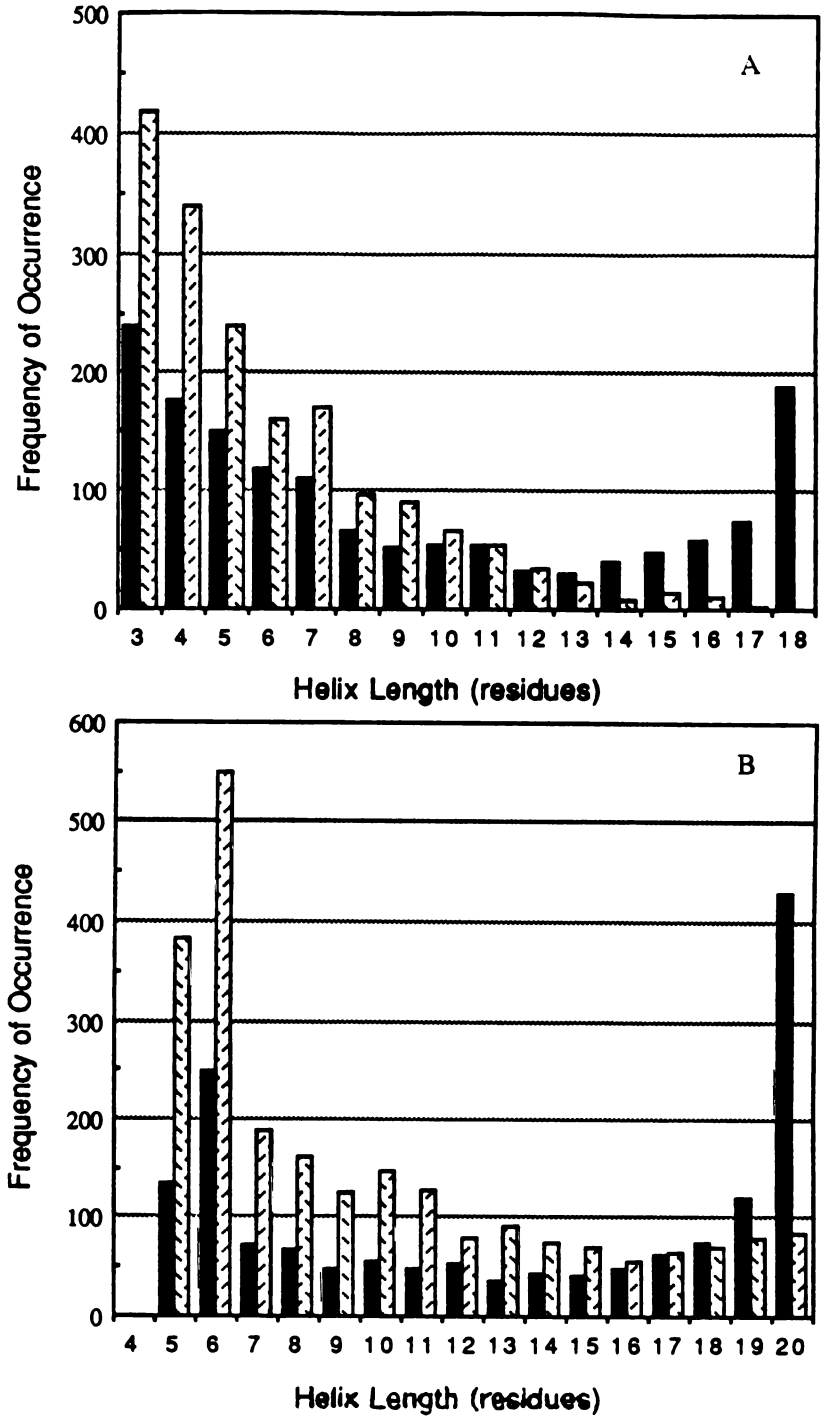


Figure 4.6: Histograms of helix length using different definitions of helicity and different simulation times. (A) Based on dihedral definition given in the text. Solid bars are for 0-0.5 ns portion of the simulation and hatched bars represent the 3.5-4 ns segment. (B) Based on the hydrogen bond definition of helicity. Helix length given by number of hydrogen bonds + 4. The coloring scheme is as described above.

decreased to 10 at the end of the simulation.

To determine the importance of mainchain hydrogen bonds in stabilizing the peptide, we computed the average percentage of time the normal (i→i+4) hydrogen bonds (16 in total for this peptide) were intact during the simulation. For this analysis the cut-off to be considered an hydrogen bond was 3 Å, disregarding the angle of approach. The average percentage of time the hydrogen bonds were intact, based on the distance constraint, was 91% (averaged over all time and all 16 possible hydrogen bonds). The hydrogen bonds were broken for 1.2 ps on average with 15.8 ps time intervals between hydrogen bond breaks. The hydrogen bond stability was slightly higher in the center of the helix than at the ends. At the carboxy-terminus of the peptide $i \rightarrow i+5$ hydrogen bonds became more prevalent and were occupied 10% of the time (i=13-15). In the center of the molecule these hydrogen bonds were only populated 4% of the time (i=4-12). These hydrogen bonds were even less populated at the amino-terminus, 2% of the time.

Energetics of the Peptide

A number of properties were derived from the energies of the intermediate structures generated during molecular dynamics. Figure 4.7 shows the percentage of helix (based on ϕ,ψ values as described above) for particular structures and their corresponding potential energies. There is no real correlation between the potential energy and the percentage of helix, although the 100% helix structures were in general of lower energy than structures with lower helical content. There is a cut-off of approximately -200 kcal/mole. When below this threshold, the structures had a limited range of helicity that was generally high (>70%). Above the threshold there were a large number of structures adopting many different helical contents.

From the potential energies the internal partition function was calculated using the following relationship:

$$Q = \Sigma \exp(-(\varepsilon_i - \varepsilon_0) / RT)$$

where ε_0 is the lowest energy found for the entire ensemble of 8000 structures (-244.2

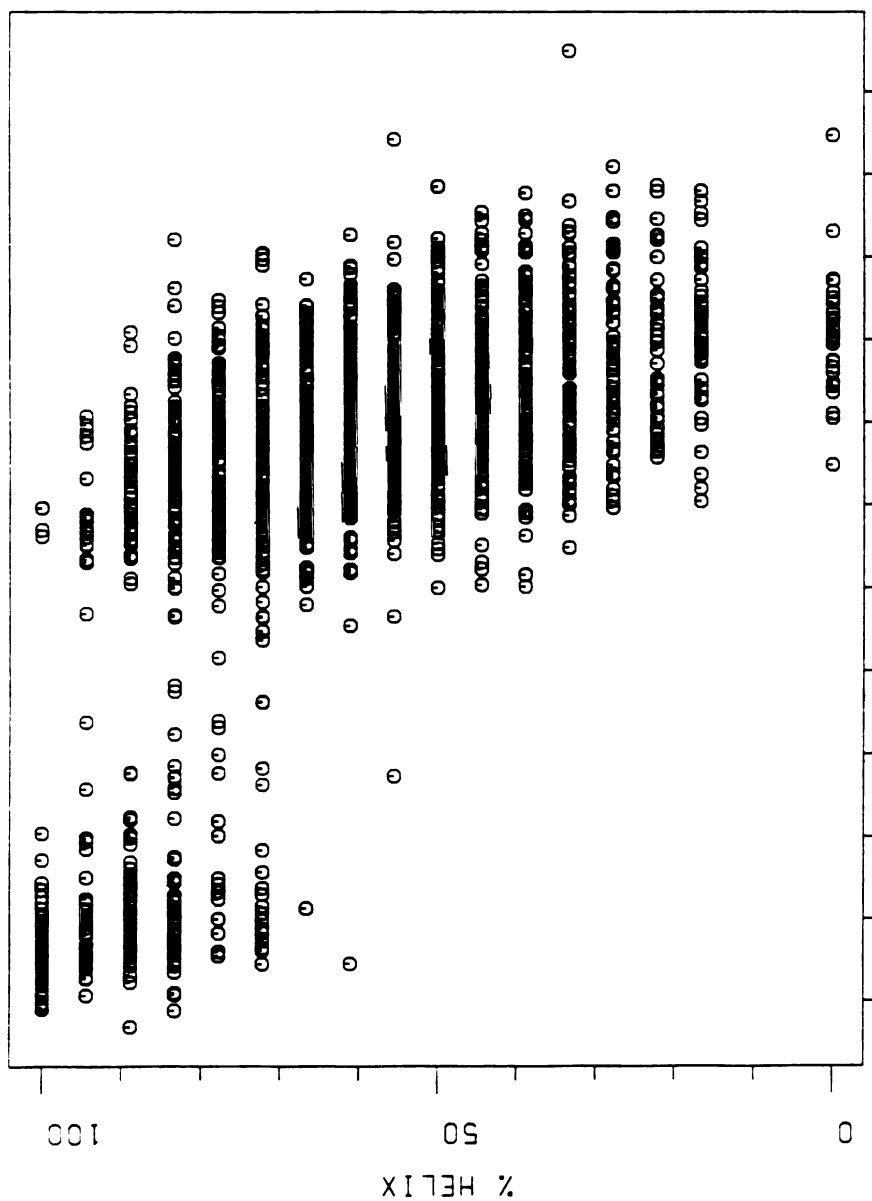


Figure 4.7: Potential energy of a structure plotted against its helix content.

kcal/mole, 94% helix), T = 400 K, R is the gas constant and the sum was over all structures. (Partion functions are usually calculated using internal energies. Since we are correcting the energies using ε_0 and the temperature was maintained by coupling to an heating bath, the kinetic energies cancel.) The partition function was found to be 1.89 and corresponded to a free energy of -0.5 kcal/mole (F = -RT ln Q). The average potential energy during the simulation was -180.4 kcal/mole, or 63.8 when corrected by ε_0. The configurational entropy was calculated from the partition function and the corrected potential energy (S = R ln Q + E / T), yielding a value of 161 eu. Also, we assume that the potential energy is equal to the enthalpy, which is reasonable for simulations *in vacuo*. Using the approach described above, the free energy of the 0% and 100% helical structures was determined by using the same ε_0 as for the full ensemble but only summing over the subset of structures. The free energy for 0% helix was found to be 58.4 kcal/mole, the uncorrected mean potential energy was -159.2 kcal/mole and the mean entropy was 67 cal/mole-K. The 100% helical structures gave F = 0.5 kcal/mole, $< H >$ = -232.3 kcal/mole and S = 28 eu. Although the 100% helix structures had the lowest mean potential energy, the system composed of mixed structures had the lowest free energy. The free energy change in going from the 100% helical structures to the mixed system (with all structures, 62% helix overall) was -1.0 kcal/mole with ΔH = 51.9 kcal/mole and ΔS = 133 eu. Therefore, this process was favorable and entropy driven. Considering the 100% $\rightarrow$ 0% helix case, ΔF = 57.9 kcal/mole, ΔH = 73.0 kcal/mole, and ΔS = 39 eu. So, here again the entropy change favored the nonhelical state, although to a lesser degree. The enthalpy change, however, was large and unfavorable, thereby causing the large positive free energy change. The thermodynamic variables derived using the partition function are only useful for internal comparison of different populations explored during the simulation. These results cannot be compared to experiment because of the lack of solvent and limited sampling of conformational space. Even the delta quantities must be viewed cautiously; the coil state was undersampled relative to the helix.

Even the low energy states are significantly underrepresented; Q is much closer to 1 (the limit for sampling of high energy states, e.g. $\varepsilon_i >> \varepsilon_0$) than to n (the limit for $\varepsilon_i = \varepsilon_0$). In a short simulation (150 ps) of a dipeptide of alanine under the same conditions Q was 120. The partition function for polyalanine is extremely low given that the helical conformation was the most dominant state. This is because all of the energies are relative to the lowest energy, which is for one structure that happened to optimize a number of interactions and is not truly representative of the helical state.

From the fluctuations in the potential energy, the internal heat capacity of the peptide was calculated ($C = R (<\delta E>/RT)^2$). The heat capacity of all 8000 structures was 0.82 cal/gm-deg. The corresponding values for 0% helix and 100 % helix were 0.10 and 0.06 cal/gm-deg, respectively. The mixed structures were, therefore, best able to absorb, or accomodate heat. That the heat capacity of the 0% helix structures is so low is probably due to insufficient sampling of that state and the presence of residual structure in the random coil population. The heat capacities that can be calculated for our system are internal heat capacities. Given our lack of explicit water molecules, we cannot calculate accurate experimentally relevant heat capacities. Nor should changes in heat capacities determined from our simulation necessarily relate to experiment because the contribution to the heat capacity due to peptide-environment interactions should be different for the helical and random coil states.

Structural Transitions

As mentioned above, the transitions between completely helical and nonhelical structures were relatively rare. But, each residue made many such transitions during the simulation. We calculated the number of times each residue converted from residing in the helical region of ϕ,ψ space ($\pm$ 20° of ideal values) to a nonhelical conformation and then back to being helical again. These values were averaged over different intervals of time and are given in Table 4.4. The number of transitions increased with time and reached a limiting value of ~270 transitions/ns after 2 ns, which is ~100 transitions per

TABLE 4.4
Times and Thermodynamic Properties of Various Transitions During the Molecular Dynamics Simulation

	Simulation Time (ns)			
	0 - 1	1 - 2	2 - 3	3 - 4
h → c → h				
$\langle \#\ tran/res \rangle_{h \rightarrow h}$	167	201	270	275
$\langle time \rangle_{h \rightarrow c}$ (ps)	4.0	3.1	2.0	2.0
$\langle time \rangle_{c \rightarrow h}$ (ps)	0.8	1.0	0.8	0.8
$\langle K \rangle_{c \rightarrow h}$	5.2	3.2	2.4	2.3
$\Delta G_{c \rightarrow h}$ (kcal/mole)	-1.3	-0.9	-0.7	-0.7
hh → cc → hh				
$\langle \#\ tran/res \rangle_{hh \rightarrow hh}$	59	80	108	112
$\langle time \rangle_{hh \rightarrow cc}$ (ps)	10.8	9.0	6.3	6.0
$\langle time \rangle_{cc \rightarrow hh}$ (ps)	1.2	2.1	1.5	1.5
$\langle K \rangle_{cc \rightarrow hh}$	8.8	4.3	4.2	4.0
$\Delta G_{cc \rightarrow hh}$ (kcal/mole)	-1.7	-1.2	-1.1	-1.1
hhh → ccc → hhh				
$\langle \#\ tran/res \rangle_{hhh \rightarrow hhh}$	19	28	38	43
$\langle time \rangle_{hhh \rightarrow ccc}$ (ps)	33.7	27.7	20.1	17.9
$\langle time \rangle_{ccc \rightarrow hhh}$ (ps)	2.1	4.9	3.1	3.0
$\langle K \rangle_{ccc \rightarrow hhh}$	16.2	5.6	6.7	5.9
$\Delta G_{ccc \rightarrow hhh}$ (kcal/mole)	-2.2	-1.4	-1.5	-1.4
hhh → hhc → hhh				
$\langle \#\ tran/res \rangle_{hhh \rightarrow hhh}$	68	75	104	103
$\langle time \rangle_{hhh \rightarrow hhc}$ (ps)	5.8	5.4	3.8	4.0
$\langle time \rangle_{hhc \rightarrow hhh}$ (ps)	1.2	2.0	1.8	1.7
$\langle K \rangle_{hhc \rightarrow hhh}$	4.9	2.7	2.1	2.3
$\Delta G_{hhc \rightarrow hhh}$ (kcal/mole)	-1.3	-0.8	-0.6	-0.7
hcc → hhc → hcc				
$\langle \#\ tran/res \rangle_{hcc \rightarrow hcc}$	26	38	55	54
$\langle time \rangle_{hcc \rightarrow hhc}$ (ps)	4.3	4.4	4.1	4.0
$\langle time \rangle_{hhc \rightarrow hcc}$ (ps)	11.5	11.3	7.3	7.0
$\langle K \rangle_{hcc \rightarrow hhc}$	2.6	2.6	1.8	1.8
$\Delta G_{hcc \rightarrow hhc}$ (kcal/mole)	-0.8	-0.8	-0.4	-0.4
ccc → chc → ccc				
$\langle \#\ tran/res \rangle_{ccc \rightarrow ccc}$	10	20	29	27
$\langle time \rangle_{ccc \rightarrow chc}$ (ps)	13.9	11.0	8.9	10.0
$\langle time \rangle_{chc \rightarrow ccc}$ (ps)	26.1	15.3	12.8	12.5
$\langle K \rangle_{ccc \rightarrow chc}$	1.9	1.4	1.4	1.2
$\Delta G_{ccc \rightarrow chc}$ (kcal/mole)	-0.5	-0.3	-0.3	-0.2

nanosecond more than at the beginning of the simulation. The number of transitions dropped by over 100 transitions per nanosecond when two residues were required to convert between the helical and nonhelical conformations (hh→cc→hh, Table 4.4). The number of transitions dropped further when three residues were considered (hhh → ccc → hhh), but again increased with time.

Table 4.4 also contains the average amount of time between the different states, or the amount of time the residue(s) spends in a particular state before converting. The single residue transitions give a measure of the inherent background motions in the structure. In the first nanosecond of the simulation the residues spent a much greater time in the helical state than in the coil, or nonhelical, state. After a nonhelical conformation was adopted, the residue quickly reverted back to the helical conformation ($\langle K \rangle_{c \rightarrow h}$=5.2, greatly favoring the helical state, Table 4.4). The nonhelical region became more favorable as the simulation progressed ($\langle K \rangle_{c \rightarrow h}$=2.3), but the helical conformation was still favored. The equilibrium constants were averaged over all of the residues but there were differences along the sequence. From ratios of equilibrium constants, the N-terminal region (residues 1-4) favored the nonhelical conformation three times as much as in the center of the structure. The C-terminal end of the peptide (residues 15-19) was slightly more helical than the N-terminus, with the coil state favored 2.5 to 1 compared to the center of the structure.

It should be noted that because of our definition for assigning helical regions, we cannot determine the time necessary for the actual conversion, it is merely the time interval between the structures collected. Instead, the rate constants have been approximated as reciprocals of the observed time intervals between structures. These pseudo-rate constants were used to determine the equilibrium constants and consequently the free energy changes. The average free energy change for the conversion from a nonhelical to an helical conformation reached a plateau of ~-0.7 kcal/mole after 1 ns. Hence, transitions at the single residue level should be easily accessible even at lower temperatures.

The disparity between the amount of time spent in the helical versus nonhelical conformation was magnified further when more residues were considered. For example, the average equilibrium constant almost doubled for the cc→hh transitions compared to the single residue transitions (comparing values after the system had unambiguously equilibrated, 2-4 ns, Table 4.4). The free energy for the conversion of 2 residues was more favorable (ΔG = -1.1 kcal/mole) than the single transition (ΔG = -0.7 kcal/mole), but the effect was not additive. In considering the conversion of three residues, ccc→hhh, the free energy became more favorable (~-1.5 kcal/mole after 2 ns, Table 4.4).

We then considered a slightly different transition, something akin to propagation. The free energy was calculated for converting a nonhelical residue to an helical one with two neighboring helical residues to, presumably, aid in the process. We found that the free energy changes for this process were roughly equivalent to the values for the conversion itself, without regard to the state of the preceding residues. This finding is surprising but may be, due to the fact that the peptide was mostly helical during the simulation. Although the free energy for adding an helical residue to an existing helical segment was equivalent to the same transition at the single residue level, the number of transitions was comparable to that seen for the two residue transitions.

To test how important the two preceding helical residues were to propagation of the helix, the following conversion was investigated: hcc → hhc (Table 4.4, ΔG = -0.4 kcal/mole). The free energy change for this process was less favorable than for a single residue transition (ΔG = -0.7 kcal/mole). Since the free energy for the previous propagation transition was equivalent to the conversion of a single residue, the difference between these two propagation processes is probably due to the fact that hcc → hhc occurs for residue i at the helix-coil interface. The hhc → hhh conversion neglects the state of residue $i + 1$. The equilibrium constant for propagation is comparable to the s parameter of Zimm-Bragg theory. From our simulation s ~ 2 (averaged over the last 2 ns for the two propagation processes, hcc→hhc and hhc→hhh).

The final transition considered in Table 4.4 is for nucleation of an helix within a nonhelical region, ccc → chc. The free energy change for this process was favorable but less than the single residue conversion, c → h. Hence, this transition was more difficult than the established background level. This process is equivalent to the nucleation step of the Zimm-Bragg model, therefore $\langle K \rangle_{ccc \rightarrow chc} = \sigma s$, where s ~ 2 for our simulation and σ is the nucleation parameter. From our simulation σ = 0.6.

To determine whether or not the various transitions discussed above occurred cooperatively, the probabilities of the transitions occcurring independently was estimated and compared to their observed frequency. The equilibrium constant for an helical to nonhelical transition was used as a measure of the probability of the change to occur. For each type of transition shown in Table 4.4, only the values from the last nanosecond were considered to ensure that the system had completely equilibrated and because there were more occurrences of the events late in the simulation.

For the single residue transitions the probability that an helical residue will convert to a nonhelical residue is $\langle K \rangle_{h \rightarrow c} = P_c = 0.43$. Since the residue must either convert to a nonhelical residue or remain helical the total probability is 1 and the probability of remaining helical is $P_h = 0.57$. Therefore, the helical state is favored slightly over the coil, or nonhelical, state. For the transitions involving two residues, hh → cc, the probability should be the product of the individual probabilities for a transition ($P = P_c P_c = 0.18$) if the events are independent. From the simulation, the probability was 0.25. So, the events were not independent----it was easier to make the second nonhelical residue after the first had converted. The same local cooperativity was found when three residues were considered, ccc → hhh. If the three transitions were to occur independently, a probability of 0.08 would be expected. Instead a frequency of 0.17 was observed from the simulation. We can assume that the first residue to convert to a nonhelical state does so independently because the propagation step, hhc → hhh, gave a probability of 0.43, which is expected for an independent transition. Therefore, conversion of the other two

residues, after the first transition, was facilitated by the neighboring coil residue.

This same local cooperativity was seen for the other propagation step investigated, hhc $\rightarrow$ hcc. Again, a probability of $P_c = 0.43$ is expected for an independent transition, but a value of 0.57 was observed from the simulation. Therefore, this result also indicates that the helix to coil transition is more probable when there is a neighboring nonhelical residue. The final case that we considered was for nucleation of an helix, chc $\rightarrow$ ccc. A probability of 0.8 was found from the simulation. Therefore, this process was cooperative such that the neighboring nonhelical residue facilitated the converstion of h $\rightarrow$ c, although the process was still energetically favorable.

DISCUSSION

We have performed a molecular dynamics simulation of a 20 residue peptide of alanine in an attempt to characterize the types of motions the structure undergoes and to investigate the mechanism of the structural transtions. This peptide was predominantly helical during the simulation---62% helix during the entire 4 nanoseconds. The α-helix is an important structural element of proteins, therefore an understanding of helix dynamics can aid in interpreting the motions of proteins. Furthermore, the helix-coil transition is of interest because of its probable role in protein folding. In the discussion that follows our results are compared to experimental results of related small peptides, to previous simulations, and to earlier predictions from polymer theory. Lastly, we present the relevance of these results to protein folding.

Comparison to Experiment

There are still only a handful of small peptides that form isolated α-helices in solution. Unlike long peptides, these small peptides do not show cooperative melting curves (Bradley personal communication; Mitchinson and Baldwin, 1986; Marqusee and Baldwin, 1987; Marqusee et al., 1990). The transitions observed during our simulation were not cooperative, either. We never saw direct transitions between 0% and 100% helical structures. Instead, intermediate states were populated. This was also true, in general,

for the less restrictive transtions between 90% and 10% helix and 80% and 20% helix. In general during our simulation, transitions occurred in a stepwise manner with one residue transforming at a time. We did, however, observe local cooperativity between neighboring residues (discussed further below). The transition times for these processes ranged from 10^{-11} - 10^{-8} sec (Table 4.2). Experimentally, the range is 10^{-9} - 10^{-6} sec. Our values overlap with the experimental range but are high, mostly due to the lack of solvent (discussed further below).

Experimentally, small peptides are only marginally stable as helices in solution. As a result, one observes fractions of helix as measured by circular dichroism (CD). Specific helical regions have been identified in small peptides by two-dimensional nuclear magnetic resonance experiments (2-D NMR) (Osterhout et al., 1989; Bradley et al., 1990), as opposed to observing 100% helical structures in equilibrium with 0% helix structues. In our simulation, we found that the mixed state, with various helical contents, was lower in free energy than the 100% helix state. That the mixed state was more favorable was because of the favorable entropy, or conformational freedom, of the mixed state, while the average potential energy was lower for the 100% helix state. The free energy difference two states was 1 kcal/mole and therefore interconversion was facile. The mixed state had a much lower free energy than the completely nonhelical state (~ 60 kcal/mole). Hence, our results are consistent with the experimental data as our peptide favors the mixed helical state, as do the experimentally characterized peptides. However, the free energy difference involving the nonhelical state have almost assuredly been overestimated because of the limited sampling of that state.

For many years it was thought that small peptides could not adopt helical conformations in solution (Epand and Scheraga, 1968). This was due in part to experiments, conducted to determine the s and σ parameters from the Zimm-Bragg model (Zimm and Bragg, 1959), where s represents an equilbrium constant for propagation of the helix and σ describes nucleation of an helix. The experiments were performed using the host-guest

method in which the effect of introducing an amino acid (guest) into a water soluble homopolymer (host) with a well-characterized helix-coil transition. Using this method, s = 1.06 and $\sigma = 8 \times 10^{-4}$ for L-alanine (Scheraga, 1973). These values indicate that nucleation of helical segments is difficult and that the residue only favors the helical state slightly over the coil conformation. In contrast, s derived from our simulation was ~ 2 and σ ~ 0.6. These results show that the helical state was greatly favored over the nonhelical conformation and that nucleation was less probable than propagation (e.g. $\sigma = 1$ for no barrier to nucleation) but the values are much higher than those derived using the host-guest method. Hence, it is much easier to make helices from alanine in our simulation than suggested by the host-guest results and nucleation of short fragments of helix are relatively facile. The lack of solvent and use of a macroscopic dielectric function can facilitate the helix to coil transitions and may overestimate s and σ. However, there is recent experimental work supports our findings. Baldwin and co-workers have synthesized a variety of alanine-based peptides that are stable as isolated α-helices in aqueous solution (~80% helix content) (Marqusee and Baldwin, 1987; Marqusee et al., 1990; Padmanabhan et al., 1990). They report a preliminary s value of ~ 2 for alanine (Marqusee et al., 1990). This value is in accord with the value from this simulation. Also, the fact that the peptide even adopts the helical conformation with a transition that is weakly cooperative, suggests that σ is greater than the value derived from host-guest experiments. Unfortunately, an experimental value from a small peptide is not available to compare to the value derived from our simulation. It has been argued that because of the small nucleation parameter from host-guest studies, short helical fragments will not exist but instead formation of long helices is favored. The fact that we see segments of helix broken by nonhelical regions is supported qualitatively by NMR data on long peptides. The NMR results indicate that, under conditions that support interconverting helical and random coil configurations of polyalanine, there are random coil segments in the midst of helical regions (Glick et al., 1966).

Comparison to Simulations

Equilibrium Properties

A number of techniques have been employed to study α-helix dynamics theoretically: normal mode calculations (Fanconi et al., 1971; Levy and Karplus, 1979); molecular dynamics (Levy et al., 1982); Monte Carlo simulations (Skvortsov et al., 1971); and harmonic dynamics in dihedral space (Go and Go, 1976). Both Go and Go (1976) and Skvortsov et al. (1971) found dihedral fluctuations of the backbone of 7-8 °. The fluctuations (15-16°) that we observed are more consistent with the values found by Levy and Karplus (1979) using harmonic dynamics (12-15°). In other simulations an 18 residue peptide of polyalanine, was reported to have root-mean-square displacements of the end-to-end length of 0.24 Å (Suezaki and Go, 1976) and 0.38 Å (Peticolas, 1978). The corresponding value from our simulation was 2.4 Å. Our value is much higher than those found earlier. This is because the previous studies were either of short duration or search available conformational space exhaustively but only in the harmonic limit for one particular structure. Our simulation, on the other hand, was sufficiently long that we sampled both the helical state and numerous nonhelical conformations.

A number of motions were seen in our MD simulation that had been observed previously in normal mode studies. Many of the end-to-end distances in Figure 4.1 that are less than the ideal α-helix length correspond to accordian-like motions of the whole helix, which is made up of two types motion occurring either separately or simultaneously. These two motions involve a bending of the helix causing the ends of the helix to move in space and a contraction of the structure leading to improved hydrogen bonds or shifting of hydrogen bonds from $i + 4 \rightarrow i + 5$ spacings, resulting in an helix with a larger cross section such that the ends of the structure only move along the helix axis. Fanconi et al. have also observed accordian-like motions in normal mode calculations of an α-helix (1971). The lowest nonzero frequency mode of a polymer is expansion and contraction (Schaufele and Shimanouchi, 1967). For helices of finite length these

accordian-like motions become optically active and have been observed in Raman scattering of hydrocarbons (Schaufele and Shimanouchi, 1967). Levy and Karplus also observed this type of contraction, yielding an helix with a large cross section and improved hydrogen bonds (1979). They note that there is asymmetry in the energy contour such that it is much easier to compress than stretch an helix. We found this in our simulation as well, almost all of the deviations from the ideal length involved compression of the helix (Figure 4.1). In addition to the accordian-like motion, we observed fraying of the ends of the helix. Levy and Karplus also observed this motion, reporting higher dihedral fluctuations at the ends of the structure than in the center (1979).

Our simulation shows similar results for helix motion as previous studies. We have the advantage, though, that we were also able to sample nonhelical phase space to observe unwinding of the helix and then reformation of the helical state. Most unfolding steps started because of unwinding of the ends of the structure but some involved unfolding emanating from the center of the structure. This process usually started as the result of a kink that was then propagated. None of the transitions that we observed were strictly cooperative; however, there was strong local cooperativity in that the fluctuations of adjacent dihedral angles were highly correlated. This has also been observed by Levy and Karplus (1979) and Go and Go (1976). Go and Go find that fluctuations were correlated as far as 6 residues apart, although the magnitude beyond 4 is small. Levy and Karplus (1976), on the other hand, only see correlations spanning 3 residues. Our results (Figure 4.3) are in accord with those of Levy and Karplus.

Mechanism of Transition

A number of investigators have explored simulation of the helix-coil transition. McCammon and co-workers have addressed this question using dynamics with simplified models of the polypeptide chain (Flory virtual bonds connecting soft spheres) with inclusion of terms to approximate the effects of solvent (McCammon et al., 1980; Pear et al., 1981). Their structure was fixed in space except that the last two or five residues were

allowed to move. In both simulations unwinding of a residue occurred with, or closely followed, unwinding of the next residue, such that when two or more residues left the helix that they did so sequentially. Furthermore, they found that correlation of motion affects the transition rates. Correlations between dihedrals have also been suggested to play a role in the dynamics of polymers in solution to minimize large solvent frictional forces caused by the tails being forced to follow rigidly the rotation of the transforming bond (Helfand, 1984). In these systems, too, transitions occur sequentially with an activation energy of approximately 1 barrier height; so, the events are essentially independent overall but locally cooperative (Skolnick and Helfand, 1980). Furthermore, Helfand and co-workers have found that immediately following the transition of one bond there is a strong increase in the transition rate of the second neighbor's bonds, lasting for short periods of time (Helfand, 1984). We also observed this effect as transformation of a residue to the coil state was facilitated after a neighboring group became nonhelical. The reason for the strong correlations between dihedral fluctuations of neighboring residues is that in order to localize the mode, the transition of the central bond must be accompanied by distortion of neighboring degrees of freedom (Hall and Helfand, 1982; Helfand et al., 1981). The anticorrelation between ϕ_i and ψ_{i-1} is an example of how compensatory motion can localize distortions. The correlation of the motions of these dihedral angles causes rotation of the plane of the amide group with only a small effect on the overall chain direction. That one residue can respond to a change in its neighbor is understandable in peptides because the intrinsic barrier to rotation about ϕ and ψ are relatively small (~1 kcal/mole) (Tsuji et al., 1976; Gruenewald et al., 1979).

McCammon et al. also calculated the parameter s for their system (1980). They saw insensitivity of s with respect to chain position, which is in accord with the Ising model assumptions. We, however, observed differences in s along the sequence. The s parameter was approximately 2.5-3 times higher at the ends (4 residues at each end) than in the center of the structure. McCammon co-workers (1980) did not see this effect because

they only allowed motion in the region we would define as the end residues. Thus, they never investigated the interior of the helix since it was fixed.

McCammon and co-workers also determined the rate constants for the helix-coil transition in their peptide (1980). They found rates of 10^8 - 10^9 sec^{-1} for individual residues. The experimental results suggest rate constants of 10^7 - 10^{10} sec^{-1} for a variety of polypeptides in different solvents (Gruenewald et al., 1979; Bosterling and Engel, 1979; Tsuji et al., 1976; Inoue et al., 1979). Our values range from 10^{10} to 10^{12} sec^{-1} for individual residue transitions. Our results are obviously high. There are at least two factors leading to these high values: the use of high temperature to facilitate bond rotations and the lack of solvent which would damp motion.

Czerminski and Elber have used a different approach to study helix-coil transitions (1989). They completed a reaction path study of a tetrapeptide in which they delineated the conformational transitions between the 112 stable states of the molecule. They find that the number of available routes for a transition is significantly lower than in a random search and occur via local dihedral flips. As has been observed with polymers, Czerminski and Elber find that a transition is close to completion before the next transition is initiated. They also observed a quasi-melting point 5 kcal/mole above the lowest energy minimum. Below that point the molecule is trapped in one or a few minima and above the threshold a large number of configurations are observed. Our simulation showed similar behavior, with a quasi-melting point ~40 kcal/mole above the lowest energy (Figure 4.7).

Comparison to Polymer Theory

Statistical Properties of Random Polypeptides

Much work has been devoted to the statistical properties of random flight chains, or freely rotating chains. The model is generally just a chain of Flory virtual bonds, vectors joining C_α atoms, that are unperturbed by excluded volume effects. Many conformations for the chains are generated and these are then evaluated statistically. One of the most

important properties derived using this approach is the mean squared end-to-end distance, $\langle r^2 \rangle$, because it reflects the distribution of conformations. From the end-to-end distance the characteristic ratio can be calculated, which gives a measure of the inherent stiffness of a real chain compared to the random flight chain. The characteristic ratio is calculated from the end-to-end distribution by the following relationship:

$$C_n = \langle r^2 \rangle_0 / nl^2$$

where n=number of resides, l= the distance between virtual bonds. The dimensions of real chains increase with n to some limiting value. The characteristic ratio calculated from our simulation was 3 with a maximum value of 5. The limiting value for polyalanine as n $\rightarrow$ infinity is 9.0, which has been determined experimentally (Brant and Flory, 1965) and estimated theoretically (Brant et al., 1967). When C_n becomes independent of n the chain is behaving like a random walk polymer, which occurs when n > 35 residues. The characteristic ratio is between 6-7 for a 20 residue peptide of alanine (Schimmel and Flory, 1967). Thus, small peptides are not expected to behave like random walk chains. But, our value is low even compared to the value for short chains. This discrepancy indicates that we did not sample the random coil state very well in our simulation. Instead, our peptide remained very compact. This is probably due to our use of a distance dependent dielectric constant, which overestimates electrostatic interactions and leads to compaction of structures (discussed in Chapter 3), resulting in the peptide becoming trapped in certain regions of conformational space. Further simulations with higher temperatures and longer simulations times could be performed in an attempt to circumvent the local minimum problem.

The distribution of the end-to-end lengths is also of use. The distribution derived from our simulation (Figure 4.1B) deviates markedly from the Gaussian distribution expected for a freely rotating chain. All real chains are described well by a Gaussian function when n $\rightarrow$ infinity (Cantor and Schimmel, 1980). Short chains, however, often exhibit deviations from this behavior. The distribution from our simulation indicates that

our structure is stiff and that our profile is a composite of two distributions, helix and coil.

Statistical Models for Helix-Coil Transitions

Previous work in this area has held as its tenet the tendency to confine helical units to one long sequence rather than to scatter short segments. The main reasons for this are that the propagation of long helices circumvents the entropy cost associated with bringing residues together to make the first hydrogen bond for each segment, since the entropy loss of adding successive residues is lower. Also, due to the alignment of peptide dipoles, the dipolar stabilization is maximal if a given residue is followed by 10 helical residues. These factors explain the cooperativity observed for helix-coil transitions of long peptides. Short peptides like the one studied here do not show cooperative transitions, however. It is our aim to see how well the theory correlates with properties observed during our simulation.

In the zipper model of helix-coil transitions only one helical segment is allowed per chain and nucleation is much less favorable than propagation. The nucleation parameter, σ, determines how sharp the transition is and is usually assigned a value of 10^{-3} - 10^{-4}. Schellman used this model to study how distributions of helix lengths change during the helix-coil transition in a small 20 residue peptide (1958). He found that short helices (1-2 hydrogen bonds) contribute very little to the thermodynamic properties of the transition and the overall picture of the distribution function. The distributions observed showed a peptide with essentially all residues in the helical state. As the transition occurred the population shifted to lower helix lengths but instead of significantly populating small segments of helix, the peptide with no helix content became prevalent. Contrary to Schellman's results, short helical segments were highly populated in our simulation both early in the simulation when the peptide was predominantly helical and later when mixed random coil and helical structures were in equilibrium (Figure 4.6). Also, in our simulation small helices can contribute to the thermodynamic properties, provided there are

multiple helices. When just a single short helix existed, its contribution to the partition function was small. This result is in accord with Schellman's work since the zipper model only allows the existence of one helix. Schellman also reports that fluctuations can protect the peptide against total unfolding because the free energy of the fluctuant states is lower than that of the helix with all hydrogen bonds satisfied (1958). We also found this to be the case in our simulation.

A more accurate statistical method for evaluating helix-coil transitions involves a matrix approach. In this model, helical and coil units are free to occur anywhere along the chain, in contrast to the zipper model. The Ising model for a one-dimensional crystal is generally used in this treatment. One of the best known examples is the model of Zimm and Bragg (1959), which clearly shows how changes in s and σ affect the transition and types of conformations observed. Many other workers have also used the Ising model to address different aspects of the helix-coil transition (ex. Flory and Miller, 1959; Peller, 1959a, 1959b; Schwarz, 1965, 1968). Although this model allows multiple helices, they are not probable for short chains (Zimm and Bragg, 1959). Furthermore, the appearance of breaks in an helical region causes the random coil content to increase by conversion of the adjacent helical regions to nonhelical conformations (Peller, 1959a). This was observed in our simulation as well.

For quite some time, the Zimm-Bragg model (1959) along with host-guest data (Scheraga, 1973) led people to believe that small isolated helices would not form in solution. After Baldwin's research group clearly showed that they do form (Bierzynski et al., 1982; Shoemaker et al., 1985, 1987), Scheraga and co-workers reported an extension of the Zimm-Bragg theory (Vasquez et al., 1987). The new formulation incorporates the effect of specific peptide charge-dipole interactions on helix stability and is able to predict reasonably accurately the overall helix probabilities of various C-peptide derivatives investigated by Baldwin and co-workers. Our results, on the other hand, suggest that small peptides may form helices in the absence of these specific side chain interac-

tion or interactions with the helix dipole. To test whether that the helix stability we observed is generally valid, as opposed to being an artifact of our simulation conditions, further simulations with other amino acids and mixed sequences should be performed. In addition, the simulations should be performed with explicit water molecules. Currently, this is not a feasible endeavor because of the large amounts of computer time that would be necessary to get efficient sampling. Instead, the use of other macroscopic dielectric models may be warranted. We have had good results on another small, helical peptide with one such model for which the dielectric constant is sigmoidally dependent on the distance between the charges (Chapter 3). Despite these limitations, our hypothesis is borne out by recent studies of alanine-based peptides that form α-helices in solution without specific side chain interactions by Marquesee and Baldwin (1990).

There is considerable disagreement betwen the values for conversion of a single residue from the helical to the nonhelical state, both in terms of the experimental results and theoretical predictions. These discrepancies are discussed by Zana, who suggests that the most probably rate constant for conversion is 10^8 sec^{-1} (1975). Gruenewald et al. (1979) suggest that 10^{10} sec^{-1} is an upper boundary value for this process. A value of 10^{11} sec^{-1} was determined from our simulation. Therefore, there is a discrepancy between our value and those found experimentally and theoretically. That we have simulated a small peptide while previous work was performed on long chains may account for the difference. Also, our value of σ is considerably higher (by 2-3 orders of magnitude) than the value used to predict properties from the theory and to interpret the experimental data.

The propagation of an helix is considered to be made up of two steps: rotation about the dihedral angles to place the residues into the helical region of conformational space and subsequent hydrogen bond formation (Gruenewald et al., 1979). Of these two steps, rotation of φ and ψ is the limiting factor. Therefore, our use of high temperature and an *in vacuo* model, which facilitates rotation, leads to an overestimation of the transition time.

Also, we used a more sensitive definition (based on ϕ, ψ values) of helix content than was used previously. When we calculated the helix content using the hydrogen bond definition that is generally used in statistical mechanics literature, the fractional helicity was four residues higher than the corresponding values using our definition. Hence, our definition defines some conformations as nonhelical that are considered helical by the hydrogen bond definition. We favored the dihedral definition for calculating the structural properties of the peptide because proteins have approximately 50% regular structures (Levitt and Greer, 1977; Kapsch and Sander, 1983), while approximately 90% of the mainchain hydrogen bonds are intact (Baker and Hubbard, 1984). Our more local definition of helix content can also be justified when considering the experimental methods generally used to characterize small peptides: CD and 2-D NMR. Woody and co-workers have suggested that distortions of helices can lead to diminished signal intensities (Manning et al., 1988). Distortions of the helix were prevalent throughout our simulation and were better characterized by the ϕ, ψ definition of helicity. The nuclear Overhauser effect is even more short-range and coupling constants are very sensitive to the conformation of the individual peptide units.

Relevance to Protein Folding

That isolated helical fragments can exist in solution is supportive of the framework model of protein folding. This model proposes that secondary structure is formed early in folding and that these preformed, marginally stable units of secondary structure coalesce to form tertiary structure (Ptitsyn and Rashin, 1975; Karplus and Weaver, 1976; Kim and Baldwin, 1982). This hypothesis implies that secondary structure should be present under conditions where folding occurs spontaneously. It is, therefore, of interest to study the inherent motions and transitions that small helices undergo. There is another finding that suggests that studies of peptide fragments are relevant to the folding of proteins; the rate of the helix-coil transition has the same temperature dependence in small molecules as in the backbone of macromolecules (Morawetz, 1979). Also, Czerminski and Elber (1989)

argue that since the transitions they observe between the minima of a tetrapeptide proceed via localized conformational changes, simulations on small molecules are likely to be of importance to the processes occurring within large polypeptides.

Our results indicate that both the s and σ parameters describing the likelihood of propagation and nucleation, respectively, of helices have been underestimated both experimentally and theoretically for small peptides. Since alanine has a strong helical propensity by a variety of measures (Chou and Fasman, 1978; Williams et al., 1987), the s and σ parameters are expected to be lower for other residues. Even so, this should not be just an artifact of a simulation of polyalanine and we expect that other residues will show like behavior.

Our simulation shows that the peptide spends the bulk of its time fluctuating between different conformations with intermediate helix contents. Transitions between highly ordered and highly disordered structures were rare, but they occurred very quickly. Our distribution of conformations favored collapsed states. Hence, our transitions to structures with high helical content were from highly fluctuating, compact structures. Thus, folding in our system is from a collapsed, heterogeneous population of structures with varying degrees of secondary structure, akin to a molten globule state (Dolgikh et al., 1981). From this state, there was fine tuning of the dihedral angles and hydrogen bonds to form more specific units of secondary structure. In the case of a protein, packing of secondary structure would follow. This progression of events represents a combination of the framework model and the collapse model. Molten globules have been proposed as early intermediates in protein folding (Dolgikh et al., 1981; Ohgushi and Wada, 1983). They are thought to accelerate folding by restricting the amount of conformational space that needs to be sampled by allowing further folding to occur in a condensed state. Our nonhelical population was not a true random coil population when compared to expected properties from polymer theory, instead it was more compact. Recent experimental studies showing residual structure in denatured proteins (Dobson et al., 1990) suggest that

experimentally the so-called random coil population may be closer to the molten globule state than to states expected from early statistical mechanical theories. Recent statistical mechanical studies by Chan and Dill (1990) show that any flexible polymer molecule will adopt secondary structure as it is driven to compactness. In our case, the peptide is compact by the use of an *in vacuo* model with $\varepsilon = r$, but, experimentally, small peptides would also be expected to be driven to compactness by the tendency to reduce their surface area. To test our proposal for folding, it would be worthwhile to perform related simulations of other homopeptides as well as heteropeptides both using *in vacuo* and solution models.

REFERENCES: For Chapters 2-4

1. Baker, E.N. and R.E. Hubbard (1984). Prog. Biophys. Mol. Biol. 44: 97.
2. Bash, P.A., Singh, U.C., Langridge, R., and P.A. Kollman (1987). Science 236: 564.
3. Bash, P.A., Singh, U.C., Brown, F.K., Langridge, R., and P.A. Kollman (1987). Science 235: 574.
4. Bierzynski, A., Kim, P.S., and R.L. Baldwin (1982). Proc. Natl. Acad. Sci. USA 79: 2470.
5. Bosterling, B. and J. Engel (1979). Biophys. Chem. 9: 201.
6. Bradley, E.K., Thomason, J.F., Cohen, F.E., and I.D. Kuntz (1989). In Techniques of Protein Chemistry. Academic Press, Inc., T.E. Hugli, Editor, 531.
7. Bradley, E.K., Thomason, J.F., Cohen, F.E., and I.D. Kuntz (1990). J. Mol. Biol., in press.
8. Brant, D.A. and P.J. Flory (1965). J. Am. Chem. Soc. 87: 2788.
9. Brant, D.A., Miller, W.G. and P.J. Flory (1967). J. Mol. Biol. 23: 47.
10. Brooks, B.R. (1989). Chemica Scripta. 29A: 165.
11. Brown, J.E. and W.A. Klee (1971). Biochem. 10: 470.
12. Brown, F.K. and P.A. Kollman (1987). J. Mol. Biol. 198: 533.
13. Cantor, C.R. and P.R. Schimmel (1980). Biophysical Chemistry Part III, W.H. Freeman and Company, San Francisco.
14. Chan, H.C. and K.A. Dill (1990). Proc. Natl. Acad. Sci. USA, in press.
15. Chou, P.Y. and G.D. Fasman (1978). Adv. Enzym. 47: 45.
16. Conway, B.E., Bockris, J.O., and I.A. Ammar (1951) Trans. Faraday Soc. 47: 756.
17. Cooke, R. and I.D. Kuntz (1974). Ann. Rev. Biophys. Bioeng. 3: 95.

18. Crawford, J.L., Lipscomb, W.N., and C.G. Schellman (1973). Proc. Natl. Acad. Sci. USA 70: 538.

19. Creighton, T.E. (1983). Proteins: Structure and Molecular Principles, W.H. Freeman and Company, New York.

20. Czerminski, R. and R. Elber (1989). Proc. Natl. Acad. Sci. USA 86: 6963.

21. Debye, P. (1929). Polar Molecules, Dover Publications, New York.

22. Dobson, C.M. Henley, C., Radford, S.E., Baum, J., and P.A. Evans (1990). Paper presented at American Association for the Advancement of Science Protein Folding Seminar.

23. Dolgikh, D.A., Gilmanshin, R.I., Brazhnikov, E.V., Baychkova, V.E., Semisotnor, G.V., Venyaminov, S.Y., and O.B. Ptitsyn (1981). FEBS Lett. 136: 311.

24. Epand, R.M. and H.A. Scheraga (1968). Biochem. 7: 2864.

25. Fanconi, B., Small, E.W. and W.L. Peticolas (1971). Biopolymers 10: 1277.

26. Flory, P.J. and W.G. Miller (1966). J. Mol. Biol. 15: 284.

27. Gilson, M.K., and B.H. Honig (1987). Nature 330: 84.

28. Glick, R.E., Stewart, W.E., and L. Mandelkern (1966). Biochim. Biophys. Acta. 120: 302.

29. Go, M. and N. Go (1976). Biopolymers 15: 1119.

30. Gruenewald, B., Nicola, C.U., Lustig, A., Schwarz, G. and H. Klump (1979). Biophys. Chem. 9: 137.

31. Harvey, S.C. (1989). Proteins 5: 78.

32. Hall, C.K. and E. Helfand (1982). J. Chem. Phys. 77: 3275.

33. Helfand, E. (1984). Science 226: 647.

34. Helfand, E., Wasserman, Z.R., Weber, T.A., Skolnick, J. and J.H. Runnels (1981). J. Chem. Phys. 75: 4441.

35. Hingerty, B.E., Ritchie, R.H., Ferrell, T.L. and J.E. Turner (1985). Biopolymers 24: 427.

36. Hol, W.G.J., van Duijnen, P.T., and H.J.C. Berendsen (1978). Nature 273: 443.

37. Hol, W.G.J., Halie, L.M., and C. Sander (1981). Nature 294: 532.

38. Hol, W.G.J. (1985). Prog. Biophys. Molec. Biol. 45: 149.

39. Inoue, S., Sano, T., Yakabe, Y., Ushio, H. and T. Yasunaga (1979). Biopolymers 18: 681.

40. Janin, J., and C. Chothia (1980). J. Mol. Biol. 143: 95.

41. Jorgensen, W.L., Chandrasekhar, J., Madura, J. Impey, R.W., and M.L. Klein (1983). J. Chem. Phys. 79: 926.

42. Kabsch, W. and C. Sander (1983). Biopolymers 22: 2577.

43. Karplus, M. and D.L. Weaver (1976). Nature 260: 404.

44. Kim, P.S. and R.L. Baldwin (1982). Ann. Rev. Biochem. 51: 459.

45. Kurland, R.J., and E.B. Wilson (1957). J. Chem. Phys. 27: 585.

46. Levitt, M. and J. Greer (1977). J. Mol. Biol. 114: 181.

47. Levitt, M. and M.F. Perutz (1988). J. Mol. Biol. 201: 751.

48. Levitt, M. and R. Sharon (1988). Proc. Natl. Acad. Sci. USA 85: 7557.

49. Levy, R.M. and M. Karplus (1979). Biopolymers 18: 2465.

50. Levy, R.M., Perahia, D., and M. Karplus (1982). Proc. Natl. Acad. Sci. USA 79: 1346.

51. Manning, M.C. Illangasekare, M. and R.W. Woody (1988). Biophys. Chem. 31: 77.

52. Marqusee, S., and R.L. Baldwin (1987). Proc. Natl. Acad. Sci. USA 84: 8898.

53. Marqusee, S., Robbins, V.H., and R.L. Baldwin (1990). Proc. Natl. Acad. Sci. USA, in press.

54. Matthew, J.B. (1985). Ann. Rev. Biophys. Chem. 14: 387.

55. McCammon, J.A. and S.H. Northrup (1980). Biopolymers 19: 2033.

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

57. Mehler, E.L. and G. Eichele (1984). Biochemistry 23: 3887.

58. Mitchinson, C. and R.L. Baldwin (1986). Proteins 1: 23.

59. Momany, F.A., McGuire, R.F., Burgess, A.W., and H.A. Scheraga (1975). J. Phys. Chem. 79: 2361.

60. Morawetz, H. (1979). Science 203: 405.

61. Norten, A.H. and H.A. Levy (1969). Science 165: 447.

62. Ohgushi, M. and A. Wada (1983). FEBS Lett. 164: 21.

63. Osterhout, J.J., Baldwin, R.L., York, E.J., Stewart, J.M., Dyson, J.H., and P.E. Wright (1989). Biochemistry 23: 7059.

64. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T.M. and R.L. Baldwin (1990). Nature 344: 268.

65. Pear, M.R., Northrup, S.H., McCammon, J.A., Karplus, M. and R.M. Levy (1981). Biopolymers 20: 629.

66. Peller, L. (1959a). J. Phys. Chem. 63: 1194.

67. Peller, L. (1959b). J. Phys. Chem. 63: 1199.

68. Perutz, M.F. (1978). Science 201: 1187.

69. Pethig, R. (1979). Dielectric and electronic properties of biological materials. John Wiley & Sons, Chichester.

70. Peticolas, W. (1978). Methods Enzymol. 61: 425.

71. Ptitsyn, O.B. and A.A. Rashin (1975). Biophys. Chem. 3: 1.

72. Quiocho, F.A., Sack, J.S., and N.K. Vyas (1987). Nature 329: 561.

73. Ramstein, J. and R. Lavery (1988). Proc. Natl. Acad. Sci. USA 85: 7231.

74. Rao, S.N., Singh, U.C., Bash, P.A., and P.A. Kollman (1987). Nature 328: 551.

75. Rees, D.C. (1980). J. Mol. Biol. 141: 323.

76. Richardson, J. (1981). Adv. Prot. Chem. 34: 167.

77. Rogers, N.K., and M.J.E. Sternberg (1984). J. Mol. Biol. 174: 527.

78. Rogers, N.K. (1986). Prog. Biophy. Molec. Biol. 48: 37.

79. Russell, A.J., and A.R. Fersht (1987). Nature 328: 496.

80. Russell, A.J., Thomas, P.G., and A.R. Fersht (1987). J. Mol. Biol. 193: 803.

81. Ryckaert, J., Ciccotti, G. and H.J.C. Berendsen (1977). J. Comp. Phys. 23: 327.

82. Sali, D., Bycroft, and A.R. Fersht (1988). Nature 335: 740.

83. Schaufele, R.F. and T. Shimanouchi (1967). J. Chem. Phys. 47: 3605.

84. Schellman, J. A. (1958). J. Phys. Chem. 62: 1485.

85. Scheraga, H.A. (1973). Pure and Applied Chem. 36: 1.

86. Schimmel, P.R. and P.J. Flory (1967). Proc. Natl. Acad. Sci. USA 58: 52.

87. Schwarz, G. (1965). J. Mol. Biol. 11: 64.

88. Schwarz, G. (1968). Biopolymers 6: 873.

89. Sheridan, R.P., Levy, R.M., and F.R. Salemme (1982). Proc. Natl. Acad. Sci. USA 79: 4545.

90. Shoemaker, K.R., Kim, P.S., Brems, D.N., Marqusee, S., York, E.J., Chaiken, I.M., Stewart, J.M., and R.L. Baldwin (1985). Proc. Natl. Acad. Sci. USA 82: 2349.

91. Shoemaker, K.R., Kim, P.S., York, E.J., Stewart, J.M., and R.L. Baldwin (1987). Nature 326: 563.

92. Singh, U.C., Weiner, P.K., Caldwell, J., and P.A. Kollman (1986). AMBER 3.0, University of California, San Francisco.

93. Singh, U.C., Brown, F.K., Bash, P.A., and P.A. Kollman (1987). J. Am. Chem. Soc. 109: 1607.

94. Skolnick, J. and E. Helfand (1980). J. Chem. Phys. 72: 5489.

95. Skvortsov, A.M., Birshtein, T.M. and A.O. Zalenski (1971). Mol. Biol. 5: 69.

96. Srinivasan, A. and W. Olson (1980). Fed. Am. Soc. Exp. Biol. 39: 2199.

97. Sternberg, M.J.E., Hayes, F.R.F., Russell, A.J., Thomas, P.G., and A.R. Fersht (1987). Nature 330: 86.

98. Sueki, M., Lee, S., Powers, S.P., Denton, J.B., Konishi, Y., and H.A. Scheraga (1984). Macromolecules 17: 148.

99. Suezaki, Y. and N. Go (1976). Biopolymers 15: 1119.

100. Tanokura, M. (1983). Biochim. Biophys. Acta. 742: 576.

101. Tsuji, Y., Yasunaga, T., Sano, T. and H. Ushio (1976). J. Am. Chem. Soc. 98: 813.

102. Van Duijnen, P.Th., Thole, B.Th., and W.G.J. Hol (1979). Biophysical Chemistry 9: 273.

103. Vasquez, M., Pincus, M.R., and H.A. Scheraga (1987). Biopolymers 26: 351.

104. Wada, A. (1976). Adv. Biophys. 9: 1.

105. Warshel, A. (1979). Photochem. Photobiol. 30: 285.

106. Warwicker, J., and H.C. Watson (1982). J. Mol. Biol. 157: 671.

107. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, Jr., S., and P. Weiner (1984). J. Am. Chem. Soc. 106: 765.

108. Wierenga, R.K., De Maeyer, M.C.H., and W.G.J. Hol (1985). Biochemistry 24: 1346.

109. Williams, R.W., Chang, A., Juretic, D. and S. Loughran (1987). Biochim. Biophys. Acta. 916: 200.

110. Wuthrich, K. (1986). NMR of Proteins and Nucleic Acids. John Wiley and Sons, New York.

111. Zana, R. (1975). Biopolymers 14: 2425.

112. Zimm, B.H., and J.K. Bragg (1959). J. Chem. Phys. 31: 526.

CHAPTER 5: Free Energy Component Analysis: A Study of the Glu 165 → Asp 165 Mutation in Triosephosphate Isomerase

The ability to selectively modify individual amino acids in proteins has been of great use in furthering our understanding of the underlying forces governing enzyme action. The best candidate for studying these interactions and how they change when mutations are introduced is an enzyme that is well characterized in mechanistic and energetic terms. Triose phosphate isomerase (TIM) is one such enzyme. Knowles and co-workers have extensively analyzed the TIM catalyzed reversible isomerization of dihydroxyacetone phosphate (DHAP) to glyceraldehyde 3-phosphate (GAP) (reviewed by Knowles and Albery, 1977). We chose to study this system because of the wealth of kinetic data on both native TIM and a variety of mutants.

The catalytic mechanism of TIM is shown schematically in Figure 5.1. After substrate binding, the carboxyl group of Glu 165 abstracts the pro-R proton from the C1 position of DHAP, resulting in an enzyme-bound enediol, or enediolate. A proton is then delivered to the C2 position of the substrate, yielding GAP. An electrophilic residue is thought to polarize the carbonyl group in the enzyme-substrate complex, thereby facilitating proton abstraction (Belasco and Knowles, 1980) (Figure 5.2). It has also been suggested that this electrophilic residue stabilizes the developing negative charges on the oxygens at C1 and C2 during formation of the enediol (or enediolate) by providing the substrate with a positive electrostatic environment. Based on the X-ray structure of Banner *et al.* (1971, 1975, 1976), Lys 13 and His 95 appear to be good candidates. On the basis of mutagenesis experiments, Ser 96 also appears to be catalytically important. Knowles and co-workers have engineered a Ser 96 –> Pro mutant with interesting consequences; proline in this position causes a decrease in the activity of wild-type TIM but increases the activity of the Asp 165 mutant (Hermes et al., 1987).

TIM has been called a "perfect" enzyme (Knowles and Albery, 1977), partly because the rate determining transition state for the reaction is that for product dissocia-

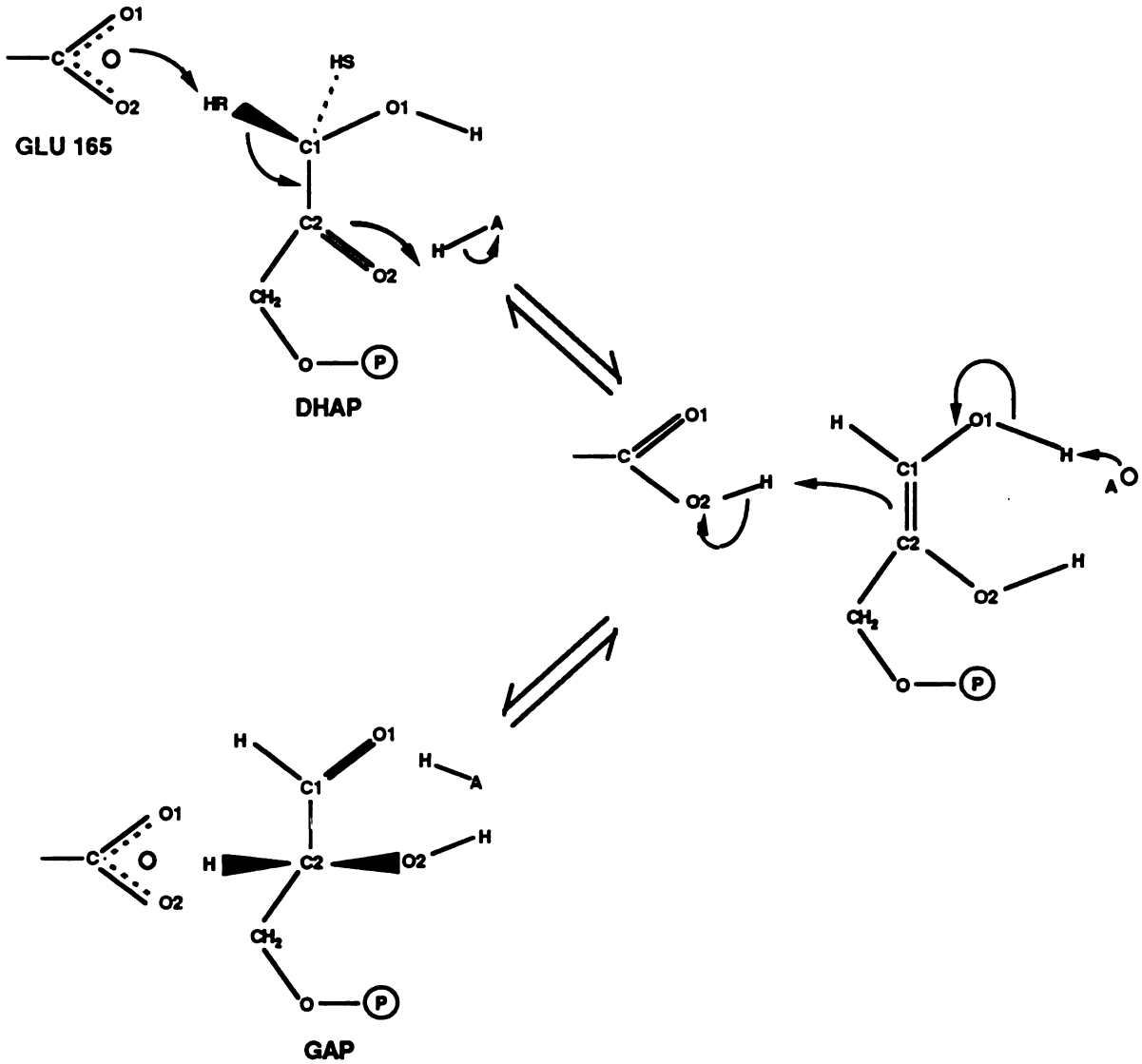


Figure 5.1. Reaction catalyzed by Triose Phosphate Isomerase (TIM). HA is an electrophilic residue(s)---probably lysine 13 and/or histidine 95---that stabilizes the developing negative charge on the substrate carbonyl oxygen (O2) during formation of the enediol/enediolate intermediate.

tion and not a chemical step. When Glu 165 is replaced by Asp, using site-directed mutagenesis techniques, the catalytic activity drops roughly three orders of magnitude (Straus et al, 1985). The Asp 165 mutant employs the same catalytic mechanism as wild-type TIM; however, the process is no longer diffusion-controlled (Blacklow et al., 1988). The rate determining transition state for the Asp mutant is that for formation of the enediol (or enediolate) intermediate, such that the mutant enzyme conforms to the classical Michaelis-Menten scheme. Binding of substrate, enediol/enediolate (as inferred by inhibitor binding), and product are only minimally affected upon mutation (Raines et al., 1986).

Our goal was to examine molecular interactions in the active site that might explain the drop in k_{cat} upon replacing Glu 165 with Asp, as well as to examine substrate binding. To this end, we employed the free energy perturbation method, which uses molecular dynamics to evaluate a statistical mechanically derived formulation of free energy (Singh et al., 1987). The free energy perturbation method has been shown to be very effective in calculating solvation free energies (Bash et al., 1987a), binding free energies (Bash et al., 1987b), and catalytic free energies (Rao et al., 1987) that are in good agreement with relevant experiments.

To date, this method has been applied to systems for which the X-ray data are very accurate, thereby allowing a reasonably complete representation of the aqueous environment around the macromolecule-ligand complex. These applications have involved changing the residue of interest in the noncovalent and covalent complexes and comparing the calculated free energies to those inferred from experimental measurements of the differences between k_{cat} and K_M for the wild-type and mutant structures (Rao et al., 1987).

The available X-ray crystal structures of chicken muscle TIM and its complexes are only known to low resolution [2.5 Å (Banner et al., 1975, 1976) and 6.0 Å (Phillips et al., 1977), respectively]. Thus, our objective was, by necessity, qualitative in nature. Com-

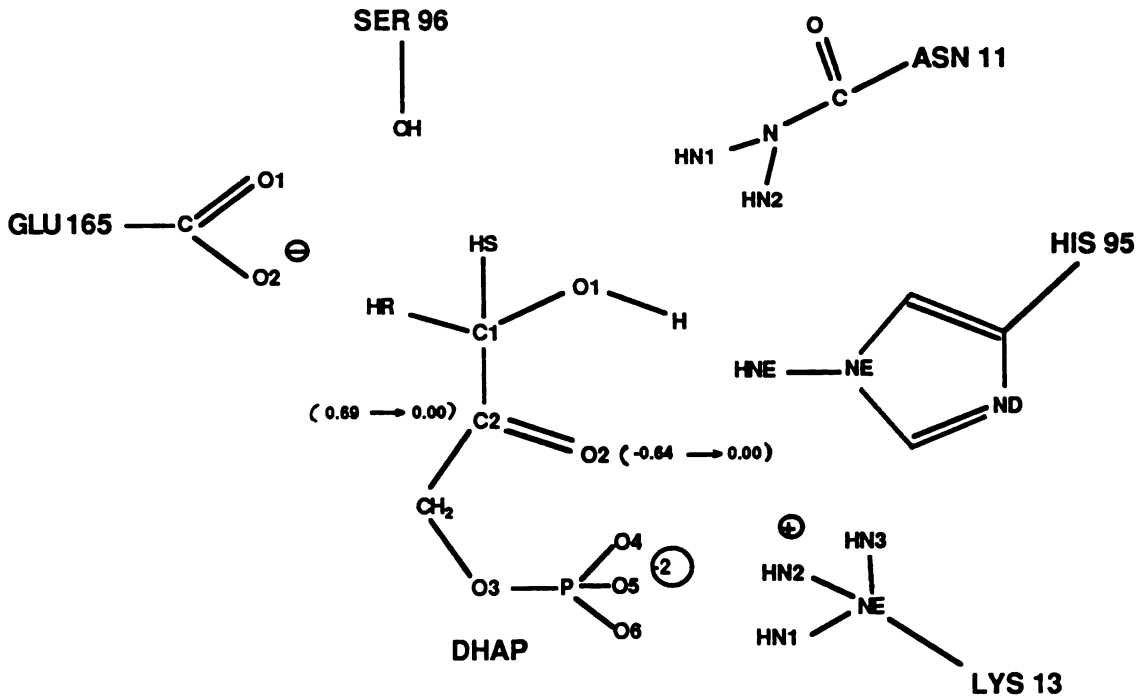


Figure 5.2. Schematic Representation of Active Site of TIM. Substrate (DHAP) and key active site residues of TIM. The O2 dipole perturbation is depicted. Numbers in parentheses represent initial, unperturbed charges and final, perturbed charges.

pared to earlier studies, we took a slightly different approach, which we refer to as a free energy component analysis. We calculated the free energies for perturbing the charge distributions of various residues in the active site (Asn 11, Lys 13, His 95, Ser 96, Glu 97) and portions of the substrate (O1 and O2 dipoles) in the covalent and noncovalent complexes of both the wild-type and mutant enzymes. Asn 11 and Glu 97 were selected for study in addition to the key catalytic residues (Lys 13, His 95, Ser 96) discussed above because of their proximity to the substrate. Cys 125 is another potential hydrogen bonding group in the active site, but it was not considered because it is over 7 Å from the catalytically important substrate atoms. (See Figure 5.2 for a schematic representation of the relative positions of these residues in the active site.) By taking this approach of systematically changing electrostatic interactions in the active site, we hoped to avoid the problems noted above and also arrive at a more detailed interpretation of the roles of those residues that are important in substrate binding and catalysis. This approach is qualitative rather than quantitative in nature but can point out the importance of particular interactions that can be tested experimentally.

By analyzing the free energies for perturbing various groups in the wild-type TIM and DHAP noncovalent complex and the mutant noncovalent complex (Glu 165 replaced by Asp), we examined the interactions important for substrate binding in the two structures. A comparison of the covalent wild-type substrate complex (carboxylate of residue 165 linked to pro-R proton of DHAP) and the covalent mutant-substrate complex (which are models of the transition structure for enolization), allowed us to evaluate qualitatively the interactions that stabilize the transition state, and in turn aid in interpreting the drop in k_{cat} upon mutation. k_{cat} can be further broken down to yield k_{enol}, as the rate constants for individual steps in the conversion of DHAP to GAP have been determined by Knowles and co-workers for the Asp165 mutant TIM (k_{enol}=2.0 sec^{-1}) (Raines et al., 1986) and wild-type TIM (k_{enol}=2.0 x 10^3 sec^{-1}) (Knowles and Albery, 1977). We compared our results to changes in k_{cat}, but one reaches the same conclusions if the rate con-

stants for enolization are considered.

METHODS

Calculation of Free Energy Changes

Except where noted below, calculations were performed using AMBER version 3.0 (Singh et al., 1986a). We calculated Gibbs (G) free energy changes using equation 1, where ΔH is the difference in the Hamiltonian between two states, ΔG is the free energy difference between these states, R is the gas constant, T is the absolute temperature, and the symbol $< >_{ref}$ indicates that an ensemble average is taken with respect to some reference state. (See Singh et al., 1987 for a more thorough discussion of the method.)

$$\Delta G = - RT \ln < \exp (-\Delta H / RT) >_{ref} \qquad \text{Equation 1}$$

In the cases discussed here, H represented the interaction energy of the perturbed group with its surroundings. These interaction energies were calculated at intermediate points along the conversion pathway using an empirical force field. Molecular dynamics at 300 K was used to generate the ensemble of structures.

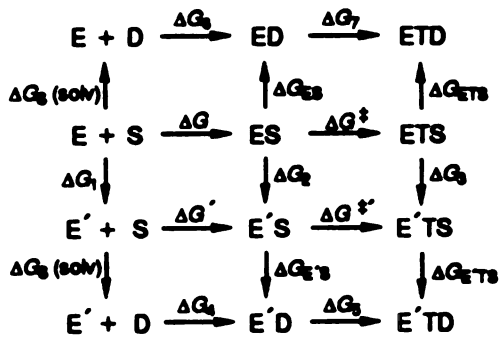
In the applications of this approach published to date, we have reported free energies due only to the inter-group interactions (Singh et al. , 1987; Bash et al., 1987a, 1987b; Rao et al., 1987). There are many cases where it is advantageous not to include intra-group effects. For example, when one mutates the charges on the oxygen of R-CO-NH-R to zero, there is a very large energy associated with a change in O...H nonbonded interactions. To separately calculate any inter-group effect of the oxygen interacting with its environment, one would need to carry out the mutation of the oxygen for the fragment by itself (which includes only intra-group effects) and in the presence of its environment (which includes both intra- and inter-group effects). This separation is easy to do operationally but involves finding a small difference between large numbers. Instead, one can define CO-NH as the perturbing group, even though the properties of the N-H group do not change, and only consider the inter-group interactions (the intermolecular interac-

tions of this group with its environment). So, when we included particular atoms as part of the system that changes while holding their molecular mechanical parameters constant, the free energy change for the "perturbation" of these atoms was zero. This approach was taken to examine interactions between Lys 13 and specific groups in the active site.

General Formalism for Component Analysis

Perturbation of Substrate Atoms

Consider the following pathways for binding and catalysis of ligands by an enzyme E and a site specific mutant E', where S is the substrate and D denotes a dummy substrate.



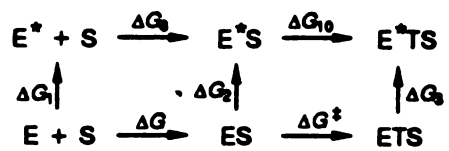
ES corresponds to the enzyme-substrate noncovalent complex and ETS to the transition state for the enzyme catalyzed reaction. We are interested in the difference in binding free energy, $\Delta\Delta G_{bind}=\Delta G' - \Delta G$ and catalytic free energy $\Delta\Delta G_{cat}=\Delta G^{+\prime} - \Delta G^{+}$ between the enzyme and its site-specific mutant. As noted by Rao *et al.* (1987) in their studies of subtilisin mutants, it is usually easier to calculate ΔG_1, ΔG_2, and ΔG_3 rather than ΔG, $\Delta G'$, ΔG^{+}, $\Delta G^{+\prime}$ and to use the fact that free energy is a state function to determine $\Delta\Delta G_{bind}= \Delta G_2 - \Delta G_1$ and $\Delta\Delta G_{cat}= \Delta G_3 - \Delta G_2$. However, in some cases, such as here for TIM, the direct determination of ΔG_1, ΔG_2, and ΔG_3 is more difficult and, instead, an indirect path can be used to estimate contributions to $\Delta\Delta G_{bind}$ and $\Delta\Delta G_{cat}$. This involves mutating the substrate S or transition state TS into dummy atoms D or TD. By using the above

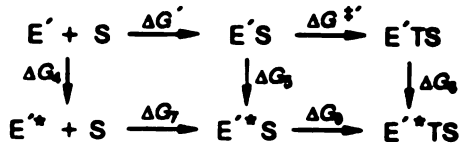
thermodynamic cycle, we see that $\Delta\Delta G_{bind}$= ($\Delta G_{E'S}$- ΔG_S(solv)+ΔG_4) - (ΔG_{ES} - ΔG_S(solv)+ΔG_6) collapses to $\Delta\Delta G_{bind}$= $\Delta G_{E'S}$- ΔG_{ES}, assuming that the wild-type and mutant enzyme have the same affinity for dummy substrate (e.g. $\Delta G_4 = \Delta G_6$). Similarly, $\Delta\Delta G_{cat}$= $\Delta G_{E'TS}$- ΔG_{ETS} - ($\Delta G_{E'S}$ - ΔG_{ES}). Thus, the determination of $\Delta\Delta G_{bind}$ and $\Delta\Delta G_{cat}$ involves the mutation of the substrate atoms into dummy atoms in both the noncovalent complex and in a model transition state structure. We also can mutate selected substrate atoms into dummy atoms to estimate the contributions of groups of substrate atoms to binding or catalysis.

In the computational implementation of this approach, we mutated only the electrostatic partial charges of the substrate S into those of the dummy molecule, D, to determine particular electrostatic contributions to $\Delta G_{E'S}$, ΔG_{ES}, $\Delta G_{E'TS}$, and ΔG_{ETS}. Therefore, we assumed that the van der Waals contribution to the free energies was equal for ΔG_{ES} and $\Delta G_{E'S}$ and for ΔG_{ETS} and $\Delta G_{E'TS}$. This appears to be a reasonable approximation, because the van der Waals contributions involve "disappearing" the same atoms in each of these cases. Computationally, though, the determination of van der Waals changes involves much more extensive sampling than electrostatic changes and given our simple model of the enzyme active site (no water inclusion) such a determination would likely be inaccurate and involve large statistical errors. Thus, such an approximation makes sense in this case.

Formalism for Perturbing Enzyme Atoms in Native vs. Mutant Enzymes

We also seek a formalism for estimating the contributions of various protein atoms to differential binding and catalysis of an enzyme and its site-specific mutant. We begin with the following thermodynamic cycle:





The nomenclature for the scheme above is essentially the same as for the substrate perturbations with the addition of E*, E*S and E*TS, which correspond to the native enzyme, its complex with substrate and transition state structure with particular atoms changed to dummy atoms, and E'*, E'*S and E'*TS correspond to the mutant enzyme with the same sets of changes. Here our goal is to calculate the following:

$\Delta\Delta G^*_{bind}$ = contribution to binding from atoms changed to dummies in mutant vs. native

$$=\Delta G' - \Delta G + \Delta G_8 - \Delta G_7$$

$\Delta\Delta G^*_{cat}$ = contribution to catalysis from atoms changed to dummies in mutant vs. native

$$=\Delta G^{\ddagger\prime} - \Delta G^{\ddagger} + \Delta G_{10} - \Delta G_9$$

and to estimate the experimental $\Delta\Delta G_{bind}$ and $\Delta\Delta G_{cat}$ by the sum of the $\Delta\Delta G^*_{bind}$ and $\Delta\Delta G^*_{cat}$, respectively, for all of the functionally important atoms in the molecule. Interactions between perturbed residues are overcounted when summing the contributions to $\Delta\Delta G_{bind}$ and $\Delta\Delta G_{cat}$. This is remedied, though, by subtracting the contribution to the free energy of each combination of perturbed groups from the total.

To proceed further, we made the assumption that we can formally break down the ΔG values into two components: (a) the change in intragroup energies for changed atoms, and (b) the interaction of changed atoms with the rest of the enzyme, water, and S or TS. Therefore, $\Delta G_i = \Delta G_{ia} + \Delta G_{ib}$ for any particular atom that we change. We also make the assumption that the change in intragroup energies are equal in the different environments.

$$\Delta G_{4a} = \Delta G_{5a} = \Delta G_{6a}$$

$$\Delta G_{1a} = \Delta G_{2a} = \Delta G_{3a}$$

Finally, we are left with

$$\Delta\Delta G^*_{bind} = \Delta G_{5b} - \Delta G_{2b} - \Delta G_{4b} + \Delta G_{1b}$$

$$\Delta\Delta G^*_{cat} = \Delta G_{2b} - \Delta G_{3b} - \Delta G_{5b} + \Delta G_{6b}$$

In the simulations reported here, we assumed that $\Delta G_{1b}=\Delta G_{4b}$. This is because we could not calculate a realistic estimate of the difference in interactions of the perturbed atoms with the rest of the protein between the native and mutant structures due to large scale side chain movements because of the lack of substrate, both with and without solvent present. Even so, this is probably a reasonable assumption given that the environment around a particular residue is very similar in the two structures.

Computational Details

We employed the windowing method of perturbation to calculate free energies, which involves breaking up the perturbation into discrete steps (windows), as described by Singh *et al.* (1987). Each progression from the unperturbed structure to the perturbed structure, for the various simulations described below, was carried out using 5-21 windows. 200 equilibration steps and 400 steps of data collection were performed at each window, with a step size of 1 femtosecond. Thus, the total time course for each perturbation was between 3 and 12.6 picoseconds (psec); different simulation times were used to ensure that the calculated free energies were independent of the length of time for the conversion. We were limited to fairly short simulation times as the structures drifted quite a bit from their starting configurations with long simulations, particularly when we changed charges critical in maintaining side chain or substrate orientations. The length of time for each simulation used here represents a compromise between minimizing hysteresis and attaining sufficient sampling. Only those residues within 10 Å of residue 165 were allowed to move. SHAKE (Ryckaert et al., 1977) was used for all bonds, and all free energies were calculated at 300 K. A distance dependent dielectric constant ($\varepsilon = r_{i,j}$,

where r is the intercharge separation between atoms i and j) and a 10 Å nonbonded cutoff were used for the calculations.

The reported free energy changes represent the average of at least two independent simulations. The uncertainties quoted with the average free energies are not true uncertainties but actually reflect the hysteresis within a particular run and between different runs. For each simulation from the unperturbed to the perturbed charge we calculated two free energy changes, one for forward-looking sampling at each window and one for backward-looking sampling. The uncertainties reported here are the largest difference between either forward and backward sampling free energy changes or calculated free energy changes from different runs. Currently there is no way to rigorously determine the uncertainties for free energies calculated with this method.

Generation of Structures

We performed calculations on four TIM models with DHAP: (1) wild type TIM (Glu 165) with noncovalently bound DHAP; (2) mutant TIM (Asp 165) with noncovalently bound DHAP; (3) wild type TIM with covalently bound DHAP; and (4) mutant TIM with covalently bound DHAP. The model for the wild-type noncovalent enzyme-substrate complex was the final structure, after 10.5 psec of molecular dynamics at 300 K, reported in an earlier study (Brown and Kollman, 1987). [This earlier study used the refined crystal coordinates of chicken muscle triose phosphate isomerase as the starting structure for molecular dynamics (Banner et al., 1975, 1976).] The mutant structure was obtained by replacing Glu 165 with Asp, maintaining the original wild-type side chain orientation. Another orientation of residue 165 was also used to test the dependence of the calculated free energy changes on the structure. This orientation was generated during a trial simulation aimed at perturbing Glu into Asp directly. Standard united-atom parameters (hydrogens on carbon atoms are incorporated into the van der Waals radius of the carbon) were used for the TIM dimer (Weiner et al., 1984). All structures contained the appropriate counterions on charged surface residues; however, explicit solvent

molecules were not present. Structures were further equilibrated, to different extents but up to 7.5 psec, prior to perturbation calculations to ensure that the calculated free energies were not excessively dependent on starting structure. The covalent complexes were constructed, from pre-equilibrated noncovalent complexes, by imposing a covalent bond between the pro-R hydrogen of the substrate (H1R) and a carboxyl oxygen of residue 165 (O2 of Glu or Asp). (See Figure 5.3.) The resulting structures were then equilibrated for 1-5 psec at 300 K.

The partial charges for DHAP were determined in a single-point 4-31G* *ab initio* calculation (Brown and Kollman, 1987) with the refined crystal coordinates of Banner and co-workers (1975, 1976), using the UCSF-G80 electrostatic potential fitting routine (Singh and Kollman, 1984). The substrate charges for the covalent enzyme-substrate complex were based on the gas phase transition structure constructed by Alagona *et al.* (1984), with the addition of a phosphate group. Two nonstandard AMBER atom types were assigned for DHAP covalently bound to TIM. The parameters for the new atom types, TC for C1 and TH for the pro-R hydrogen (Figure 5.3), are given in Table 5.1; all other substrate atoms were assigned standard all-atom parameters (Weiner et al., 1986).

Description of Protein Residue Perturbations

Potential hydrogen bonding sites of specific amino acid residues in the active site and portions of the substrate were removed by zeroing the charge on the hydrogen and/or oxygen atoms. The perturbed groups fall into three categories: neutral, polar protein residues; charged protein residues; and the substrate.

Neutral, Polar Protein Residues

In the case of the neutral, polar residues, overall charge neutrality was maintained in going from the perturbed to the unperturbed state. To maintain charge neutrality is was necessary to change the charges of some atoms covalently connected to the potential hydrogen bonding atoms. Three residues fall into this group of perturbed residues: Asn 11, His 95, and Ser 96. The pertinent original and all perturbed charges for these residues

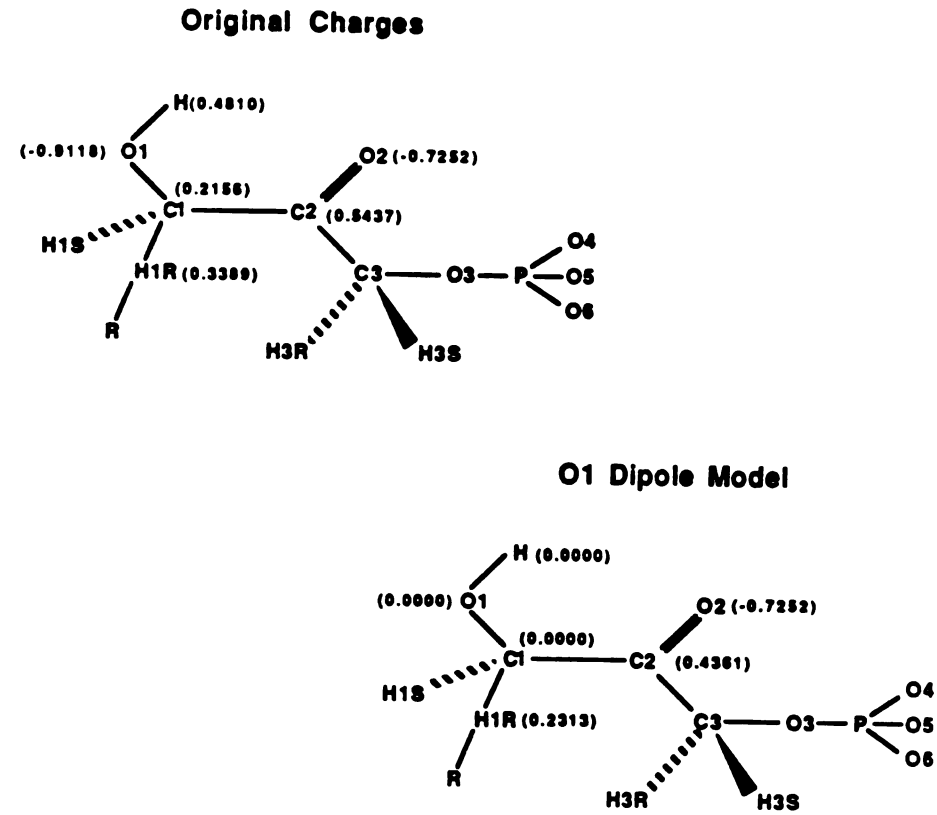


Figure 5.3. Models for Perturbing Charges and Hydrogen Bond Dipoles in Substrate (DHAP). The structures shown are for DHAP covalently bond to TIM through O2 of Glu or Asp (here represented as R). Partial charges are given in parentheses. The perturbed charges are explicitely shown here; all other charges are given in Table 5.2.

TABLE 5.1
Parameters for Atom Types TC and TH[a]

Bond Parameters		
Bond	K_r	r_{eq}
TC-TH	331.0	1.10
TC-C	317.0	1.43
TH-O2	331.0	1.60

Angle Parameters		
Angle	K_θ	θ_{eq}
TH-TC-OH	135.0	109.5
O2-TC-C	0.0	109.5
C	0.0	1.0
C	0.0	109.5
O	0.0	109.5
T	0.0	120.0
TI	135.0	109.5
O2	0.0	179.9
TC	70.0	118.6
TH-	135.0	109.5
OH-TC-C	63.0	109.5
TC-C-O	80.0	120.4
TC-OH-HO	55.0	108.5
C-O2-TH	70.0	120.0

Torsion Parameters			
Torsion	$V_n/2$	γ	n
X-TC-TH-O2	0.0	0.0	2.0
X-O2-TC-X	0.0	0.0	3.0
OH-TC-C-X	20.0	180.0	2.0

Nonbonded Parameters		
Atom	R^*	ε^*
TH	0.10	0.000

[a]Nonstandard atom type parameters that are the same as the standard parameters are not listed here (e.g. when TC and CT, and TH and HC have the same parameters).See Weiner et al. (1986) for standard parameters and Weiner et al. (1984) for explanation of terms.

are given in Figure 5.4.

Charged Protein Residues

The perturbation of the charged residues involved neutralizing a full charge and the changes were localized to the actual atoms of interest. Two fully charged residues in the active site were neutralized---Lys 13 and Glu 97 (Figure 5.4). The perturbation of Glu 97 was straightforward and involved only changing the charges on the oxygens. The determination of interactions between lysine 13 and its environment, on the other hand, was complicated by the strong interactions between the ε-amino group and the phosphate group of DHAP. To separate the interactions between Lys 13 and non-phosphate portions of the substrate, the phosphate group, and other active site residues, we included various atoms in the perturbing group, which remain unchanged during the simulation. Only results for the perturbation of Lys 13 in the covalent complex are reported. The structural consequences of changing the Lys charges were too drastic in the noncovalent structure to provide meaningful results.

Four separate perturbation calculations were performed to explore interactions between Lys 13 and its environment: (1) a 30% decrease in the partial charges of the N_ε hydrogens; (2) a 30% decrease in the partial charges of the hydrogens and the entire substrate is defined as part of the perturbing group (without any of its parameters changing); (3) a 30% reduction in the charge of the hydrogens and the phosphate group is part of the perturbing group; and (4) the partial charges on the hydrogens are zeroed and the phosphate group is part of the perturbing group. The first perturbation is shown schematically in Figure 5.4. The decrease in hydrogen charge of 30% is arbitrary. Our concern was to calculate free energies of reasonable magnitude (e.g. not too large) so that the uncertainties were not too large.

Substrate Perturbations

In addition to perturbing active site residues, portions of the substrate were perturbed to determine their interactions with the enzyme active site. Figure 5.3 shows the

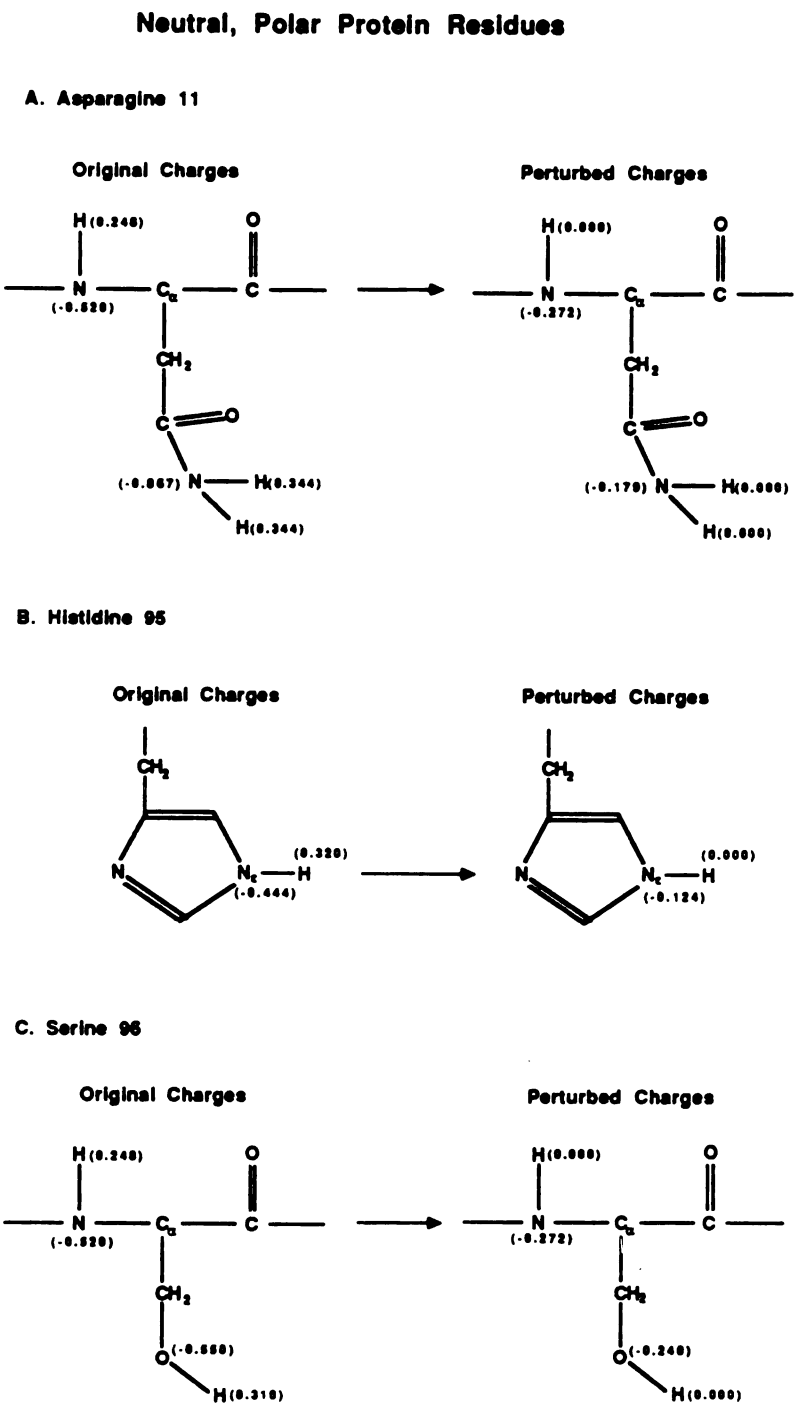
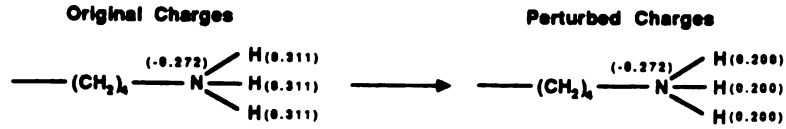


Figure 5.4. Models for Perturbing Active Site Hydrogen Bond Dipoles and charges of TIM. Partial charges are given in parentheses. All perturbed charges and the pertinent original charges are depicted.

Charged Protein Residues

D. Lysine 13



E. Glutamate 97

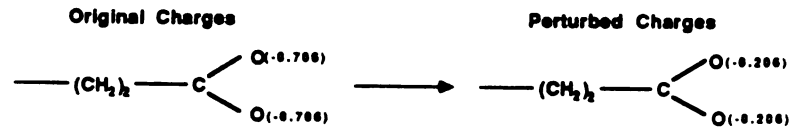


TABLE 5.2
Original and Perturbed Partial Charges for DHAP Dipoles[a] in Noncovalent and Covalent Complexes with TIM

Noncovalent					
Atom Number[a]	Atom Type[b]	Original Charge	Perturbed Charge		
			O1 Dipole	O2 Dipole	O1/O2
H	HO	0.4452	0.0000	0.4452	0.0000
O1	OH	-0.7955	0.0000	-0.7955	0.0000
C1	CT	0.2966	0.0000	0.2966	0.0000
H1R	HC	-0.0150	-0.0150	-0.0150	-0.0150
H1S	HC	-0.0125	-0.0125	-0.0125	-0.0125
C2	C	0.6879	0.6879	0.0000	0.0000
O2	O	-0.6367	-0.6367	0.0000	0.0000
C3	CT	0.1714	0.1714	0.1714	0.1714
H3R	HC	0.0119	0.0119	0.0119	0.0119
H3S	HC	0.0121	0.0121	0.0121	0.0121
O3	OS	-0.5971	-0.5971	-0.5971	-0.5971
P	P	1.4452	1.4452	1.4452	1.4452
O4	O2	-1.0045	-1.0045	-1.0045	-1.0045
O5	OH	-1.0045	-1.0045	-1.0045	-1.0045
O6	OH	-1.0045	-1.0045	-1.0045	-1.0045
Total Charge		-1.9999	-1.9463	-2.0512	-1.9975
Covalent					
Atom Number[a]	Atom Type[b]	Original Charge	Perturbed Charge		
			O1 Dipole	O2 Dipole	O1/O2
H	HO	0.4810	0.0000	0.4810	0.0000
O1	OH	-0.9118	0.0000	-0.9118	0.0000
C1	TC	0.2156	0.0000	0.0341	0.0000
H1R	TH	0.3389	0.2313	0.3389	0.0041
H1S	HC	0.0041	0.0041	0.0041	0.0041
C2	C	0.5437	0.4361	0.0000	0.0000
O2	O	-0.7252	-0.7252	0.0000	0.0000
C3	CT	0.2764	0.2764	0.2764	0.2145
H3R	HC	-0.0286	-0.0286	-0.0286	-0.0286
H3S	HC	-0.0286	-0.0286	-0.0286	-0.0286
O3	OS	-0.5971	-0.5971	-0.5971	-0.5971
P	P	1.4452	1.4452	1.4452	1.4452
O4	O2	-1.0045	-1.0045	-1.0045	-1.0045
O5	OH	-1.0045	-1.0045	-1.0045	-1.0045
O6	OH	-1.0045	-1.0045	-1.0045	-1.0045
Total Charge		-1.9999	-1.9999	-1.9999	-1.9999

[a] See Figure 5.3.

[b] See Brown and Kollman (1987).

original and perturbed charges for DHAP in the covalent complex. (See Table 5.2 for the complete set of charges.) The O1 dipole model involved perturbing the dipole resulting from the hydroxyl group at C1. The charges of other atoms in the substrate were also altered to maintain the original overall charge. The O2 dipole model involved zeroing the charges of the carbonyl group. (See Figure 5.2 for an illustration of the O2 dipole change.) The combined O1/O2 dipole model entailed the simultaneous perturbation of both dipoles. The various charges for perturbation of the substrate in the noncovalent complexes are shown in Table 5.2 and are analogous to those described above.

RESULTS

Neutral, Polar Protein Residues

Asparagine 11

The free energy changes for zeroing the backbone N-H and side chain N-H dipoles of Asn 11 (Figure 5.4a) are given in Table 5.3. Asn 11 interacts strongly with its environment in all four structures. The wild-type noncovalent complex is stabilized by 2 kcal/mole over the mutant complex (12.1 vs. 10.1 kcal/mole, Table 5.3). This difference appears to be due to less repulsive interactions between one of the Asn hydrogens and a side chain hydrogen of Gln 63 in the wild-type complex than in the mutant complex. The distance between the two hydrogens is 3.8 Å in the wild-type structure and 2.33 Å in the mutant; this difference of 0.85 Å could easily account for a 2 kcal/mole difference in the calculated ΔG. The interactions between the substrate atoms and Asn 11 appear to be very similar in the two structures; as can be seen in Figure 5.5, the orientations of Asn 11 in the wild-type enzyme and in the mutant relative to other active site residues are essentially the same. For example, the distance between a side chain amide hydrogen and O1 of the substrate in the noncovalent wild-type and mutant complexes are 1.85 and 1.82 Å, respectively (Table 5.6). The free energy changes for perturbing Asn 11 in the covalent structures are essentially the same, 14.9 vs. 14.2 kcal/mole (Table 5.3), although Asn 11 makes an hydrogen bond with O1 of the substrate in the wild-type complex but shifts to

TABLE 5.3
Free Energy Changes Upon Perturbing Charges and Hydrogen Bond Dipoles in Protein (kcal/mole)

Asparagine 11: NH and NH_2 (Figure 5.4A)		
Protein Model	Substrate Model	ΔG^a
Glu 165	noncovalent	12.1 ± 0.1
Asp 165	noncovalent	10.1 ± 0.2
Glu 165	covalent	14.9 ± 1.2
Asp 165	covalent	14.2 ± 1.3
Histidine 95: N_ε-H (Figure 5.4B)		
Protein Model	Substrate Model	ΔG^a
Glu 165	noncovalent	4.4 ± 0.2
Asp 165	noncovalent	4.8 ± 0.4
Glu 165	covalent	4.2 ± 0.4
Asp 165	covalent	3.5 ± 1.7
Serine 96: NH and OH (Figure 5.4C)		
Protein Model	Substrate Model	ΔG^a
Glu 165	noncovalent	13.2 ± 0.7
Asp 165	noncovalent	11.0 ± 0.4
Glu 165	covalent	12.3 ± 1.0
Asp 165	covalent	12.9 ± 1.6
Glutamate 97: O1 and O2 (Figure 5.4E)		
Protein Model	Substrate Model	ΔG^a
Glu 165	noncovalent	23.7 ± 0.5
Asp 165	noncovalent	22.4 ± 0.3
Glu 165	covalent	19.7 ± 0.6
Asp 165	covalent	23.6 ± 4.0

[a] The free energy changes reported in this table conform to the formalism in the methods section by noting the following: changes in the wild-type enzyme in the noncovalent complex represent ΔG_2; perturbation of the Asp mutant in the noncovalent complex correspond to ΔG_5; free energies for the wild-type covalent complex represent ΔG_3; and changes in the mutant covalent complex correspond to ΔG_6.

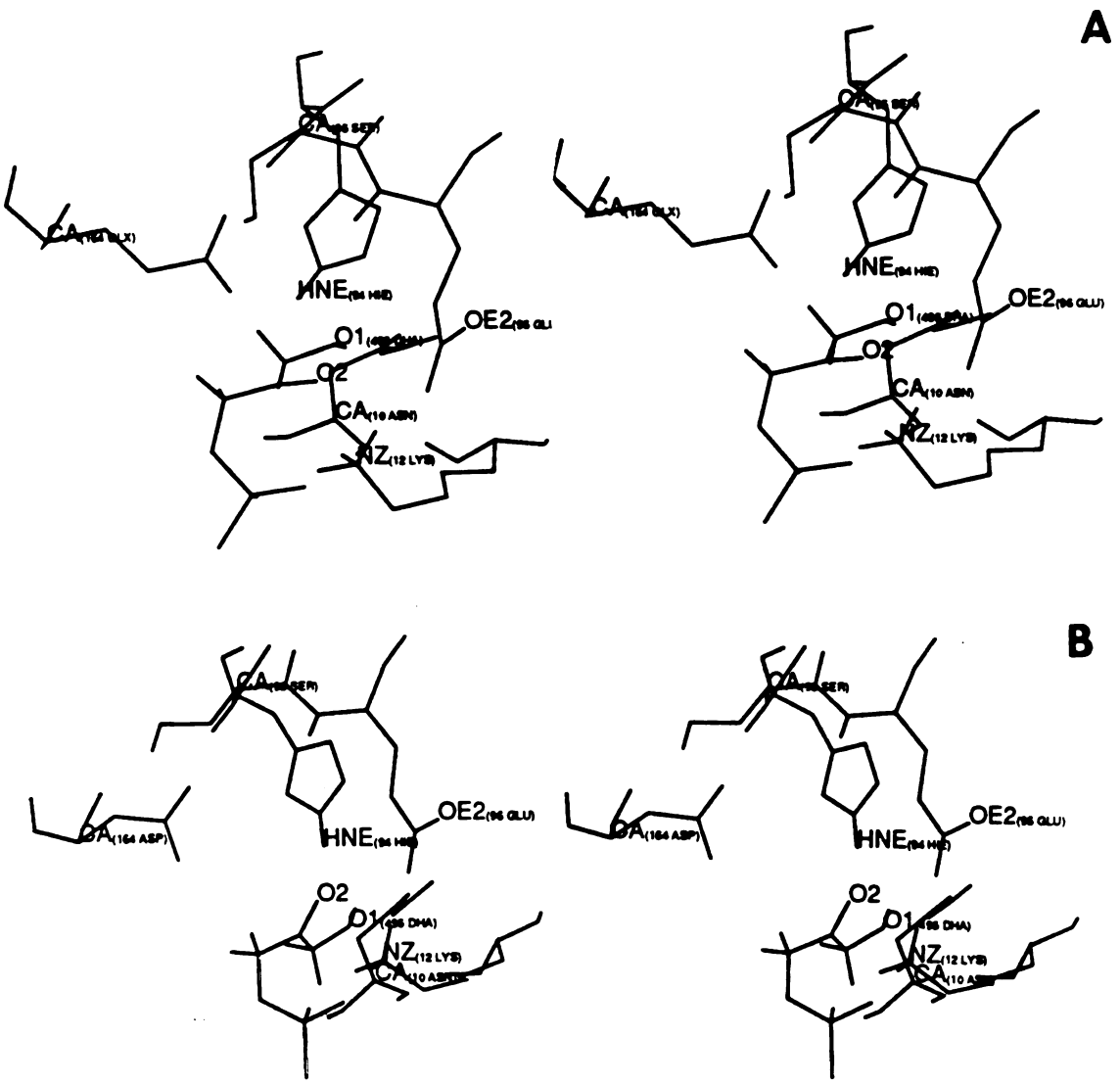


Figure 5.5. Stereoviews of the Active Site Region of Noncovalent DHAP-TIM Complexes. A. Wild-type TIM (Glu 165) and DHAP. The structure is the final structure after 15 psec of molecular dynamics at 300K. The numbering and labels are slightly different than in the text. Residue 164 (GLX) is Glu 165. Residue 495, DHA, is the substrate DHAP. The numbering of the residues differs from the numbering in the text by one: His 95 is Hie 94, Asn 11 is Asn 10, Ser 96 is Ser 95, Lys 13 is Lys 12 and Glu 97 is Glu 96 in this figure. B. Mutant TIM (Asp 165) and DHAP. The structure is a molecular dynamics (at 300 K) snapshot after 15 psec. Residue 164 in this figure is Asp 165. The other residues are the same as described above.

O2 in the mutant complex.

Histidine 95

The free energies for perturbing the hydrogen bond donor site of histidine 95 are unfavourable, which is consistent with the suggestion that it interacts favourably with the substrate by providing a positive electrostatic environment. However, our results indicate that the hydrogen bond interacts in a similar manner in all four of the structures (all of the free energies are approximately 4 kcal/mole, Table 5.3). In the absence of discrimination between wild-type and mutant TIM, the drop in activity cannot be explained in terms of differential histidine interactions. As mentioned above, these free energies represent interactions between His 95 and its environment. So, differences between the free energies for perturbation of wild-type and mutant structures can be due to any number of interactions. In some cases perturbation of different structures can yield different energetic contributions while maintaining the same overall free energies. It appears as though this might be the case with the His 95 perturbations. For example, the HN_{ε}-O2 (of the substrate) distances in the Glu structures (noncovalent 2.21 Å, covalent 2.43 Å) are longer than in the Asp structures (noncovalent 1.70 Å, covalent 1.76 Å). (See Table 5.4.) This might suggest that His 95 of the Asp 165 structure would better stabilize the transition state, but the geometry of the hydrogen bond is less than optimal. (Compare structures A and B of Figure 5.6.) Also, HN_{ε} of His 95 interacts more strongly with O1 in the wild-type structure than in the mutant structure. Even though it appears that His 95 is in a better position to stabilize O2 of the substrate in the mutant transition structure (in terms of distances between HN_{ε} and O2 of the substrate), there are compensating interactions in the wild-type structure that result in similar overall free energies for perturbation of the two covalent structures. It may be important to maintain His 95 poised between O1 and O2, which the wild-type does very effectively and the mutant does not (Figure 5.6 and Table 5.6).

Serine 96

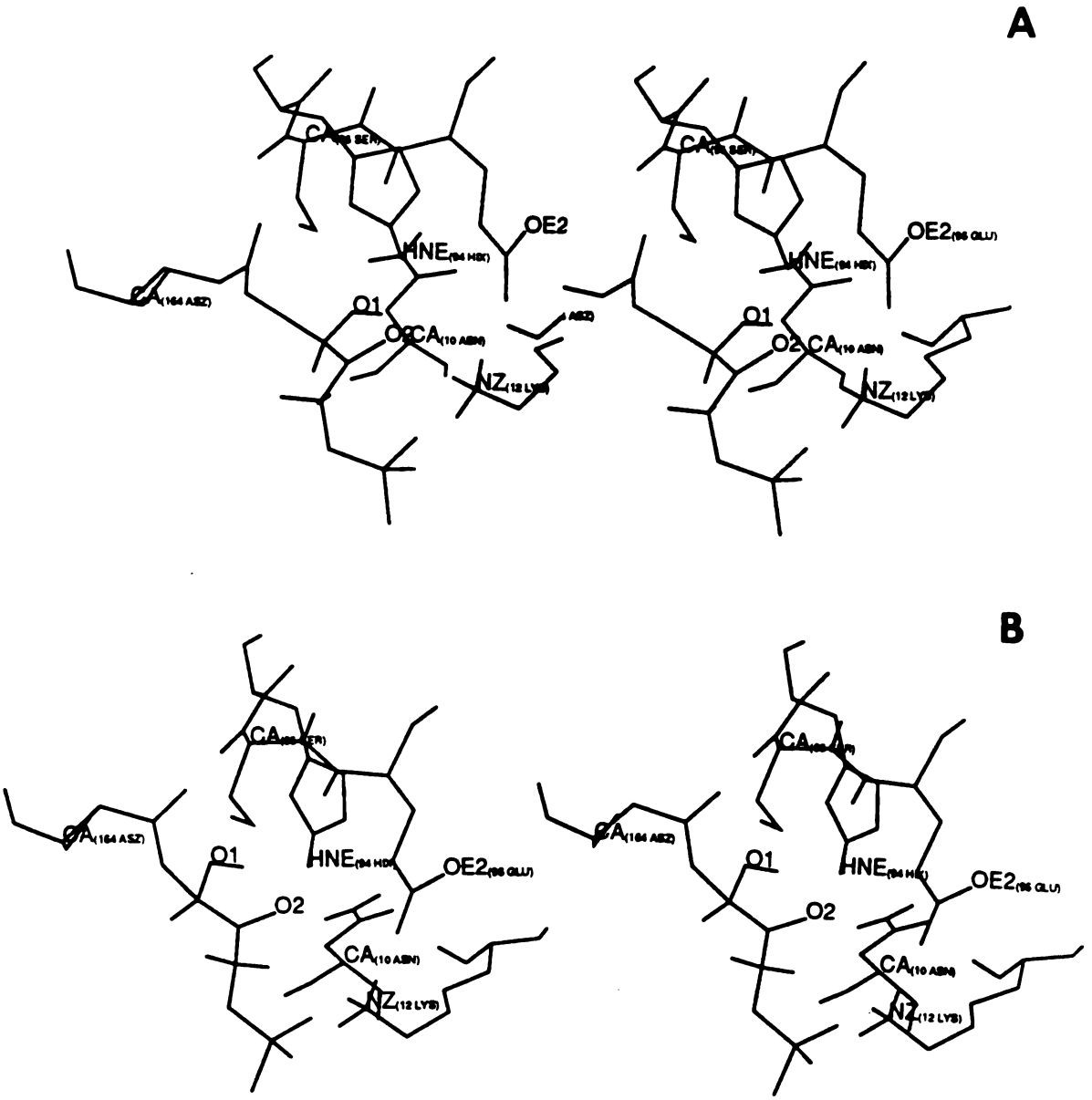


Figure 5.6. Stereoviews of Active Site Regions of Covalent DHAP-TIM Complexes. A. Wild-type TIM (Glu 165) and DHAP. The structure is a molecular dynamics snapshot after 15 psec at 300 K. The labeling is essentially the same as in Figure 5 except for the following differences: residue 164, ASZ, is Glu 165 covalently linked through O2 to H1R of the substrate. B. Mutant TIM (Asp 165) and DHAP. The structure is a molecular dynamics snapshot after 15 psec at 300 K. The labeling is the same as described above. Residue 164 (ASX), however, represents Asp 165 covalently linked to DHAP.

Serine 96 interacts strongly with its environment in all four structures. The free energy for this interaction in the wild-type noncovalent complex is 2.2 kcal/mole more favourable than in the mutant complex (13.2 vs. 11.0 kcal/mole, Table 5.3). The environment around Ser 96 is very similar in the two noncovalent structures with the exception that Ser 96 is in a more favourable position to form an hydrogen bond with O1 of Glu 165 in the wild-type structure. The distance between the hydroxyl proton of Ser 96 and O1 of residue 165 is 0.8 Å longer for mutant TIM than for wild-type TIM (Table 5.4); there is less than a 0.2 Å difference between the Ser HN and O1 distances. It is important to note that the calculated free energies are a measure of the interactions between the perturbed residue and the rest of the protein as well as the substrate. Therefore, although our intention is to interpret the free energies in terms of probable molecular interactions, we need structural confirmation that the proposed hydrogen bonds exist.

Charged Protein Residues

Lysine 13

The results for the various lysine 13 perturbations are presented in Table 5.4. The first perturbation represents the interaction of Lys 13 with its environment, including the entire substrate. Lys 13 stabilizes the Asp 165 transition structure by almost 3 kcal/mole over the wild-type structure (31.1 vs. 28.4 kcal/mole). The free energies for the second perturbation (H charge 0.311 → 0.200 and substrate → substrate) are a measure of interactions between Lys 13 and other TIM residues but do not include interactions with the substrate. In this case, the wild-type covalent complex is stabilized by approximately 1 kcal/mole, although both values are fairly low. The third perturbation (H charge 0.311 → 0.200 and phosphate → phosphate) represents interactions between Lys 13 and other TIM residues in the active site and all of the substrate atoms except the phosphate group. This interaction stabilizes the wild-type covalent complex over the mutant structure by 6 kcal/mole (11.0 vs. 4.9 kcal/mole).

TABLE 5.4
Free Energy Changes Upon Perturbing Hydrogen Bond Dipoles of Lysine 13 (N_ε-H) and Portions of Substrate (kcal/mole)

Perturbation	Substrate Model	Protein Model	ΔG^b
(1) H charge 0.311→0.2[a]	covalent	Glu 165	28.4 ± 0.1
		Asp 165	31.1 ± 0.2
(2) H charge 0.311→0.2 substrate→substrate	covalent	Glu 165	3.2 ± 0.4
		Asp 165	2.2 ± 0.1
(3) H charge 0.311→0.2 phosphate→phosphate	covalent	Glu 165	11.0 ± 0.2
		Asp 165	4.9 ± 1.5
(4) H charge 0.311→0.0 phosphate→phosphate	covalent	Glu 165	26.1 ± 1.2
		Asp 165	10.3 ± 1.2

Differences Between The Perturbations Above

Difference	Protein Model	ΔΔG	Interaction
(1)-(2)	Glu 165	25.2	Lys 13 and substrate
	Asp 165	28.9	
(1)-(3)	Glu 165	17.5	Lys 13 and phosphate portion of substrate
	Asp 165	27.0	
(3)-(2)	Glu 165	7.7	Lys 13 with substrate minus phosphate group
	Asp 165	1.9	

[a] See Figure 5.4d.

[b] The free energy changes reported here for the wild-type complex correspond to ΔG_3 in the free energy cycle given in the methods section. The values for the mutant enzyme represent ΔG_6.

Table 5.4 also contains free energy changes for interactions between Lys 13 and portions of the substrate. The difference between the free energy changes ($\Delta\Delta G$) for perturbation 1 (that is, Lys interactions with the environment) and perturbation 2 (Lys interactions with everything but the substrate) represents the free energy of interaction between Lys 13 and DHAP. The free energies for this interaction, for both the wild-type and mutant covalent structures, are large; the interaction is almost 4 kcal/mole more favourable for the mutant. The difference between perturbations 1 and 3 (Lys interactions with everything but the phosphate portion of the substrate) reflects interactions between Lys 13 and the phosphate group of DHAP. This interaction is much stronger for mutant TIM (by approximately 10 kcal/mole) than for the wild-type enzyme. The difference between perturbations 2 and 3 is a measure of the interactions between Lys 13 and the non-phosphate portion of the substrate. For this case, the wild-type covalent state structure is favoured by almost 6 kcal/mole compared to the mutant structure. These results indicate that the substrate interacts strongly with the lysine in the mutant transition structure, although in a nonproductive manner (e.g. Lys 13 interacts strongly with the phosphate group instead of polarizing O2). As can be seen in Table 5.6, the N_{ε}-O2 distance is 1.3 Å greater for the mutant transition structure than for the wild-type structure. In contrast, the ε-amino group of Lys 13 is closer to the phosphate group in the mutant structure. (See relevant distances in Table 5.6 and compare structures A and B of Figure 5.6.)

Glutamate 97

The free energy change for neutralizing fully charged Glu 97 in the noncovalent wild-type complex is 1.3 kcal/mole higher than in the mutant complex. The added stability in the wild-type case may be due to more effective interactions between O2 of Glu 97 and O1H of the substrate (this distance is ~0.8 Å shorter in the wild-type structure compared to the mutant, Table 5.6). This Glu 97 substrate interaction is the only interaction that we found to be very different between the two structures. Even in the wild-type

TABLE 5.5
Free Energy Changes Upon Perturbing Charges and Hydrogen Bond Dipoles in Substrate (kcal/mole)[a]

O1 Dipole		
Protein Model	Substrate Model	ΔG^b
Glu 165	noncovalent	2.8 ± 0.9 / 2.1 ± 0.8
Asp 165	noncovalent	1.5 ± 0.1 / 1.1 ± 0.5
Glu 165	covalent	-1.9 ± 0.2
Asp 165	covalent	-3.4 ± 0.9
O2 Dipole		
Protein Model	Substrate Model	ΔG^b
Glu 165	noncovalent	6.8 ± 0.9 / 3.9 ± 0.4
Asp 165	noncovalent	6.2 ± 1.0 / 2.8 ± 0.9
Glu 165	covalent	9.4 ± 0.2
Asp 165	covalent	7.9 ± 0.5
O1/O2 Dipole		
Protein Model	Substrate Model	ΔG^b
Glu 165	noncovalent	6.4 ± 1.2 / 7.8 ± 0.8
Asp 165	noncovalent	7.0 ± 0.7 / 8.0 ± 0.3
Glu 165	covalent	15.6 ± 0.6
Asp 165	covalent	1.9 ± 0.3

[a] See Figure 5.3.

[b] Two free energy changes are listed for each noncovalent complex. The first represents the wild-type side chain orientation of residue 165 (regardless of whether it is Glu or Asp) and the second refers to a slightly different orientation of residue 165 generated during a trial run aimed at perturbing Glu directly into Asp. Also, to conform to the formalism outlined in the methods section, the free energy changes reported for the wild-type noncovalent structure represent ΔG_{ES}, those for the mutant noncovalent structure represent $\Delta G_{E'S}$, those for the wild-type covalent complex are ΔG_{ETS}, and those presented for the mutant covalently bound substrate structure represent $\Delta G_{E'TS}$.

TABLE 5.6
Distances (in Å) Between Active Site Residues
for Wild type TIM and Mutant TIM (Asp 165)

Residues Involved	Atoms Involved[a]	Noncovalent		Covalent	
		GLU[b]	ASP[c]	GLU[d]	ASP[e]
Asn 11, Substrate	HN_{δ} , O1	1.85	1.82	2.12	3.81
Lys 13, Substrate	N_{ε} , O2	2.75	3.61	3.00	4.31
Lys 13, Substrate	N_{ε} , O3	4.18	4.22	4.44	4.06
Lys 13, Substrate	N_{ε} , O4	2.48	2.53	2.64	2.52
Lys 13, Substrate	N_{ε} , O5	2.60	2.59	2.84	2.53
Lys 13, Substrate	N_{ε} , O6	4.54	4.37	4.68	4.47
Lys 13, Glu(Asp)165	N_{ε} , O2	5.51	6.91	7.10	7.07
His 95, Substrate	HN_{ε} , O1	2.76	3.36	2.70	3.04
His 95, Substrate	HN_{ε} , O2	2.21	1.70	2.43	1.76
His 95, Substrate	HN_{ε} , O1H	2.26	2.55	2.10	2.27
His 95, Glu(Asp)165	HN_{ε} , O2	1.78	4.46	4.47	4.61
Ser 96, Glu(Asp)165	HOG , O1	1.72	2.48	2.20	1.77
Ser 96, Glu(Asp)165	HN , O1	1.69	1.85	2.46	1.67
Glu 97, Substrate	O2, O1H	3.87	4.66	5.28	6.70
Glu 97, Lys 13	O1, HN_{ε}	1.74	1.69	1.82	1.67
Glu(Asp)165, Substrate	O2 , O2	3.26	4.40	4.16	4.31
Glu(Asp)165, Substrate	O1 , O1	4.08	6.33	3.39	3.40
Glu(Asp)165, Substrate	O2 , H1R	3.08	5.29	1.20	1.20
Glu(Asp)165, Substrate	O2 , C1	3.54	5.51	2.70	2.70

[a]Where there is more than one choice for the distances involving hydrogens, the lowest distance is reported.

[b]Figure 5.5a.

[c]Figure 5.5b.

[d]Figure 5.6a.

[e]Figure 5.6b.

enzyme this distance is fairly large and Glu 97 does make closer contact with Lys 13 than with the substrate but the distances and side chain orientations are essentially the same in the wild-type and mutant structures. (The distance between Glu 97 O1 and $N_\epsilon H_1$ of Lys 13 is 1.74 Å for the wild-type compared to 1.69 Å for the mutant, Table 5.6.)

The free energy changes for perturbing the O1 and O2 charges of Glu 97 in the covalent structures differ by 3.9 kcal/mole, with the mutant structure favoured over the wild-type complex (Table 5.3). It is not clear whether this is a real effect, though. Unfortunately, the perturbation simulations of the mutant complex were not stable, yielding a large uncertainty in this value. Given the large uncertainties, the two values are essentially the same.

Substrate Perturbations

We also examined interactions in the active site from the point of view of the substrate by perturbing catalytically important portions of the molecule. Free energy changes for perturbing the O1 dipole of the substrate within both the Glu 165 and Asp 165 covalent structures are favourable (Table 5.5). The free energies for removing the O2 dipole in both covalent structures are fairly large and positive. The Glu covalent complex gains 9.4 kcal/mole in stabilization energy from the O2 dipole and is destabilized 1.9 kcal/mole by the O1 dipole, yielding a net 7.5 kcal/mole stabilization due to the environment around O1 and O2 of the substrate. The analogous overall stabilization free energy for the Asp covalent complex is 4.5 kcal/mole. Thus, on the basis of the O1 and O2 dipoles, the Glu transition state model is stabilized by 3 kcal/mole over the Asp covalent complex.

The free energies for the O1 dipole perturbation in the noncovalent wild-type and mutant complexes are essentially the same. Two values are given for each enzyme-substrate complex: the first represents the wild-type side chain orientation of residue 165 (regardless of whether it is Glu or Asp) and the second represents a slightly different orientation of residue 165 generated during a trial run aimed at perturbing Glu directly

into Asp. We can see in Table 5.5 that there is some dependence on structure. For example, the free energies for disappearance of the O2 dipole differ by 2.9 kcal/mole in the noncovalent wild-type complexes. If the same orientations are compared for the wild-type and mutant structures for a particular perturbation, however, the free energies are within the uncertainties. Despite the conformational dependence of the free energies, like orientations result in similar free energies for the wild-type and the mutant. Therefore, both mutant and wild-type substrate binding appear to be affected to the same degree by loss of the O1 and O2 dipoles.

The results for the simultaneous "disappearance" of both the O1 and O2 dipoles for the covalent complexes, as shown in Table 5.5, are not the sum of the single perturbations. To maintain the total charge of the substrate constant, we altered the charges of other substrate atoms. Different atoms were perturbed, and to different extents, in the transition state models. Hence, there is no reason to expect the results to be additive. Another explanation for the nonadditivity is that the position of the substrate changes during the perturbation simulation in the absence of the electrostatic interactions that aid in anchoring the substrate in the active site. The difference between the free energies for the wild-type and mutant structures further illustrates the favourable positions of the residues, which stabilize the O1 and O2 dipoles, in the wild-type structure. Our results show that the mutant stabilizes the O1 and O2 dipoles much less efficiently than native TIM. However, the difference between the two is large and cannot be correlated in a quantitative way with the experimentally observed 4 kcal/mole difference in k_{cat}.

DISCUSSION

Our models for the noncovalent substrate-mutant TIM (Asp 165) complex, after equilibration, show the substrate interacting with most active site residues to the same extent as in the wild-type structure, instead of being pulled in towards Asp 165. (Compare structures 5A and 5B.) For example, the distance between O2 of Glu or Asp and the pro-R-hydrogen of the substrate is much shorter for wild-type TIM than the mutant TIM

(3.08 Å and 5.29 Å, respectively), while the distances between the phosphate oxygens and the lysine group are essentially the same (Table 5.6). The calculated free energies for perturbing portions of the substrate DHAP in the active site of the wild-type and mutant noncovalent complexes are nearly the same when like orientations are compared (Table 5.5), suggesting that substrate binding is not significantly altered upon mutation of Glu 165 → Asp providing that other substrate-enzyme interactions do not differ greatly. This is in agreement with the experimental results.

The results for perturbing active site protein residues are a bit more ambiguous, because there can be many compensating interactions for any particular interaction that we observe, any comparison to the experimental results is tenuous unless all possible interactions in the active site are evaluated, since $\Delta\Delta G_{bind}$ and $\Delta\Delta G_{cat}$ correspond to the effect of all of the individual free energy changes. The perturbations of the substrate should be measures of enzyme-substrate interactions that can be related, at least qualitatively, to the experimental results while the protein residue perturbations are most useful for ascertaining which residues are important for substrate binding and catalysis. We found differences between the free energy changes of interaction for charge perturbations of Asn 11, Ser 96, and Glu 97 in the noncovalent complexes and we have suggested possible reasons for the differences, which could be tested experimentally.

Upon mutation of Glu 165→Asp 165, k_{cat} drops by approximately three orders of magnitude (Straus et al., 1985). At least three explanations for the drop in activity of the mutant have been proposed. Alagona *et al.* (1986) suggest two plausible interpretations of the lower catalytic activity of the Asp 165 mutant. They show that a small (0.3 Å) change in the C...O distance in the transition state (normally 2.6 Å) could result in a 4 kcal/mole higher barrier to proton transfer from the substrate to the enzymic carboxylate. They also note that if the Asp 165 mutant is able to achieve the 2.6 Å distance without extra stereochemical strain, the increase in the free energy of activation might be rationalized in terms of less effective interactions between the substrate in the covalent com-

plex and electrophilic groups in the active site. The focus of this paper is the second possibility for the lower activity of the Glu 165 –> Asp mutant suggested by Alagona *et al.*, less effective interactions between some electrophilic group(s) in the active site and the substrate. We forced the substrate and enzyme to adopt transition structures by imposing a covalent bond between the two (Figure 5.3) and evaluated various interactions in the wild-type and mutant transition structures without regard to how these structures might actually be attained. We cannot simulate bond making steps using the approach outlined here, therefore the question of the distance between the attacking carboxylate oxygen of Glu or Asp and the pro-R-hydrogen remains a possible source of the drop in activity of the mutant enzyme. Nevertheless, we suggest that less effective interactions between Lys 13 and the substrate in the transition state structure of the Asp 165 mutant might explain the observed drop in catalytic activity.

The free energies reported here for perturbing the charges of the other active site residues, besides Lys 13, in the wild-type and mutant transition structures are all within the reported uncertainties (Table 5.3); the only striking difference between the structures are seen for the Lys 13 perturbations (Table 5.4). While the results for perturbing Lys 13 are not directly comparable to the substrate perturbations, they offer a plausible explanation for why the mutant is catalytically less effective than the wild-type enzyme, but this does not rule out other interactions. By looking at $\Delta\Delta G$ for the Lys 13 perturbations we were able to separate the interactions between this residue and the catalytic portion of the substrate and the phosphate group. Based on these results, in the mutant Lys 13 interacted strongly with the substrate in a nonproductive manner, by strong interactions with the phosphate group but not the catalytically important portions of the substrate.

Raines *et al.* (1986) offer an alternative explanation for the decrease in "transition state binding" based on the geometry of proton abstraction by the enzymic base. Earlier Gandour (1981) postulated that a carboxylate group is an approximately 100-fold better catalyst when the proton is transferred in a *syn* orientation (to both carboxylate oxygens

simultaneously) than in an *anti* orientation (to only one carboxylate oxygen). Given this argument, k_{cat} would be expected to decrease if Asp 165 of the mutant were to abstract a substrate proton in an *anti* orientation. Current theoretical approaches cannot definitively establish which of the three explanations is correct, or indeed if some other explanation is.

Nonetheless, our interpretation of the drop in activity of the mutant enzyme could be tested by site-specific mutation of the lysine residue. We would suggest that replacing Lys 13 by Arg might result in an increase in activity of the Asp 165 mutant. The longer side chain of Arg may facilitate stabilization of the substrate by this mutant. Furthermore, in the absence of unforseen complications, we would expect a single Lys 13→Ala mutant (Glu 165, Ala 13) and the corresponding double mutant (Asp 165, Ala 13) to have comparable, low activities. Alanine would be unable to stabilize the substrate, if the positive charge in this position is indeed crucial.

The nature of the intermediate---enediol or enediolate---has not been definitively established. Iyengar and Rose (1986) argue in favour of the enediol. On the other hand, the effectiveness of phosphoglycolate (Wolfenden, 1969) and phosphoglycohydroxamate (Collins, 1974) in inhibiting TIM suggests that the enediolate plays an important role at some stage of the reaction. One could imagine protonation of the enediolate by water with a low activation energy (Alagona et al., 1984). The covalent models examined in this study employ an enediolate. If the rate limiting transition state of the reaction involves formation of the enediol, then the models we have constructed are not appropriate for interpretting the experimental data. However, if, as mentioned above, the enediolate is important and is rapidly protonated (compared to formation of the enediolate) by solvent, then our models are relevant. (For a more thorough discussion see reference 20.)

Another limitation of our models is that solvent is not explicitly present. We attempted to compensate for this by using a distance dependent dielectric function; how-

ever, this dielectric model leads to an overestimation of the effect of changing a full charge. The commonly accepted value of the internal dielectric constant of the protein interior is between 1 and 5 (Pethig, 1979). Recent experimental work suggests, however, that the effective dielectric constant of the protein interior is between 40 and 50 (Rees, 1980; Russell and Fersht, 1987; Russell et al., 1987). The use of $\varepsilon = r_{i,j}$, then, severely underestimates the influence of the solvent on the dielectric constant of proteins. The simple addition of solvent and the use of a dielectric constant of unity would probably still result in a poor model for the heterogeneous dielectric environment within the protein. In addition, it is not appropriate to construct such an elaborate model given the low resolution of the available X-ray stuctures. Despite these shortcomings, the importance of this study lies in the development of the free energy component analysis, in which one perturbs specific portions of the enzyme or substrate. This method can give qualitative insight into the residues that are important in binding and catalysis. Thus, free energy calculations can be useful in the qualitative way described here as well as the quantitative approaches used in the studies by Bash *et al.* (1987b) and Rao *et al.* (1987).

Knowles and co-workers have engineered other interesting mutants, whose properties we can discuss in terms of the results presented here. Asp 165, Pro 96---a pseudo revertant of the relatively inactive Asp 165, Ser 96 single mutant---is a significantly better catalyst than the single mutant (Hermes et al., 1987). Our results suggest that Ser 96 forms hydrogen bonds of similar strength with both native TIM and the Asp 165 mutant in the covalent structures, although we do see a difference for the noncovalent complexes. The calculated $\Delta\Delta G^*_{cat}$ contribution from Ser 96 is 2.8 kcal/mole (favouring the Asp mutant) but this value is within the sum of the uncertainties, making any quantitative arguments tenuous. Therefore, the strength of these hydrogen bonds probably cannot be used to explain the relative catalytic activity of the Glu 165, Pro 96 and Asp 165, Pro 96 mutants (Hermes et al., 1987). Mutating Ser 96→Pro in the native enzyme reduces k_{cat} such that both Pro 96 proteins (that is, Glu and Asp in position 165) have

comparable activities. The Pro 96 mutants appear to exert their effect by changing the relative orientations of the key catalytic groups at positions 165, 95 and 13. A change in these orientations can be deleterious, as in the case of the native structure. But it may be advantageous to alter the Asp 165 enzyme to bring His 95 or Lys 13 into a better orientation to stabilize the transition structure.

It is interesting to note that a pseudo revertant of His 95 →Asn 95 is the double mutant (Asn 95, Pro 96) (Hermes et al., 1987). That Pro 96 can revert mutations at both positions 95 and 165 is consistent with the idea that the relative orientation of these groups is critical for catalysis. Actual simulations of Pro 96 mutants are required to assess the speculations presented above on the low activity of the Asp 165, Asn 95 enzyme and the partially restored activity of the Asp 165, Pro 96 and Asn 95, Pro 96 structures. We have performed molecular dynamics simulations of these mutants and Pro 96 appears to exert its effect by altering the orientations of other active site residues (Chapter 6).

CONCLUSIONS

Free energy perturbation calculations can give interesting insights into the effects of amino acid substitutions on both substrate binding and catalysis. We have shown how a free energy component analysis, in which one perturbs the properties of individual groups on the enzyme or ligand, yields detailed information, albeit qualitative, about the specific interactions important in enzyme action. In the specific application studied here, we have used a simple model without explicit inclusion of solvent and a distance dependent dielectric constant to compensate for the lack of solvent. Thus, we have not implemented the free energy component analysis in as rigorous a fashion as might have been warranted if we had a better X-ray structure and the reaction had not involved a highly charged substrate. Nonetheless, we suggest that the free energy component analysis tool can be as useful and insightful as energy component analysis has been in molecular mechanics studies (Wipff et al., 1983; Kollman et al, 1981; Singh et al., 1986b).

We examined the importance of electrostatic interactions between active site residues of triose phosphate isomerase and portions of the substrate dihydroxyacetone phosphate. Our results indicate that the charge interactions examined contribute equally to binding in the wild-type and Asp 165 mutant enzymes. This is consistent with the experimental observation that substrate binding does not change substantially upon replacement of Glu 165 by Asp. Furthermore, our results suggest that less effective interactions between Lys 13 and the non-phosphate portion of DHAP in the mutant transition state for enolization may, at least partially, explain the observed drop in catalytic activity upon mutation of Glu 165→Asp. Other explanations for the observed drop in k_{cat} have been proposed, and more simulations and the X-ray structure of the mutant are required to differentiate between the possibilities.

CHAPTER 6: Molecular Dynamics Simulations of Active Site Mutants of Triosephosphate Isomerase

The ability to selectively modify individual amino acids in proteins is of great use in furthering our understanding of the underlying forces governing enzyme action. Triosephosphate isomerase (TIM) is a good candidate for this type of approach because it is well-characterized in both mechanistic and energetic terms. TIM catalyzes one of the simplest reaction in metabolic biochemistry, the interconversion of dihydroxyacetone phosphate (DHAP) to glyceraldehyde 3-phosphate (GAP) by the transfer of a single proton. We chose to study this system using computer simulation approaches because of the wealth of binding and kinetic data on both native TIM and a variety of active site mutants.

The catalytic mechanism of TIM is shown schematically in Figure 6.1 and the complete free energy profile for this reaction has been determined (summarized by Knowles and Albery, 1977). After substrate binding, the carboxyl group of Glu 165 abstracts the pro-R proton from the C1 postion of DHAP, resulting in an enzyme-bound enediol or enediolate (the nature of the intermediate is not known). A proton is then delivered to the C2 position of the substrate, yielding GAP. An electrophilic residue is thought to polarize the carbonyl group in the enzyme-substrate complex, thereby facilitating proton abstraction (shown in Figure 6.1 as HA) (Belasco and Knowles, 1980). Furthermore, this electrophilic residue can then stabilize the developing negative charges on the oxygens at C1 and C2 during formation of the enediol (or enediolate) by providing the substrate with a positive electrostatic environment. Based on the X-ray structure of Banner et al. (1975), Lys 13 and His 95 appear to be good candidates. (See Figure 6.2 for a schematic representation of the important active site residues.) Alagona et al. (1984, 1986) have demonstrated, using quantum and molecular mechanics calculations, how these residues can facilitate catalysis; their results are consistent with the results and proposals of Belasco and Knowles (1980).

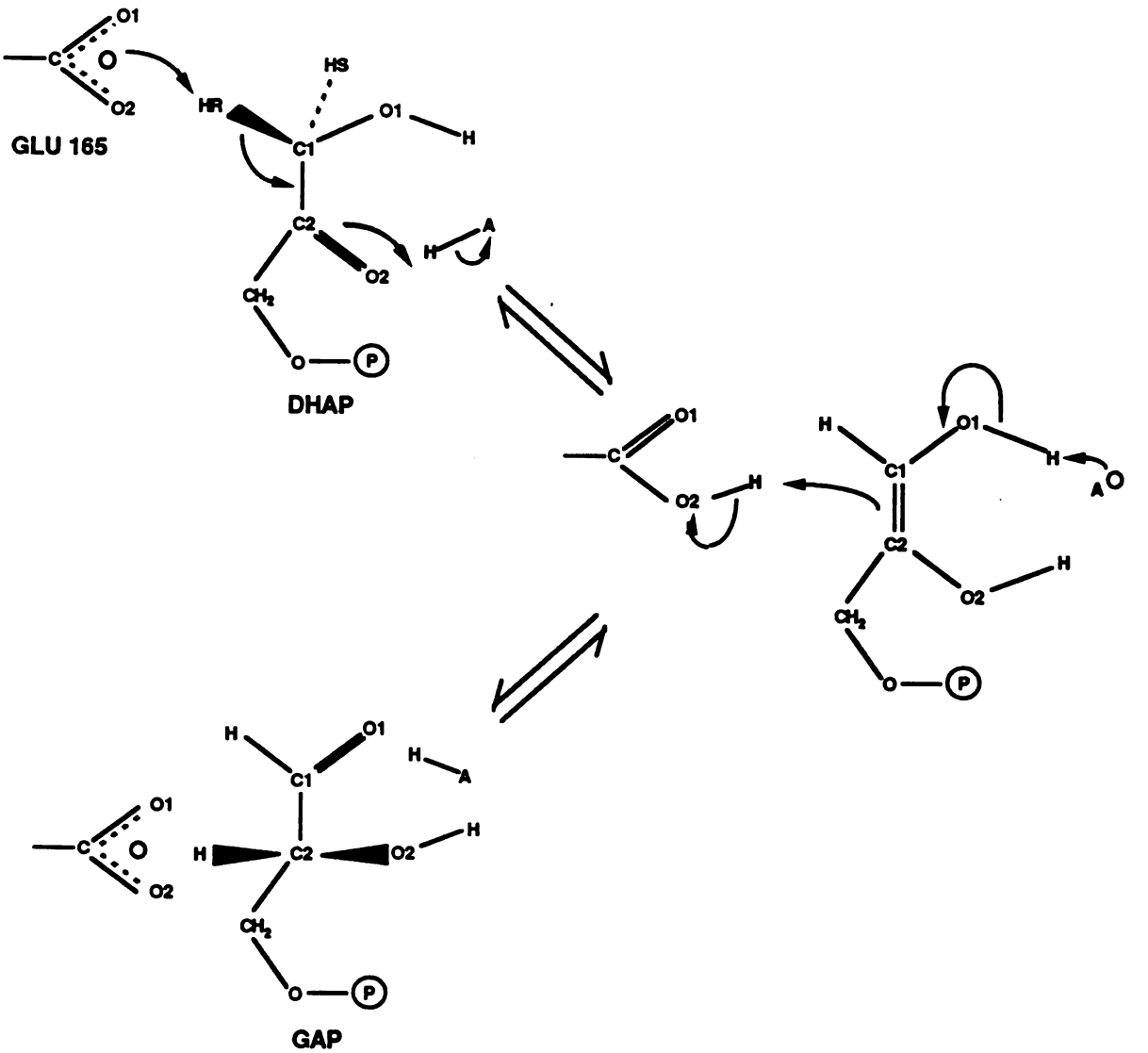


Figure 6.1. Reaction catalyzed by Triosephosphate Isomerase (TIM). HA is an electrophilic residue(s)---probably lysine 13 and/or histidine 95---that stabilizes the developing negative charge on the substrate carbonyl oxygen (O2) during formation of the enediol/enediolate intermediate.

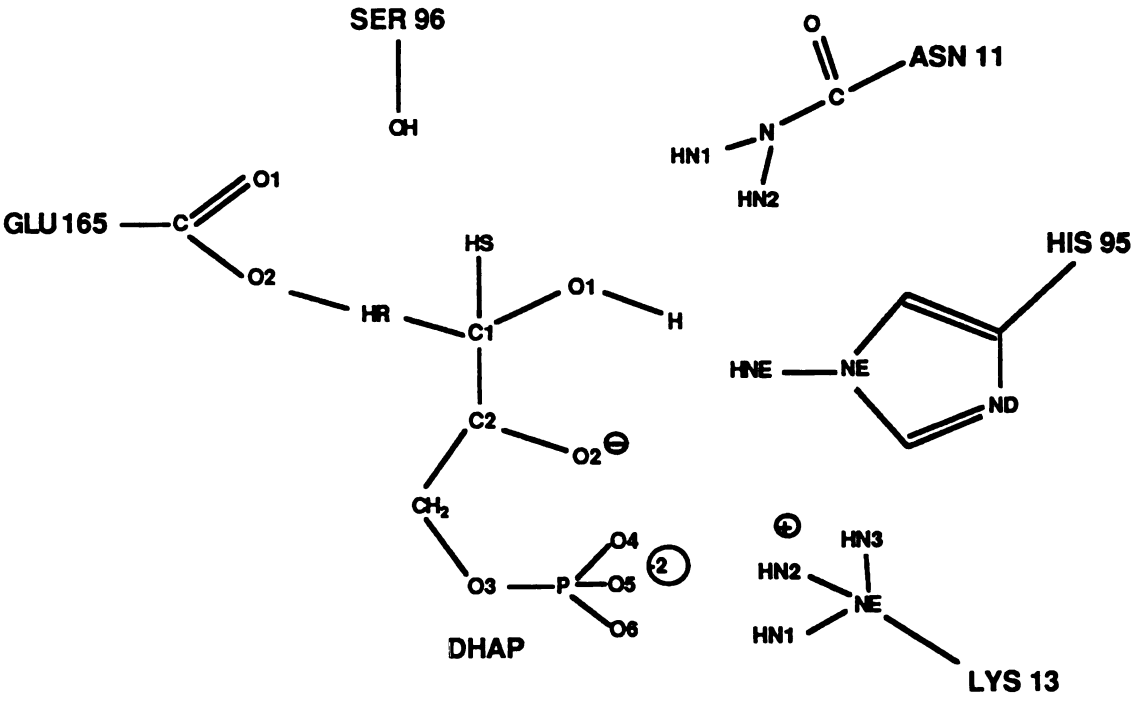


Figure 6.2. Schematic Representation of Active Site of TIM. Covalently Bound substrate (DHAP) and key active site residues of TIM.

TIM is considered to be a perfectly evolved enzyme, partly because the rate determining transition state is for product dissociation and not a chemical step (Albery and Knowles, 1976a; Blacklow et al., 1988). Therefore, there is no evolutionary pressure for improving chemical steps in the reaction. Knowles and co-workers have proposed, on the basis of their findings over the years, that there are three specific mechanisms by which catalytic efficiency could evolve (Albery and Knowles, 1976b). Comparison of isomerases from different species with different activities might provide information as to which interactions lead to improved catalytic activity and whether these mechanisms are consistent with the hypothesis of enzyme evolution of Albery and Knowles (1976b). Since there are no naturally occurring examples of less highly evolved triose phosphate isomerases, Knowles and co-workers set out to study forward evolution by generating "imperfect" isomerases using site-directed mutagenesis techniques and then randomly producing pseudorevertants with increased activity. A variety of active site residues have been altered; in each case a decrease in the catalytic activity of the mutant results, so that the reactions are no longer diffusion-controlled (Hermes et al., 1987). One of the effects of these mutations is on the transition state for enolization, such that it is now the rate-limiting step in the reaction. Hence, these mutants are now susceptible to evolutionary development. By subjecting the genes encoding two of these mutants (one in which the catalytic base Glu 165 is replaced by Asp and the second in which the active site His 95 is mutated to Asn) to heavy random mutagenesis and then selecting for transformants that synthesize isomerases with increased catalytic activity, Knowles and co-workers found two second-site suppressor mutants (Table 6.1). Surprisingly, in both cases the increase in catalytic activity is a result of the replacement of Ser 96 by Pro. [Table 6.1 contains the comparisons of k_{cat} for isomerization of DHAP; the (E165D, S96P) pseudorevertant affects the k_{cat} of isomerization of GAP to DHAP. See Blacklow and Knowles, 1990.]

Our goal was to examine molecular interactions in the active site that might explain

TABLE 6.1
Catalytic and Binding Properties of Mutant Isomerases with the Substrate Dihydroxy Acetone Phosphate Relative to Wild-Type Triosephosphate Isomerase[a]

Enzyme	Amino Acid at Position			Relative Rate of Catalysis	Relative Binding Affinity
	165	95	96		
Wild type	Glu	His	Ser	100	100
Mutant S96P	Glu	His	Pro	1.83	271
Mutant E165D	Asp	His	Ser	0.68	54
Pseudorevertant from E165D	Asp	His	Pro	0.57	1226
Mutant H95N	Glu	Asn	Ser	0.03	110
Pseudorevertant from H95N	Glu	Asn	Pro	1.13	162

[a] The values for the relative rate of catalysis were derived from the k_{cat} values of Blacklow and Knowles (1990); likewise, the relative binding affinity reflects differences in K_m.

the drop in activity of the TIM single mutants, the increase in activity of the corresponding pseudorevertants, the differences in substrate binding brought on by the Ser 96 → Pro replacement, and the ability of TIM to tolerate changes in amino acid sequence. To this end, we performed molecular dynamics (MD) simulations, using an empirical force field, of wild-type TIM and of each of the mutants: (E165D); (E165D, S96P); (H95N); (H95N, S96P); and the control mutant (S96P). All mutant structures were derived from the X-ray structure for wild-type TIM (Banner et al., 1975) with the replacement of the appropriate residue to form the mutant, as crystal structures for the mutants are not yet available. In the case of the (H95N) mutant and the (H95N, S96P) double mutant with the substrate DHAP, the effect of the Pro is primarily upon catalysis (Table 6.1). On the other hand, the ability of the enzyme to bind DHAP is compromised while catalysis is unaffected upon replacement of Ser 96 by Pro in the (E165D) mutant (Table 6.1). So, we performed simulations of the enzyme with DHAP in the binding pocket and with DHAP covalently linked to the enzyme (O2 of residue 165 was covalently linked to HR of the substrate, Figure 6.2) to serve as a model for the transition structure for enolization. Our hypothesis is that our models of the different mutants with noncovalently bound substrate should show differences in interactions between the enzyme and substrate relevant to the experimental binding data. We also assume that our models with covalently bound substrate reflect differences in the ability of the mutants to stablize the transition state for enolization, which becomes the rate-limiting step for the mutants with DHAP. The difference between these states can then be discussed in light of the kinetic data (k_{cat}).

In addition to the experimentally characterized mutant isomerases in Table 6.1, we performed MD simulations of two hypothetical mutants in which the active site Lys 13 was replaced by Arg in the (E165D) mutant and wild-type TIM. In an earlier study of the Glu 165 → Asp 165 mutation and its effect on catalysis and binding, we found that Lys 13 interactions with the substrate were disrupted in the mutant (Chapter 5). In wild-type TIM Lys 13 appeared to be important in stabilizing O2 of the substrate in the enediolate

form. In the (E165D) mutant, Lys 13 interacted strongly with the substrate but in a catalytically nonproductive manner; i.e., Lys 13 interacted with the phosphate portion of the substrate instead of with O2. We proposed, on this basis, that the replacement of Lys 13 by Arg might improve interactions with O2 in the (E165D) mutant.

In the molecular dynamics simulations presented here, we found less effective electrostatic stabilization of the transition state structures of the single mutants compared to the pseudorevertants. We found that interactions between electrophilic active site residues with both O1 and O2 of the substrate were critical for stabilization, comparing the noncovalent to the covalent complexes, and that the degree of the interaction of the enzyme with these atoms correlated qualitatively with catalytic activity. Pro 96 played an indirect role by altering the orientations of other active site residues interacting directly with the substrate. The pseudorevertants optimized interactions with O1 and O2 of the substrate by using non-wild-type interactions when necessary, pointing out the degeneracy of the electrophilic residues in the active site. In the case of the noncovalent complexes, the simulations demonstrated a change in geometry that may facilitate proton abstraction upon addition of Pro to the (E165D) mutant, and this mutation led to more favourable catalytically productive interactions in the (H95N) mutant. The motions of the active site residues were highly correlated during molecular dynamics, preventing an unique mechanistic description of how proline exerts its effect upon neighboring residues.

METHODS

Generation of Structures

Covalent Complexes

The X-ray structure of native TIM by Banner et al. (1975) was used as the starting point for this study. Construction of the covalent complex between DHAP and the wild-type TIM dimer and the corresponding (E165D) mutant complex has been described

(Chapter 5). All other structures discussed in this study were derived from these original structures, therefore they warrant some discussion here. The transition structure complexes were constructed by imposing a covalent bond between the pro-R hydrogen (HR) of the substrate DHAP and O2 of Glu 165, or O2 of Asp 165 (Figure 6.2). These models were then energy-minimized and subjected to molecular dynamics at 300 K for 2 picoseconds (ps). We took these final structures, after molecular dynamics, as our starting structures for the present study and considered them structures at time = 0 ps.

All of the calculations described below were done using AMBER (Singh et al., 1986). Standard united-atom parameters (Weiner et al., 1984) were used for the TIM dimer for all of the complexes. Both standard (Weiner et al., 1986) and nonstandard all-atom parameters were used for the covalently bound DHAP. The nonstandard parameters and charges used for the substrate were taken from Chapter 5.

The various mutant TIM complexes described in this study were generated by replacing the residue of interest in the appropriate structure described above, wild-type TIM (Glu 165) or the (E165D) mutant. To avoid biasing the results, the orientations of the swapped side chains were not reoriented but instead fit as closely as possible to the orientation of the residue being replaced. The (S96P) mutant was used directly after swapping residue 96. Further preparation for MD was necessary for the other mutants.

The following protocol was used to prepare the (E165D, S96P), (H95N), and (H95N, S96P) mutants for molecular dynamics. The structures were minimized briefly to remove any bad contacts. 1200 cycles of minimization were performed; the first 200 cycles were done using the steepest descent method of minimization and the remaining 1000 cycles utilized the conjugate gradient method. Only the swapped residue(s) was allowed to move. A non-bonded cut-off of 10 Å and a linear distance dependent dielectric constant ($\varepsilon=r$) were employed. The resulting structures were then brought to 300 K and equilibrated for 0.5 ps using molecular dynamics. A 1 femtosecond time step was used for the equilibration. Other residues were allowed to move besides the swapped

residue during the equilibration. Any residue in the wild-type complex within 12 Å of C_α of Pro 96, C_γ of Pro 96, O2 of Glu 165, or O2 of the substrate was allowed to move; 95 residues fell into this group. These same 95 residues were used for all of the complexes. The resulting complexes, after the brief equilibration described above, were referred to as the time = 0 ps structures.

The protocol described above was altered somewhat to prepare the (K13R) and (E165D,K13R) mutants for molecular dynamics, because the enzyme did not tolerate the Lys → Arg substitution as readily as the other mutations described above. 2000 cycles of steepest descent minimization were performed on each of the mutants, allowing only Arg 13 to move. Then, full minimization, allowing all residues to move, was performed to a root-mean-square (r.m.s.) energy gradient of 0.5 kcal/mole-Å. The full minimizations were carried out using the steepest descent method for the first 200 steps followed by conjugate gradient minimization with a short non-bonded cut-off of 6 Å. The resulting structures were equilibrated using the protocol described above with the exception that the structures were equilibrated for 1 ps not 0.5 ps. The equilibrated structures were the time = 0 ps structures for the molecular dynamics simulation.

Noncovalent Complexes

The noncovalent complexes---(E165D), (E165D, S96P), (H95N), (H95N, S96P)---were derived from the relevant covalent complexes. We started from the covalent complexes instead of using the procedure described above for constructing mutant structures from the wild-type enzyme, so that the simulations would begin from catalytically relevant orientations. Essentially the same procedure as described above was used to prepare the noncovalent structures for MD. The structures were minimized, allowing only the substrate to move in the case of the (E165D) mutants and allowing the substrate and residues 95 and 96 to move in the (H95N) mutants. Each structure was then equilibrated for 1 ps with the same 95 residues described above allowed to move.

Computational Details of Molecular Dynamics Simulations

The simulations began with the time = 0 ps structures at 300 K described in the previous section. The temperature was maintained at 300 K by coupling to an external bath using the method of Berendsen et al. (1984). The protocol used for the actual simulations was similar to that employed for equilibration. Only the 95 residues allowed to move during equilibration were mobile during the MD simulation. 10,000 total steps of MD were performed for each mutant on a Cray X-MP (San Diego Supercomputer Center). SHAKE (Ryckaert et al., 1977) was used on all bonds so that a stepsize of 2 femtoseconds could be employed, resulting in 20 psec trajectories for each complex from the time = 0 ps structures (e.g. equilibration time was not included in the total time). A 10 Å non-bonded cut-off was used and the pairlist was updated every 50 steps. Structures were saved every 0.2 psec during the simulations for analysis, resulting in 100 structures for each complex.

We found that the angle for proton transfer O2-HR-C1 (Figure 6.2) collapsed during MD of the covalent complexes using the substrate parameters given in Chapter 5 and the procedure described above. One expects, based on quantum mechanical calculations, that the transition state for proton transfer involves a near linear O2-HR-C1 angle (Chapter 5). So, we performed control calculations with higher force constants on this angle to maintain near linearity. The final structures after MD, using the protocol above with K_θ =0 kcal/mole-rad^2 for the proton transfer angle, were energy-minimized to a r.m.s. gradient of 0.5 kcal/mole-Å with K_θ = 100 kcal/mole-rad^2, allowing only the substrate and catalytic base to which it was attached to move. This procedure maintained orientations generated during MD while restoring the ideal proton transfer geometry. MD simulations with higher force constants (K_θ=50 or 100 kcal/mole-rad^2) were also performed from either the starting structures described above (time = 0 ps structures) or the structures minimized with K_θ = 100 kcal/mole-rad^2. The MD protocol was the same as that described above with the exception of the change of the one angular parameter.

RESULTS

We performed molecular dynamics simulations of the triose phosphate isomerases in Table 6.1. We first present the results of simulations of the noncovalent complexes, focussing on properties of the final structures after 20 ps of MD. Next we present the results of the simulations of the covalent complexes, in which we examined specific interactions in the final, static structures following 20 ps of MD and minimization. After identifying interactions that appeared to correlate with changes in activity, we focussed on the average properties of these interactions during the simulations. We present the results of the analysis of the static covalent structures first and then their average dynamical properties.

A. Noncovalent Complexes

Distances between particular interacting atoms in the final isomerase structures, with noncovalently bound DHAP, after 20 ps of MD are given in Table 6.2. There are a few features that the mutants have in common when Ser 96 is replaced by Pro. Ser 96 aids in orientation of the catalytic base by forming an hydrogen bond (Figure 6.3A), which is lost upon introduction of Pro. (See the 96 N-165 O1 distances in Table 6.2.) For example, the root-mean-square deviation of the catalytic base, between the single mutants and the corresponding pseudorevertants was at least 0.3 Å with a maximum deviation of 1.1 Å for the (H95N) mutants. As a result most of the proton abstraction distances were long, making catalysis difficult (165 O1,O2-Sub HR distance; Table 6.2). Another result of the replacement of Ser 96 by Pro is that Asn 11 shifts from its postion of interacting with O1 and O2 of the substrate to form close interactions with the phosphate group of the substrate.

Although the catalytic base drifted from its original position upon replacement of Ser 96 by Pro, interactions between residue 95 and the substrate improved. In the wild-type protein His 95 forms an hydrogen bond with the backbone amide hydrogen of Glu

TABLE 6.2
Distances (in Å) Between Active Site Residues in Starting Wild-type Triosephosphate Isomerase and Mutant Isomerases with Noncovalently Bound Substrate after 20 ps of Molecular Dynamics

Distance[a]	TIM Model				
	Wild-type	Single-Site Mutants		Double Mutants	
	E165 H95 S96	E165 H95N S96	E165D H95 S96	E165 H95N S96P	E165D H95 S96P
13 HNE-Sub O2	2.08	3.74	2.42	2.56	3.26
13 HNE-Sub O4	1.68	1.63	1.76	1.75	1.79
13 HNE-Sub O5	2.61	1.61	1.76	1.69	2.56
13 HNE-Sub O6	3.94	3.41	3.89	3.92	3.84
13 NE-Sub C2	3.59	4.86	3.70	3.39	4.33
13 NE-Sub P	3.13	3.09	3.10	3.39	3.18
95 ND(O)-97 HN	3.10	3.78	3.72	3.48	2.70
95 HNE(HN2)-97 O1,O2	4.40	6.97	1.79	1.78	3.94
95 HNE(HN1,2)-Sub O1	2.76	4.99	4.38	3.77	2.69
95 HNE(HN1,2)-Sub O2	2.21	2.67	4.90	1.84	1.84
96 N-165 O1	2.69	2.66	2.85	4.28	4.40
165 O1-Sub HR	5.01	6.43	3.16	6.29	2.74
165 O2-Sub HR	3.09	4.73	4.44	4.20	3.30
165 O1-Sub O1H	5.45	6.83	5.14	3.70	1.67
165 O2-Sub O1H	3.79	5.79	5.86	1.59	3.50
Sub O1H-Sub O4	3.89	1.65	1.70	4.87	4.98
Sub O1H-Sub O5	5.29	4.13	4.19	5.36	6.02
Sub O1H-Sub O6	6.23	3.71	4.30	6.50	6.79

[a] See figure 6.2 for atom name nomenclature. Alternate atom names in parenthesis refer to mutants; the nomenclature for the atoms of Asn 95 is the same as for Asn 11 in figure 6.2. When there is more than one possible combination of atoms, the lowest distance is given.

97 [95 ND(O)-97 HN, Table 6.2 and Figure 6.3A]. This hydrogen bond became longer in the single mutants but dropped to lower distances in the double mutants. The reformation of this hydrogen bond aided in the orientation of residue 95, yielding better interactions with O1 and O2 of the substrate. This interaction was most evident for the (E165D) mutant and its revertant (Figures 6.3B and 6.3C). This orientational effect exerted by Pro 96 was less pronounced with Asn in place of His 95 but still evident (Figures 6.3D and 6.3E). In the case of the (E165D) mutant and the (H95N, S96P) double mutant, residue 95 twisted away from the substrate to form an hydrogen bond with the side chain of Glu 97 [95 HNE(HN2)-97 O1,O2, Table 6.2]. Such an interaction is detrimental to substrate binding with His 95 since it competes with protein-substrate interactions, while binding was not compromised with Asn in the place of His 95 because of the second hydrogen available to interact with the substrate.

The (E165D) and (E165D, S96P) mutants had fairly short abstraction distances after MD but exhibited different orientations with respect to the approach of the catalytic base to the pro-R-hydrogen. The (E165D) mutant adopted a geometry for abstracting a proton from the substrate through the anti orbital of the base's oxygen (the lone pair pointing away from the second oxygen, Figure 6.3B). The (E165D, S96P) mutant, on the other hand, had a geometry consistent with syn abstraction (the lone pair pointing towards the second oxygen, Figure 6.3C). The catalytic rate is estimated to be up to three orders of magnitude higher with syn abstraction compared to anti (Gandour, 1981).

The single mutants also exhibited differences in interactions between the other electrophilic residue, Lys 13, and the substrate, compared to wild-type TIM. The substrate did not bind in an extended conformation in the mutants but, instead, was slightly kinked (Figure 6.3). Evidence of the kink is shown by the dramatic drop in the distance between O1H and the phosphate oxygens (O1H-O4,O5,O6; Table 6.2). The new substrate conformation led to differences in the Lys 13 interactions. For example, interactions between Lys 13 and O2 of the substrate became weaker, especially in the case of the (H95N)

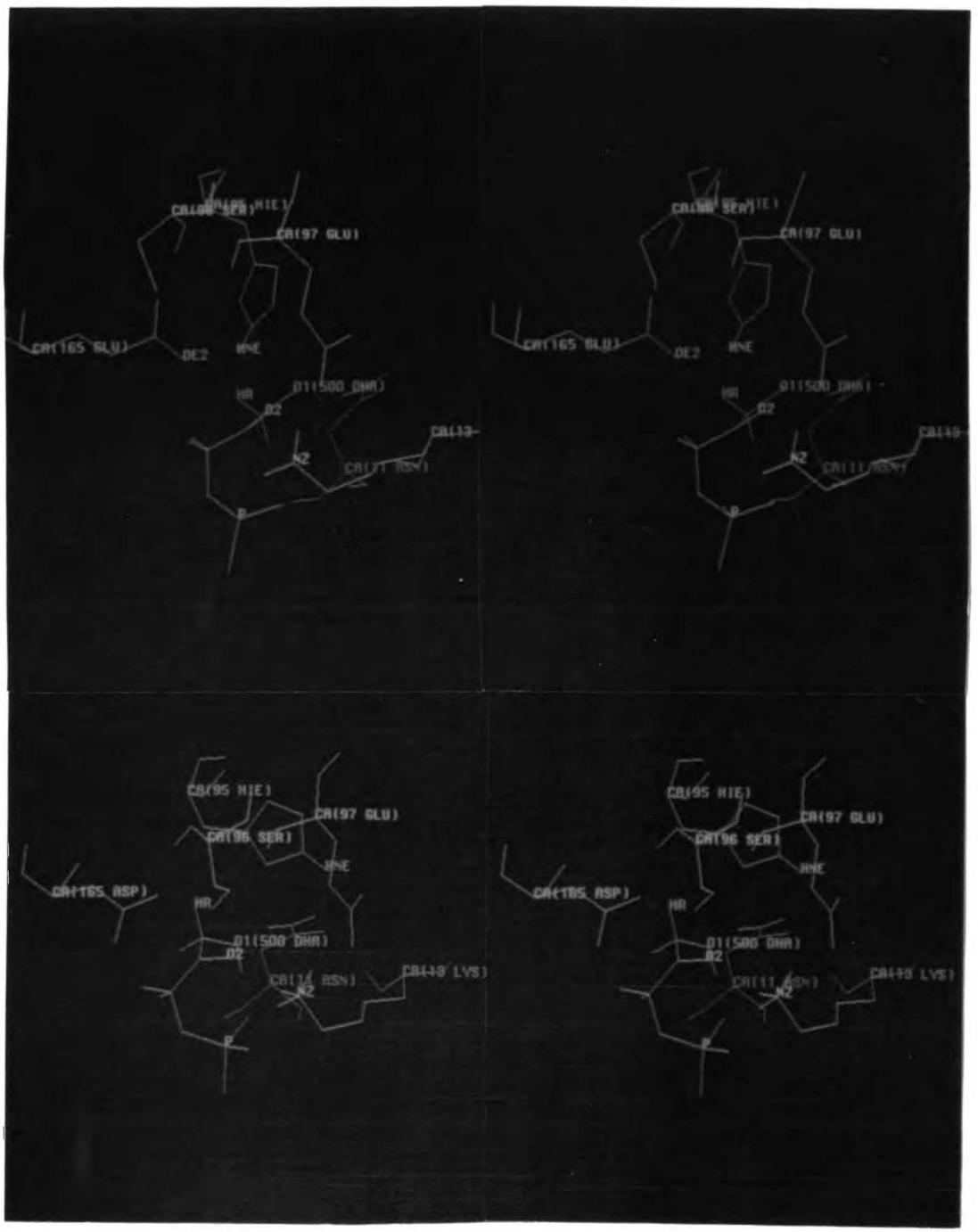


Figure 6.3. Stereoviews of noncovalent complexes: (A) Wild-type TIM; (B) (E165D); (C) (E165D, S96P) double mutant; (D) (H95N); and (E) the (H95N, S96P) double mutant. The wild-type structure is prior to MD. The mutant complexes are the final structures after 20 ps of MD. For ease of identification, wild-type TIM and the single mutants are shown in red and the double mutants are yellow.

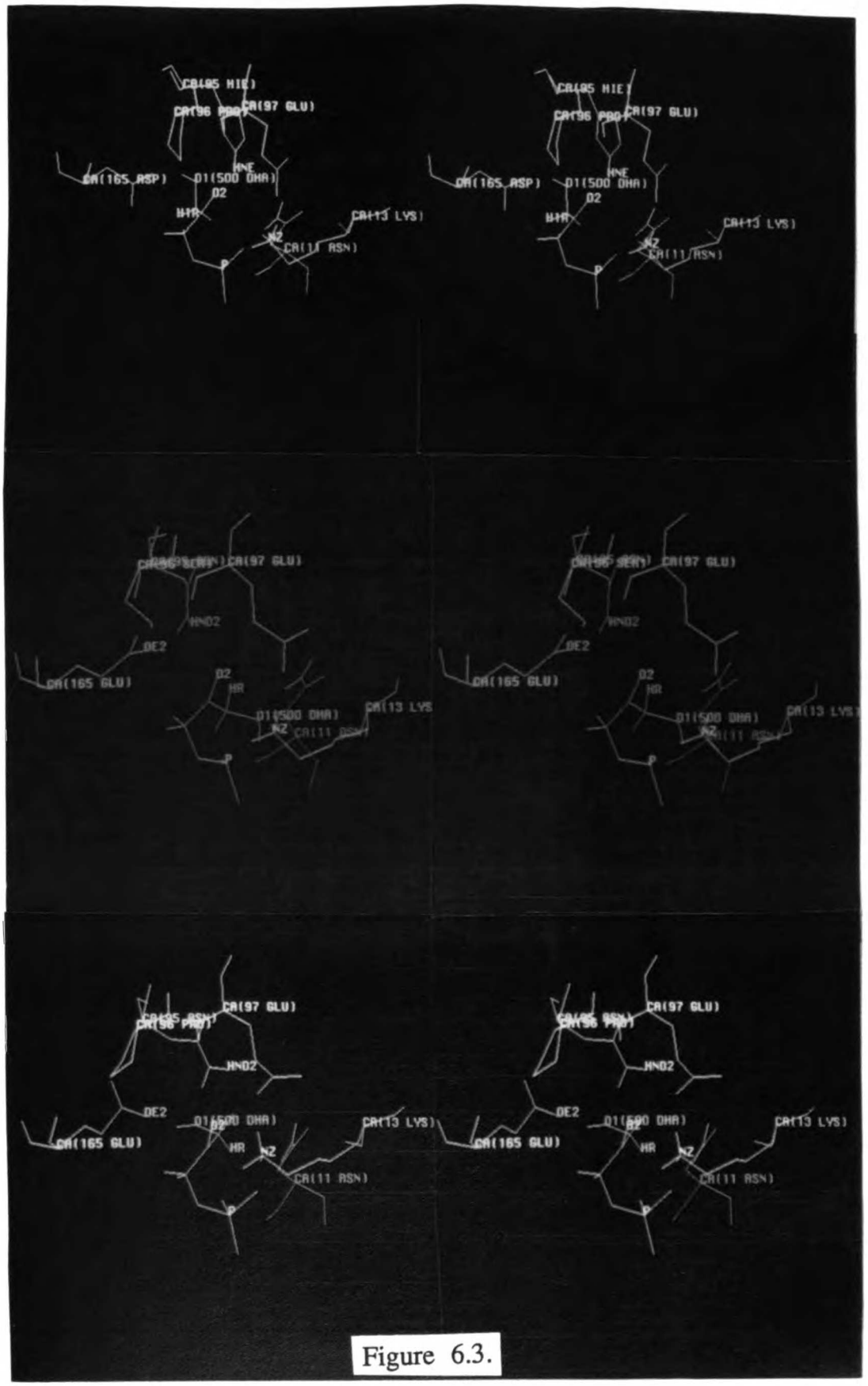

CB(85 HIE)
CA(97 GLU)
CA(96 PRO)
HNE
CA(165 ASP)
D1(500 DHA)
D2
HN
CA(13 LYS)
NZ
CA(11 ASN)
P
HND2
OE2
CA(165 GLU)
HR

Figure 6.3.

mutant (Table 6.2). Lys 13 interactions with the phosphate group, however, became much stronger (e.g. HNE 13-O5 Sub; Table 6.2). This effect of preferentially interacting with the phosphate group was most evident for the (H95N) mutant (Figure 6.3D). In this case, the distance between N_ε of Lys 13 and the phosphorus atom was comparable to that of the wild-type protein, while the distance to C2 increased by 1.3 Å.

For the (E165D) mutant the distances between N_ε and P and between N_ε and C2 were essentially the same as in wild-type TIM and the less effective interactions with O2 were due to orientational differences of the two groups. (Compare Figures 6.3A and 6.3B.) When Pro was introduced into the single mutants the substrate became more extended and the intra-substrate hydrogen bond between the hydrogen of the hydroxyl group and the phosphoryl oxygen was no longer observed. This conformational change again led to differences in interactions between Lys 13 and O2 of the substrate. In the case of the (H95N) mutant, stabilization of O2 by Lys 13 improved dramatically, leaving N_ε equidistant between C2 and P of the substrate. The introduction of proline into the (E165D) mutant, resulted in the resumption of wild-type-like interactions between Lys 13 and the phosphate group. The interactions between Lys 13 and O2 of the substrate became weaker, though.

Another interesting consequence of the collapsed substrate with the (H95N) mutants was that O1, O2 and P of the substrate were forced out of plane (Figure 6.3D). The substrate was clearly quite distorted and would require fairly substantial conformational changes to adopt a reasonable transition state structure. In fact, Glu 165 interacted strongly with O1H of the substrate instead of HR in the (H95N, S96P) double mutant (Figure 6.3E). With the kinked conformation of the substrate and with O1,O2 and P out of plane, the (H95N) mutant appeared to be in a position to facilitate intramolecular elimination of the phosphate group from the substrate, perhaps through the phosphate group's abstraction of O1H. The (E165D) and (E165D, S96P) mutants had O1, O2 and P roughly in plane---at least to the same extent as wild-type TIM. (Compare Figures 6.3A, 6.3B and

6.3C.)

B. Covalent Complexes

Control Calculations

Initially we performed MD of the covalent complexes with a force constant of 0 kcal/mole-rad^2 (K_θ) on the angle of proton abstraction of the pro-R-hydrogen (HR) by O2 of the enzymic base (<O2-HR-C1). Ideally this angle should be 180° for effective proton transfer, but we found that it collapsed during MD, although to different extents in the various complexes. After performing 20 ps of MD with K_θ=0 kcal/mole-rad^2 we adjusted the angle to 180° using energy minimization to force the substrate to adopt the transition state geometry. We then performed control calculations with an higher force constant on the angle for proton transfer, during the entire MD simulation, to see how the orientation of the active site residues was affected. These control simulations were performed on wild-type TIM and the (E165D, S96P) double mutant because the proton transfer angle in the former was least affected during MD with K_θ=0 and the latter showed the largest effect (the angle dropped from 180° to 99°). First we present the results of the control calculations, and then we describe the results of MD simulations of all of the enzymes in Table 6.1 with K_θ=0, followed by minimization to optimize the angle for proton transfer.

Table 6.3 lists the angle for proton transfer and the hydrogen bonding distances between His 95 and O1 and O2 of the substrate in the final structures after molecular dynamics with different force constants on the O2-HR-C1 angle. In the wild-type enzyme the proton transfer angle collapsed 60° with K_θ=0 kcal/mole-rad^2. A simulation from the same starting structure with K_θ=100 kcal/mole-rad^2 maintained the angle, but it showed substantially longer hydrogen bonding distances between His 95 and O1 of the substrate. When the final structure from MD with K_θ=0 was minimized with K_θ=100 kcal/mole-rad^2 to optimize the angle for proton transfer, the hydrogen bond distances were maintained. MD from the minimized structure with K_θ=100 again led to an

TABLE 6.3
Various Properties of Covalent Complexes Following Molecular Dynamics and Minimization with Different Parameters

Model and Preparation[a]	<(O2-HR-C1) (degrees)	d(95HNE-SubO1) (Å)	d(95HNE-SubO2) (Å)
Wild-type TIM			
Starting Structure	179	2.69	2.45
20 ps of MD, K_θ=0	118	2.43	1.84
20 ps of MD, K_θ=100	176	3.16	1.81
MIN with K_θ=100 following 20 ps of MD K_θ=0	174	2.39	1.76
20 ps MD with K_θ=100 following MIN with K_θ=100 and 20 ps MD with K_θ=0	175	2.98	1.75
(E165D, S96P) Double Mutant			
Starting Structure	179	3.05	1.76
20 ps of MD, K_θ=0	98	1.87	1.84
20 ps of MD, K_θ=100	177	3.21	1.75
MIN with K_θ=100 following 20 ps of MD K_θ=0	171	1.83	1.82
20 ps MD with K_θ=50 following MIN with K_θ=100 and 20 ps MD with K_θ=0	176	3.01	1.75
20 ps MD with K_θ=100 following MIN with K_θ=100 and 20 ps MD with K_θ=0	176	2.83	1.80

[a] MD = Molecular Dynamics. MIN = Minimization. K_θ = the force constant on the angle for proton abstraction by the enzymic base in units of kcal/mole-rad^2 (165 O2-Sub HR-Sub C1). See Figure 6.2 for atom names.

increase in the hydrogen bonding distance between His 95 and O1 of the substrate.

The (E165D, S96P) double mutant showed the greatest deviation of all of the complexes from the ideal proton transfer angle with MD using K_θ=0 (Table 6.3). Minimization with K_θ=100 of the structure following MD with K_θ=0 retained good hydrogen bond distances between His 95 and the substrate while optimizing the proton transfer angle. Also, the hydrogen bond distance between His 95 and O1 of the substrate increased when the force constant was increased to 50 or 100 kcal/mole-rad^2 for MD relative to the starting structures. However, with K_θ=0 the His 95-O1 hydrogen bond distance decreased during MD.

In both control systems, motion in the active site was damped and forced towards certain orientations when strict transition state geometry was imposed throughout MD. In constrast, the enzymes were able to sample more space when near the transition state geometry than with the strict transition state geometry imposed by the higher force constant. Use of K_θ=0 kcal/mole-rad^2 with the covalent structures allowed us to stay near the transition state so that catalytically relevant conformational states were sampled while reducing the number of catalytically nonproductive orientations that one would find in the noncovalent complexes. Given these factors, our analysis focusses on the MD trajectories with K_θ=0 kcal/mole-rad^2 and the corresponding final energy-minimized structures (Figure 6.4).

Analysis of Static Structures After Molecular Dynamics

Table 6.4 contains differences observed in particular interactions in the active site between the structures with the transition state geometry, e.g. the minimized structures with ideal approach for proton transfer. The only interactions found to differ considerably between the mutants was interactions between the two electrophilic groups in the active site, residues 13 and 95, and the O1 and O2 oxygens of the substrate (Figure 6.2). The clearest point to emerge from comparing changes in these energies is that every single-site mutation led to a decrease in favorable interactions between residue 95 and

TABLE 6.4
Changes in Interaction Energies of Covalent Enzyme-Substrate Complexes Upon Mutation[a]

Interaction[b]	Protein Model				
	Single Mutations from Wild-Type TIM			Double Mutations from Single-Site Mutant	
	S96–>P96	H95–>N95	E165–>D165	N95–>N95 S96 P96	D165–>D165 S96 P96
Residue 95 with All Other Atoms	↓	↑	↑	↓	↓
Residue 95 with O1 of Substrate	↓	↓	↓	↑	↑
Residue 95 with O2 of Substrate	↓	↓	–	↓	↓
Residue 13 with All Other Atoms	↓	↓	↑	↑	↓
Residue 13 with O1 of Substrate	↑	↑	↓	↓	↑
Residue 13 with O2 of Substrate	↓	↑	↓	↓	↓

[a] Arrows pointing upwards (↑) represent favorable changes upon mutation and conversely downward pointing arrows (↓) indicate unfavorable changes. No change upon mutation is denoted by (–)

[b] See Figure 6.2.

O1 or O2 (Table 6.4 and compare Figures 6.4A, B and C). In some cases the Lys 13-O1,O2 interactions were important for stabilization. Interactions between residue 95 and O1 became more favorable upon further mutation of the single mutant by replacement of Ser 96 by Pro (column for double mutants in Table 6.4).

Table 6.5 lists distances for particular interactions thought to be important for catalysis and those that exhibited large changes between different minimized structures after MD. For each single mutant the distance between O1 of the substrate and the hydrogen bond donor of residue 95 increased by approximately 0.5 Å. The hydrogen bond distance between His 95 and O2 was comparable in the different structures, though. Comparison of the double mutant to the relevant single mutant from which it was derived shows that the hydrogen bond distance between residue 95 and O1 decreased by approximately 1 Å upon introduction of Pro 96. In fact, these distances were lower than in the wild-type structure. The drastic change in His 95-substrate interactions was again, as with the noncovalent complexes, linked to the formation of an hydrogen bond between N_δ and the amide hydrogen of Glu 97 (Table 6.5 and Figure 6.4). For the Ser 96–>Pro change in the (E165D) mutant, the His 95-O2 distance remained roughly the same but the distance increased by over 2 Å in the (H95N) –> (H95N, S96P) case.

The distances between the closest hydrogen on N_ε of Lys 13 and O2 of the substrate increased by 0.2-0.4 Å going from wild-type to the single mutants (S96P) or (E165D) and the double mutant (E165D, S96P). This distance was about 0.4 Å shorter in the (H95N) mutant but increased to the wild-type distance when Pro was introduced. Asn 11 also interacted with O1 of the substrate in wild-type TIM, the (S96P) mutant, and the (H95N) mutant (Table 6.5). Asn 11 rotated slightly to interact with O2 of the substrate in the other mutants. Thus, all of the residues in the active site were highly correlated and the enzyme tolerated changes in amino acid sequence by stabilizing O1 and O2 of the substrate with other than wild-type interactions when necessary. (Compare Figures 6.4B and 6.4C to 6.4A.)

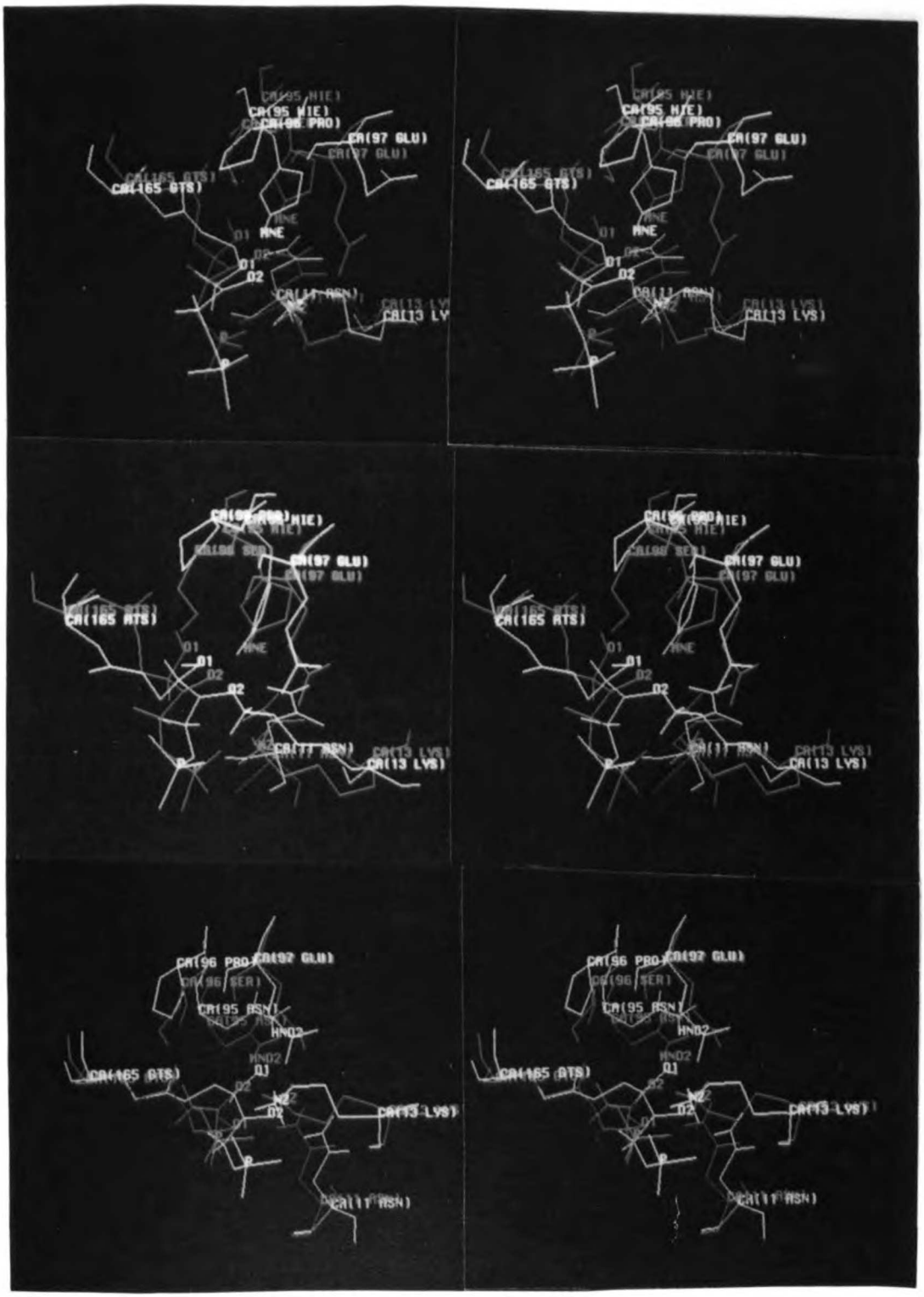


Figure 6.4. Stereoviews of covalent complexes: (A) Wild-type TIM in red, (S96P) mutant in yellow; (B) (E165D) mutant in red, (E165D, S96P) double mutant in yellow; and (C) (H95N) in red, (H95N, S96P) double mutant in yellow. All complexes have been subjected to 20 ps of MD and minimization.

TABLE 6.5
Distances (in Å) Between Active Site Residues in Minimized Wild-type and Mutant Triosephosphate Isomerases with Covalently Bound Substrate after 20 ps of Molecular Dynamics[a]

Distance[b]	TIM Model					
	Wild-type	Single-Site Mutants			Double Mutants	
	E165 H95 S96	E165 H95 S96P	E165 H95N S96	E165D H95 S96	E165 H95N S96P	E165D H95 S96P
11 HN1-Sub O1	1.69	1.71	1.70	3.48	3.73	4.13
11 HN1-Sub O2	2.48	3.86	3.60	2.41	1.89	2.04
13 HNE-Sub O2	1.97	2.35	1.63	2.18	1.99	2.30
13 HNE-Sub O4	2.11	1.62	3.99	1.82	3.62	2.92
13 HNE-Sub O5	1.65	5.42	1.59	1.63	2.62	3.58
13 HNE-97 O2	3.55	8.84	3.19	3.61	3.51	3.18
95 HNE(HN2)-Sub O1	2.39	2.85	2.84	2.85	1.86	1.83
95 HNE(HN2)-Sub O2	1.76	1.79	1.89	1.72	4.11	1.82
95 ND(O)-97 HN	2.38	1.99	3.39	3.30	3.39	1.93
165 O1-170 HN	8.15	9.25	7.40	9.75	8.28	1.75

[a] K_θ for the proton transfer angle during MD was 0 and 100 kcal/mole-rad^2 for minimization.

[b] See Figure 6.2 for atom name nomenclature. When there is more than one possible combination of atoms, the lowest distance is given.

Introduction of Pro at residue 96 into the wild-type or single mutants altered Lys 13 interactions with the phosphate group. In each case the distance for the Lys 13 interaction with O5 of the substrate (Figure 6.2) increased at least 1 Å (Table 6.5 and Figure 6.4) with the substitution of serine by proline. The changes were most dramatic for the (S96P) mutant. For this mutant the distance between Lys 13 and the carboxyl group of Glu 97, which is important for maintaining the Lys orientation, was over 5 Å greater than in the other structures.

All of the mutants retained roughly similar orientations of active site residues with the substrate, compared to wild-type, except the (E165D, S96P) double mutant. In this complex the substrate flipped around and the orientation of O1 and O2 of the substrate changed (Figure 6.4B). The extent of change can be seen by comparing the distances between the O1 oxygen of the carboxyl group of the catalytic base with the mainchain amide hydrogen of residue 170 (Table 6.5). In all of the structures except this particular double mutant, the distances were large and, in fact, increased when Pro was inserted. But for the (E165D) mutant this distance decreased 8 Å upon replacement of Ser 96 by Pro. The substrate conformation remained relatively extended during the simulations of the covalent complexes, unlike the kinking observed in the noncovalent complexes (Figure 6.4). The (H95N) mutant, however, again had O1, O2 and P of the substrate out of plane, but the substrate regained planarity with the introduction of proline (Figure 6.4C).

Dynamic Properties

After finding that there were changes in interactions for the isomerases, in the static minimized structures after MD, that appeared to be important for catalysis, we monitored the dihedral angles of residues 95 and 96 to determine whether the proline substitution affects the secondary structure. We then focussed on how the interactions stabilizing O1 and O2 of the substrate evolved during MD. The percentage of time that the hydrogen bonds between His 95 (or Asn) and Lys 13 and O1 and O2 of the substrate for the first 10 ps and second 10 ps of MD are given separately in Table 6.7 for each enzyme listed in

Table 6.1 as well as the two hypothetical Lys 13 mutants: (K13R) and (E165D,K13R).

Table 6.6 shows some properties of the dihedral angles of residues 95 and 96. Residue 95 is in an extended conformation (average ψ value was 140° in wild-type TIM) near the N-terminus of a short helix starting at residue 96 (average ϕ in wild-type TIM was -48°). The average ψ values for residue 95 were comparable in the different mutants with the exception of the (H95N) mutant, which was 20° higher than the other mutants (Table 6.6). The ϕ values of residue 96 were all uniformly helical, the largest deviation from the average wild-type angle was 6°, which is easily within the r.m.s. fluctuations in the angle during the simulation. That residue 96 was in the helical region of conformational space in all of the structures is interesting because Pro is generally considered an "helix breaker", although structural tolerance may be explained because residue 96 is at the end of the helix. In all of the mutant complexes the r.m.s. fluctuations in ψ_{95} were greater than in the wild-type structure. When Ser 96 was replaced by Pro in either of the single mutants (H95N, E165D), the fluctuations in ψ_{95} decreased, indicating that the motion of His 95 was restricted by the adjacent Pro. This type of effect was not observed in the fluctuations of ϕ_{96}. The fluctuations of ϕ_{96} were higher on average than those of ψ_{95}, which is interesting since residue 96 is involved in secondary structure while residue 95 is not. Apparently interactions with the substrate damp the motion of residue 95.

Since fluctuations of ψ_{95} were affected by the introduction of Pro 96, we computed the cross correlation coefficient for these angles (Table 6.6). These residues were most correlated in the wild-type structure. There is not a consistent pattern in the coefficients for the mutants. Most of the coefficients are negative, indicating that the residues moved in opposite directions in a type of crankshaft motion. This type of motion facilitates the maintenance of the orientations of active site residues. The smallest change in the correlation coefficients upon mutation of residues 95 or 96 was for the Asp 165 pseudorevertant---(E165D, S96P). The motions were fairly correlated in both the single and double mutant structures. Even for the other mutants that showed low correlation

TABLE 6.6
Average Dihedral Angles for Residues 95 and 96 in Wild-type and Mutant Triosephosphate Isomerases with Covalently Bound Substrate During Molecular Dynamics and Correlations between these Angles[a]

Protein Model	$\langle\psi_{95}\rangle$	$\langle\Delta\psi_{95}^2\rangle^{1/2}$	$\langle\phi_{96}\rangle$	$\langle\Delta\phi_{96}^2\rangle^{1/2}$	$\frac{\langle\Delta\psi_{95}\Delta\phi_{96}\rangle}{\langle\Delta\psi_{95}^2\rangle^{1/2}\langle\Delta\phi_{96}^2\rangle^{1/2}}$
E165 H95 S96	140.2	9.0	-48.5	13.2	-0.50
E165 H95 S96P	144.6	10.3	-35.7	16.6	-0.17
E165D H95 S96	139.7	10.8	-50.0	12.5	-0.41
E165D H95 S96P	134.6	9.8	-54.0	17.1	-0.37
E165 H95N S96	161.1	15.3	-54.6	20.8	0.07
E165 H95N S96P	144.4	13.6	-46.4	13.2	-0.18
E165 K13R H95 S96	141.8	9.1	-43.6	15.3	-0.33
E165D K13R H95 S96	143.3	11.3	-48.0	13.3	-0.38

[a] All angles are given in degrees.

between residues 95 and 96, the average dihedral angles, and therefore gross structual features, were maintained.

Table 6.7 shows the percentage of time that the hydrogen bonds between His 95 and Lys 13 and O1 and O2 of the substrate were intact during MD. Hydrogen bonds were considered intact if the donor-acceptor distances were less than 3 Å. The population of the hydrogen bond between residue 95 and O1 of the substrate decreased substantially upon mutation of a single residue, and the change was most striking for the (E165D) mutant (the percentage dropped by 50%). These hydrogen bonds became more stable with time, however. The percentage of time the residue 95-O2 hydrogen bond was intact was maintained at very high levels in all of the single mutants. The mutants recovered from the changes in sequence in less than 10 ps (in most cases reorganization, if necessary was complete within 2-6 ps) to regain favorable wild-type-like interactions.

The Lys 13-O2 interaction changed some with mutation (Table 6.7). This interaction was essentially unchanged with replacement of Ser 96 by Pro in the wild-type enzyme, decreased markedly (>50%) for the Glu 165 –> Asp mutation, and increased to 100% with the His 95 –> Asn mutation. The hydrogen bond between O2 and residue 13 was lost entirely upon replacement of Lys 13 by Arg in the wild-type enzyme.

The hydrogen bonds in the active site of the (E165D) mutant were improved upon introduction of Pro, e.g. the percentage of time the His 95-O1 hydrogen bond was intact increased 50% (Table 6.7). The His 95-O1 hydrogen bond also improved when Lys 13 was replaced by Arg in the (E165D) mutant. Similar, but even more striking, effects were observed for the (H95N) –> (H95N, S96P) mutation. Upon insertion of Pro into this single mutant, the population of the Asn 95-O1 hydrogen bond increased and Asn 95 interacted exclusively with O1; the percentage of time the hydrogen bond to O2 was intact dropped from 100 to 0%. Lys 13 was in a position to stabilize O2 in both the single and double mutants.

TABLE 6.7
Percentage of Time Hydrogen Bonds Between Active Site Residues and Substrate Were Intact in Wild-Type and Mutant Triosephosphate Isomerases with Covalently Bound Substrate During Molecular Dynamics[a]

Hydrogen[b] Bond	Protein Model and Time of MD (ps)															
	Wild-type		Single-Site Mutants								Double Mutants					
	E165 H95 S96 K13		E165 H95 S96P K13		E165 H95N S96 K13		E165D H95 S96 K13		E165 H95 S96 K13R		E165 H95N S96P K13		E165D H95 S96P K13		E165D H95 S96 K13R	
	0-10	10-20	0-10	10-20	0-10	10-20	0-10	10-20	0-10	10-20	0-10	10-20	0-10	10-20	0-10	10-20
95HNE(HN2)/SubO1	92	94	4	64	72	70	34	44	34	52	95	100	94	94	36	86
95HNE(HN2)/SubO2	100	100	26	98	62	100	100	100	100	100	10	0	92	100	96	100
13HN2/SubO2	98	94	90	90	100	100	46	42	18	0	100	100	94	98	0	0

[a] Hydrogen bonds were considered intact if the acceptor-donor distance was less than 3 Å.

[b] See Figure 6.2 for atom name nomenclature.

In the single mutants, interactions with O1 and O2 were disrupted compared to wild-type TIM. In the case of the (E165D, S96P) double mutant, introduction of Pro maximized interactions between His 95 and O1. The other pseudorevertant (H95N, S96P) showed separate residues stabilizing O1 and O2 of the substrate. Another way of viewing this is by the angles of approach of the hydrogen-bonding group to O1 and O2. Table 6.8 shows the angles between O1 and C1 of the substrate and the hydrogen bond donor of residue 95 and the corresponding angle at O2 for the different enzymes. The property of interest is the difference between these angles. For the wild-type enzyme it can be seen that the difference is relatively small (13.42°). In the (S96P) and (E165D) mutants the difference increased by approximately 20° relative to wild-type TIM. Replacement of Lys 13 by Arg in wild-type TIM caused the average difference in angles to increase by 10°. When Pro was introduced into the (E165D) mutant this difference decreased to below the value of the wild-type enzyme. Hence, MD with Pro allowed His 95 to maximize its interactions with O1 and O2 simultaneously. Replacement of Lys 13 by Arg in the (E165D) single mutant also led to a decrease in the difference between the angles, although the effect was not as striking as with the introduction of Pro into this mutant. The (H95N) mutant showed a low difference between these angles which increased upon addition of Pro, unlike the case above. For this mutant it appeared that the optimal alignment in the active site involved residue 95 interacting with O1 and Lys 13 with O2.

Some general features emerge in these mutants that correlate with changes in catalytic activity upon mutation, but we lack a structural, or mechanistic, explanation of how Pro exerts its effect on neighboring residues. The motions of residues in the active site were so highly correlated that it has been difficult to arrive at an unique step-by-step description of proline's role in altering binding and catalytic properties.

DISCUSSION

We performed molecular dynamics simulations of wild-type TIM and five mutant

TABLE 6.8
Average Orientations Between Substrate Atoms and Residue 95 in Wild-type and Mutant Triosephosphate Isomerases with Covalently Bound Substrate During Molecular Dynamics[a]

Protein Model	$\langle\theta_1\rangle$	$\langle\theta_2\rangle$	$\langle\theta_1-\theta_2\rangle$
E165 H95 S96	51.0	37.6	13.4
E165 H95 S96P	81.0	48.0	33.0
E165D H95 S96	65.4	29.2	36.1
E165D H95 S96P	48.9	45.2	3.7
E165 H95N S96	43.0	47.3	-4.3
E165 H95N S96P	27.1	78.7	-51.6
E165 K13R H95 S96	72.4	50.0	22.4
E165D K13R H95 S96	61.8	40.8	21.0

[a] θ_1 is the angle between SubO1, SubC1, 95HNE(HN2). θ_2 is the angle between SubO2, SubC2, 95HNE(HN2). See Figure 6.2. All angles are given in degrees.

enzymes that have been characterized experimentally. All of the mutants exhibit lower catalytic activity than wild-type TIM (Hermes et al., 1987). Two of these mutants give rise to second-site suppressor mutants with increased catalytic activity or binding affinity, with regard to the substrate dihydroxyacetone phosphate (DHAP). In both cases, the changes in binding and catalysis are the result of the replacement of Ser 96 by Pro. We found in our simulations of the covalent complexes (models for the transition state for enolization) that stabilization of O1 and O2 of the substrate by the nearby electrophilic residues Lys 13 and His 95 (Figure 6.2) diminished upon mutation of a single residue in wild-type TIM and that the introduction of Pro 96 into a single-site mutant improved these interactions through different cooperative changes. These same interactions, and others, also led to improved protein-substrate interactions in the noncovalent complexes.

We did not see a correlation between total energy and binding affinity or catalytic activity in our MD simulations. Catalytic activity and the degree of interaction at O1 and O2 of the substrate were correlated, however. Every single-site mutation disrupted these interactions with the substrate. The pseudorevertants showed improved hydrogen bonding to the substrate relative to the single mutants from which they were derived. There were essentially two different mechanisms of transition state stabilization. The first mechanism involved strong interactions between His 95 and O1 and O2 of the substrate with aid from Lys 13 in stabilizing O2. The (E165D, S96P) double mutant was an example of this type of stabilization and in a sense mimicked the wild-type interactions. The (E165D, S96P) mutant showed better stabilization of O1 than the wild-type, which may indicate that the optimal arrangement is for His 95 to stabilize O1 and O2, with a slight emphasis on O2. (The distances for wild-type TIM in Table 6.5 also support this idea.) Proline's ability to improve the orientation of His 95 appeared to be due to a slight change in the backbone that improved the hydrogen bond between the mainchain amide hydrogen of Glu 97 and N_δ of His 95. The second mechanism for improving substrate stabilization was for separate residues to interact with O1 and O2. Comparison of the

(H95N) mutant to its pseudorevertant (H95N, S96P) revealed improved interactions at O1 of the substrate with Pro in the place of Ser 96. After reorganization of the active site residues, Asn 95 interacted exclusively with O1 and Lys 13 stabilized O2. Even though transition state stabilization was more effective in the pseudorevertants compared to the single mutants, the adopted orientations and interactions were worse overall than in wild-type TIM, although a single interaction may have appeared to be better in a mutant.

In the noncovalent complexes we observed the same type of substrate stabilizing interactions as those described for the transition state models. The (H95N) mutant derived its tight binding of substrate from strong interactions with the phosphate group. The (H95N, S96P) double mutant also exhibited strong interactions with the phosphate group in addition to improved interactions at O1 and O2 of the substrate. Overall, substrate affinity appeared to be very similar in the two structures with the double mutant interacting slightly more strongly with the substrate. This comparison, although qualitative in nature, is consistent with with the experimental findings (Table 6.1). In the case of the (E165D) and (E165D, S96P) mutants, substrate binding again appeared to be tighter in the double mutant, which is consistent with the experimental data (Table 6.1). All of the mutants except one---(E165D)---demonstrated tighter binding than wild-type, showing the importance of ground state destabilization to catalysis.

To compare our results to the kinetic data we must consider the difference between the noncovalent and covalent structures for each protein. For the (H95N) mutants the effect of adding a proline was to greatly improve interactions in the transition state structure compared to the single mutant. There were only minor differences in the noncovalent complexes. These results are consistent with the experimental findings that the introduction of Pro into the (H95N) mutant predominantly affects catalysis and not binding (Table 6.1). With the (E165D) mutant interactions were improved in both the noncovalent and covalent complexes when Ser 96 was replaced by Pro. Although these comparisons are only qualitative, they suggest that the introduction of Pro into the (E165D)

mutant affects binding with little effect on catalysis. This is, in fact, what is observed experimentally with DHAP (Table 6.1).

When the isomerization catalyzed by TIM proceeds in the direction of GAP to DHAP (Figure 6.1), the differences between the single mutants and their corresponding pseudorevertants are manifested primarily in the transition states for enolization and not in the noncovalent GAP-isomerase complexes (Blacklow and Knowles, 1990). Experimentally, the introduction of Pro into the single mutants leads to improvements in catalysis. Although, we have not actually simulated the noncovalent and covalent complexes with GAP, our results are qualitatively consistent with the experimental findings. In our models of the transition state we found greatly improved interactions between His or Asn 95 and O1 of the substrate with the introduction of Pro. The enediolate intermediate from GAP would have the negative charge on O1. So, even though O1 is protonated in our model, we observed improved stabilization of this group with Pro in position 96, which would presumably be even more favourable when O1 is more negatively charged.

There is another interesting feature of the (H95N) mutant. We found a number of interactions in both the noncovalent and covalent complexes that may, in concert, lead to elimination of the phosphate group: tight binding of the phosphate portion of the substrate preferentially over the catalytically important groups; O1, O2 and P of the substrate were out of plane; and strong interactions with O1H of the substrate. The elimination reaction producing methyl glyoxal and inorganic phosphate is favored, for stereoelectronic reasons, when O1, O2 and P of the substrate are out of plane; whereas, isomerization is favored when these atoms are in plane (Richard, 1984). Elimination may begin with intramolecular abstraction of O1H of the substrate (Rose, 1981). Experimentally, the (H95N) mutant catalizes the elimination reaction approximately one-third of the time (Blacklow and Knowles, 1990). In our (H95N, S96P) transition state model, the substrate became more planar, and although phosphate binding was tight there were strong interactions with other portions of the substrate, too.

In our simulations, TIM appeared to be highly degenerate, in the sense that there were a number of residues in the active site that were able to, at least partially, compensate for interactions lost upon mutation. These results are consistent with the speculations of Petsko and co-workers regarding the evolution of catalytic efficiency (Alber et al., 1987). They contend that TIM has been overdesigned as a safety feature against adventitious mutations.

Crystal structures are not yet available for the various mutants but Alber et al. (1987) appear to have a low resolution structure (3 Å) for a related mutant in yeast---His 95 → Gln. They find that Gln 95 rotates slightly to interact with Glu 97 and that Lys 13 moves to a position midway between O1 and O2. Alagona et al. (1984) identified a similar hydrogen bond between Gln 95 and Glu 165 in a simulation that preceded the X-ray analysis. We found both types of interactions in our simulations of the (H95N) mutant. In the noncovalent comples, Asn 95 interacted with Glu 165 (Figure 6.3d). This interaction was disrupted upon replacement of Ser 96 by Pro to yield (H95N, S96P) (Figure 6.3e). This type of interaction constitutes a barrier to catalysis as discussed by Alagona et al. (1984). In the covalent complex, however, Asn 95 made an hydrogen bond with Glu 97 in the single mutant, which was disrupted upon introduction of Pro. Although Asn 95 formed an hydrogen bond in the single mutant with Glu 97, the other amide hydrogen interacted with O1 and O2. Thus, Asn 95 still participated in stabilization of the transition state model but less effectively than its pseudorevertant. The opposite was found with the noncovalent complexes: (H95N, S96P) contained the hydrogen bond, however, the (H95N) mutant did not. Given that Asn has another hydrogen bond donor interacting strongly with the substrate, binding is not compromised severely by this interaction.

In an earlier study we found that the interaction of Lys 13 with O2 of the substrate was less effective in the (E165D) mutant compared to wild-type TIM and suggested that this could be the reason for the drop in activity of the mutant (see Chapter 5). Further,

we suggested that replacement of Lys 13 by Arg might improve stabilization of the transition state for enolization of the Asp 165 mutant. We did not, however, find this to be the case in the simulations presented here. The replacement of Lys 13 by Arg in the (E165D) enzyme substantially improved interactions between His 95 and O1 of the substrate, while interactions between residue 13 and O2 of the substrate were lost. These were, in fact, the interactions that we predicted would improve. This simulation points out how coupled the active site residues are and how simple-minded suggestions of amino replacements may be misleading. In fact, in an attempt to engineer improved enzymes, who would have replaced Ser 96 by Pro? The Lys 13 mutants are hypothetical at this point and await experimental confirmation of their inadequacy.

Based on our results of the study described here, Asn 11 may be a good choice for mutation. We found that Asn 11 performed a variety of functions, from stabilizing O1 and O2 of the substrate to interacting with the phosphate group. However, we did not see a pattern to these interactions that correlated with activity. Mutation of Asn 11 to a non-polar residue may shed light on the role of this residue in catalysis and the cooperativity between it and other active site residues.

Some of the simulations described here have employed models with substrate covalently attached to the enzymes to mimic the transition state for enolization. The actual nature of the intermediate---enediol or enediolate---has not been definitively established. Rose (1981) argues in favor of the enediol. However, the enediolate probably plays an important role at some stage of the reaction since phosphoglycolate (Wolfenden, 1969) and phosphoglycolohydroxamate (PGH) (Collins, 1974) are effective in inhibiting TIM. Phosphoglycolate binds tightly to TIM as the trianion (Campbell et al., 1978, 1979). PGH also, presumably, has the charge configuration of the putative enediolate transition state, and the binding of PGH to TIM is substantially tighter than substrate binding (Alber et al., 1987). At neutral pH PGH may be protonated in bulk solution (the phosphate group is dianionic) (Nickbarg et al., 1988); however, the effect of the protein

environment on the protonation state of PGH is not known. The models in this study represent an enediolate. If the rate limiting transition state of the reaction involves formation of the enediol, then our models are inappropriate for interpreting the experimental data. If the enediolate is important and is rapidly protonated (compared to formation of the enediolate) by solvent, then our models are relevant. Based on isotopic exchange experiments, Maister et al. (1976) suggest that the enediol intermediate rapidly exchanges with solvent. Given that for this to occur the active site must be at least transiently accessible to solvent, water could easily protonate the enediolate in a process with a low activation energy (Alagona et al., 1984). The calculations of Alagona et al. (1984) suggest that the enzyme sufficiently stabilizes the enediolate such that even if the enediol is the ultimate intermediate, proton abstraction from DHAP does not necessarily need to occur concomitantly with protonation of the enediolate.

Initially we performed simulations of the covalent complexes with a low force constant on the angle of proton transfer. We found that the angle collapsed in these simulations. We then took the final structures after MD and imposed transition state geometry by increasing the force constant. The advantage of using this procedure over performing MD with the high force constant was that the active site was more flexible during dynamics. Since it appears that Pro 96 plays an indirect role in determining catalytic activity and binding by altering the orientations of the active site residues, it is important that the models allow sufficient movement in the active site. With the higher force constants, the substrate was quite rigid and larger scale motion in the active site did not occur. The types of motion that we observed during MD with K_θ=0 kcal/mole-rad^2 might also occur with K_θ=100 if the simulation were run for a longer period of time. In reality, an enzyme binds a substrate and the alignment of enzyme and substrate atoms occurs, for the most part, prior to formation of the transition state. To mimic this effect in a simulation one would perform MD of the noncovalent enzyme-substrate complex and then form the transition state. By using K_θ=0 kcal/mole-rad^2 reorganization of the active site residues

occurred, which mimics what might occur in noncovalent complexes. But, with this approach the sampling of catalytically nonproductive orientations is reduced by using structures near the transition state and therefore in the proper region of conformational space for looking at effects on catalysis. In imposing productive transition state geometry with minimization the orientations generated during MD were maintained.

The simulations described here were of short duration and one wonders how general the results are. Related issues are: how dependent the results are on the starting structures and how many simulations need to be performed before we can be assured that the conclusions are generally valid. These concerns can be raised for almost any MD simulation, but one can argue against a large number of long simulations when, say, a single short simulation can reproduce a number of independent, experimental parameters. Given the time scale of biochemical processes and experimental techniques for detection (usually nsec-sec) compared to the time scale of most MD simulations (generally ps), this is almost never achieved. We did not perform long simulations nor did we exhaustively test different starting structures, but our results are qualitatively consistent with the experimental data and various predictions are experimentally testable. When crystal structures of the mutants with substrate and with transition state inhibitors are available, specific interactions can be compared between our models and the experimental structures.

The simulations are flawed in another way, by the absence of water molecules and the use of a relatively poorly resolved crystal structure (2.5 Å), which is the best structure available. We have attempted to account for the lack of solvent by using a linear distance dependent dielectric constant ($\varepsilon=r$), which partially screens long-range electrostatic interactions. (For a more thorough discussion of the relative merits of this function and others used in force fields, see Chapter 3) For the types of cut-offs that we used the average dielectric constant was approximately 5. This is the magnitude of the dielectric constant that one expects in the protein interior (Pethig, 1979) and we focussed on the active site residues, which are in the interior of the protein. But, given recent experimental

work (Rees, 1980; Russell and Fersht, 1987; Russell et al., 1987) that suggests that the dielectric constant of a protein can be an order of magnitude larger at moderate distances (circa 10 Å), the dielectric constant of a protein in water and how to model this effect is still a matter of debate. (See Harvey, 1989 for further discussion.)

There is a considerable advantage to performing simulations with macroscopic dielectric functions as opposed to full inclusion of solvent, which relates to the issue of simulation time versus the time scale for molecular events. Dielectric functions allow the calculations to proceed much faster than when solvent is present (of order 10-100 times faster). Also, since electrostatic interactions are exaggerated with $\varepsilon=r$, an event may occur that would take much longer in water. In another system we showed that one gets much better sampling of conformational space with macroscopic dielectric functions compared to simulations with water. As mentioned above, we have presumably limited sampling of catalytically irrelevant orientations by using structures near the transition state geometry to compare with the experimental kinetic results instead of imposing transition state geometry on the noncovalent structure.

Despite the various limitations imposed by our approximations and the lack of structural experimental data, our results are qualitatively consistent with both the binding and kinetic data. We have pointed out interactions that appear to be important for effective catalysis and strong binding of substrate. Confirmation of specific proposed interactions must await detailed X-ray crystal structures of the mutants with DHAP and transition state inhibitors. It is encouraging that our results are consistent with the data of the Pro 96 mutants and suggest possible reasons for the changes in activity and binding, since the effect of this mutation could not have been predicted *a priori*. Our simulations reproduce the highly cooperative nature of the interactions in the active site and suggest that this approach may be useful for identifying particularly promising sites for mutation.

CHAPTER 7: Semiempirical Molecular Orbital Studies of Bond Cleavage Catalyzed by Trypsin

One of the biggest challenges in biochemistry is to understand the molecular basis of enzyme catalysis. The serine protease family of endopeptidases provides one of the best systems for addressing this question. There is more direct evidence about the mechanism of catalysis and structures of intermediates in the reaction of serine proteases than any other enzyme or enzyme family (Fersht, 1985). Serine proteases have also been the subject of numerous theoretical studies.

All serine proteases contain the catalytic triad---Asp...His...Ser (Kraut, 1977). The purpose of this arrangement of residues in the active site is presumably to make the serine sufficiently nucleophilic that it attacks the carbonyl carbon of the amide or ester substrate (Figure 7.1), thus beginning the catalytic process. Serine's attack on the substrate results in formation of a tetrahedral intermediate (TET_1, Figure 7.1), which breaks down to yield the acylenzyme intermediate, EA. The second part of the reaction begins at this point with hydrolysis of the acylenzyme intermediate, resulting in formation of the second tetrahedral intermediate, TET_2. Collapse of this intermediate yields the enzyme-product complex.

All serine proteases also contain an oxyanion hole, which is generally made up of two backbone amide hydrogens (Kraut, 1977). The main purposes of the oxyanion hole are to stabilize the oxyanion of the tetrahedral intermediates and to stabilize the developing negative charge en route to the tetrahedral conformation.

Many theoretical studies indicate that the proton is transferred from the serine to the histidine prior to attack of the substrate, such that the transferred proton is not a direct participant in the bond making and breaking events occurring in the transition state (Scheiner and Lipscomb, 1976; Warshel and Russell, 1986). This sequence of steps is inconsistent with the kinetic isotope effects seen experimentally that suggeest that there is protonic bridging in the transition state (Schowen, 1988). Many of these theoretical

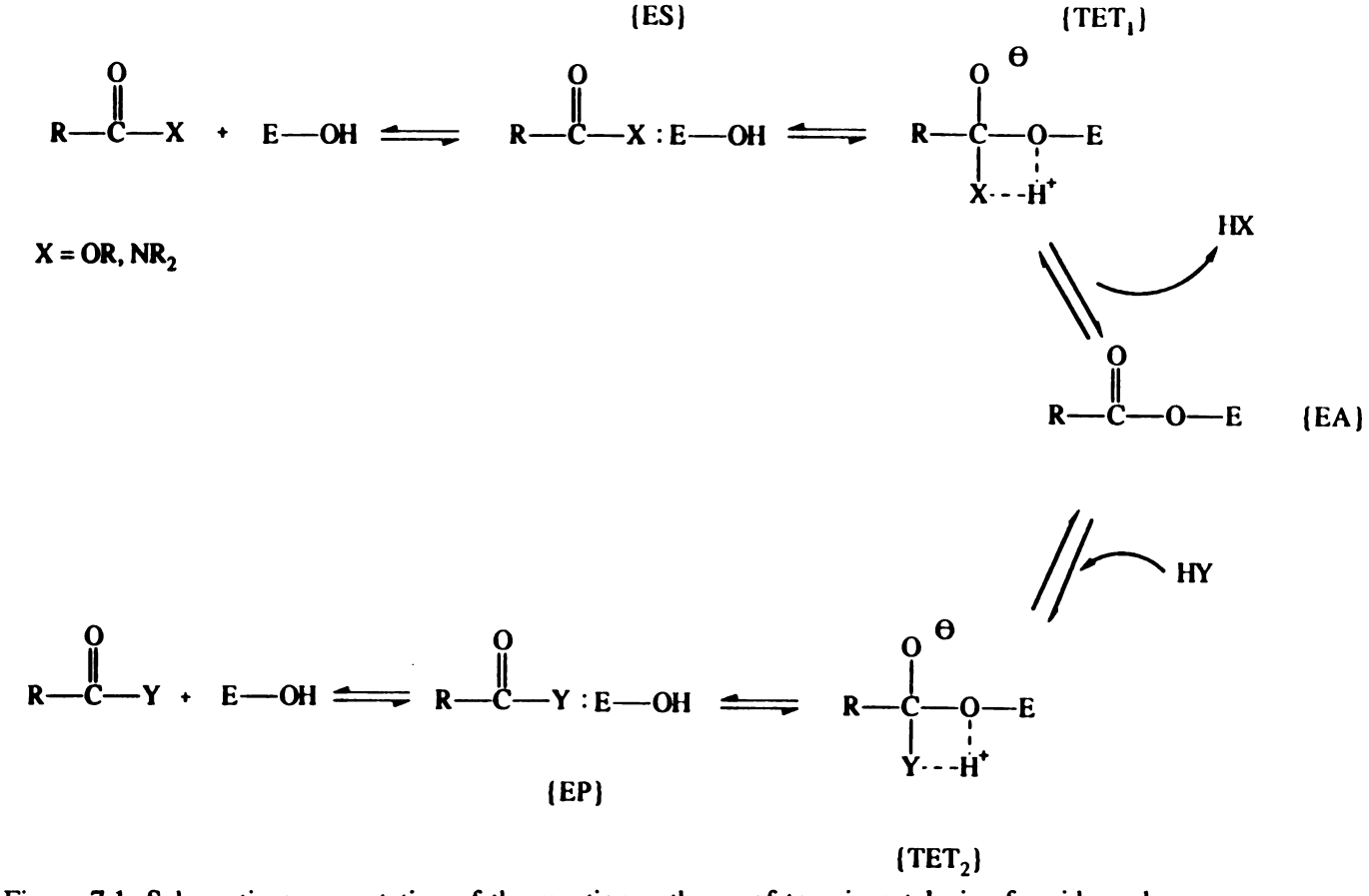


Figure 7.1: Schematic representation of the reaction pathway of trypsin catalysis of amide and ester bonds. E-OH represents the enzyme's hydroxyl group on Ser 177. The following abbreviated are used: {ES} denotes the Michaelis comples; {TET_1} represents the first tetrahedar intermediate; {EA} represents the acylenzyme intermediate; HY represents a water molecule in our scheme; {TET^2} is the second tetrahedral intermediate following water attack of the acylenzyme intermediate; and {EP} denotes the enzyme-product complex.

studies also support the double proton transfer mechanism, whereby the Asp abstracts a proton from the histidine following transfer from the serine (ex. Dewar and Storch, 1985; Scheiner et al., 1975; Umeyama et al., 1973). This mechanism was first suggested to be wrong on the basis of theoretical studies by Hayes and Kollman (1979). This mechanism has also been refuted experimentally (Kossiakoff and Spencer, 1981; reviewed by Steitz and Shulman, 1982). Most of the theoretical studies of serine protease catalysis have only dealt with the first part of the reaction pathway---acylation (Figure 7.1) (ex. Warshel and Russell, 1986). In addition, many of the previous studies used severely truncated forms of the catalytic triad (or have completely left portions out, such as the histidine, Dewar and Storch, 1985). For these reasons we decided to explore the entire trypsin-catalyzed pathway in detail with the functional forms of all of the catalytic triad residues present.

This paper presents the results of semiempirical molecular orbital calculations of model compounds pertinent to the trypsin-catalyzed reaction and calculations of the actual reaction pathway. We employed two molecular models initally, AM1 (Dewar et al., 1985) and PM3 (Stewart, 1989). We found AM1 to be unsatisfactory for the model compounds and further studies were performed with PM3. The calculations of the reaction pathway included slightly truncated versions of the residues making up the catalytic triad extracted from the trypsin X-ray crystal structure (Figure 7.2). The oxyanion hole was represented by two water molecules in the positions of the mainchain groups making up the oxyanion hole in the crystal structure. A tripeptide substrate was modelled into trypsin's active site, from which the pertinent atoms were extracted for the quantum mechanical calculations. Acylation was investigated using both amide and ester substrates.

We found that formation of the first tetrahedral intermediate was the rate-limiting step in the reaction pathway with both substrates (Figure 7.3). The lowest energy path for formation of the tetrahedral intermediate was for serine to approach the substrate followed by coupled heavy atom movement and proton transfer to complete the reaction;

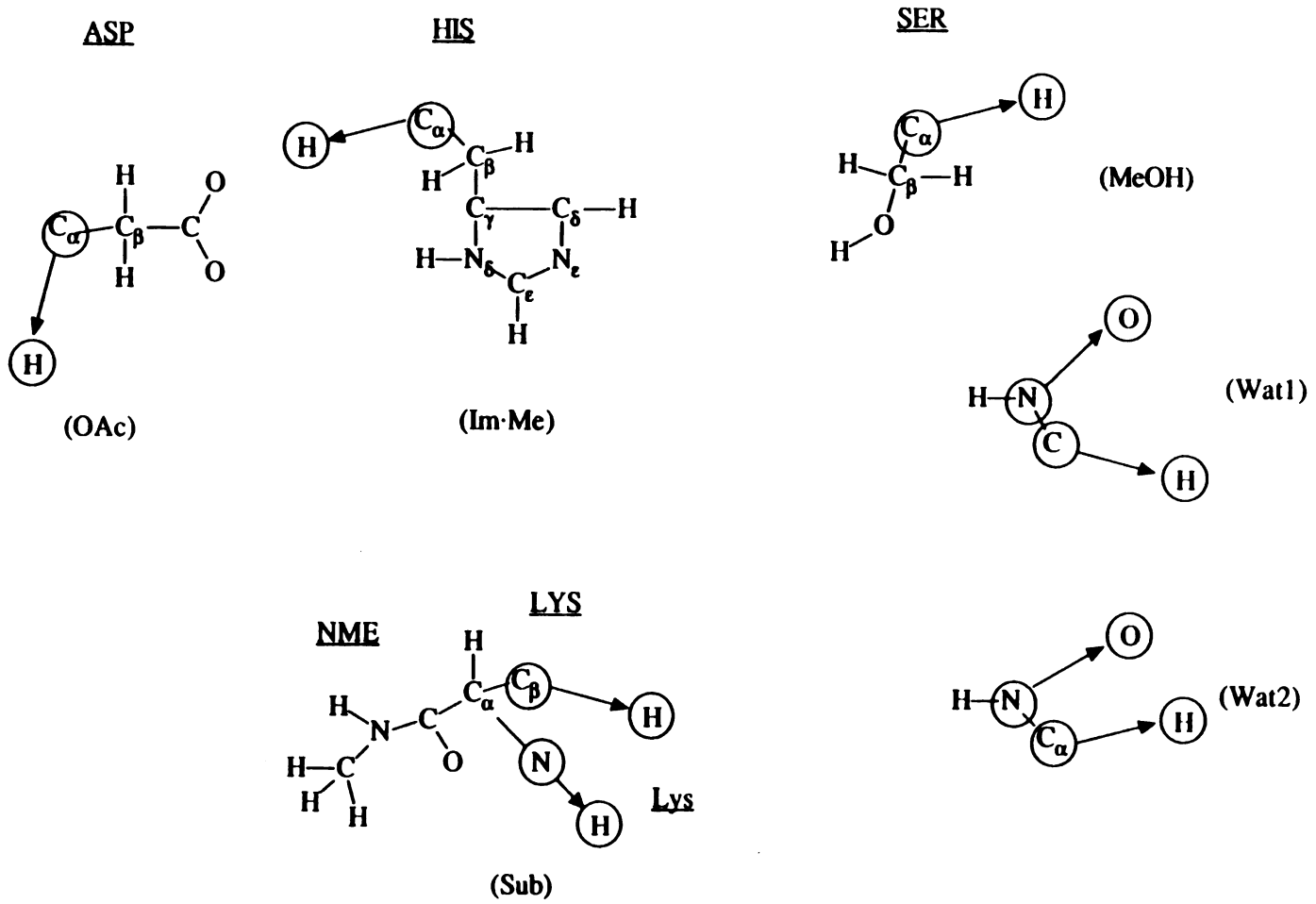


Figure 7.2: Groups important for catalysis by trypsin. The atom changes illustrated here are for adapting the residues from the refined crystal structure of trypsin to reduced representation (names in parenthesis) for the quantum mechanical calculations. See text for details.

we believe our description of this sequence of events to be novel. When one considers the effect of fluctuations in the enzyme on the process, deacylation becomes the rate-limiting step in ester hydrolysis. The oxyanion hole and Asp greatly stabilized the tetrahedral intermediates, and presumably their transition state structures. Interactions between the substrate and oxyanion hole were stronger, however, with the amide substrate than the ester. Our results are discussed in light of the available experimental data.

METHODS

Preparation of the Serine Protease Model

We used the bovine trypsin crystal structure of Chambers and Stroud (1977, 3ptp, 1.5 Å resolution). We added the substrate Ace-Phe-Val-Lys-Nme using the computer graphics package MIDAS (Jarvis et al., 1985). This tripeptide was used because it is catalytically very efficient (Pozsgay et al., 1981) and because it was used in an earlier study from this research group (Weiner et al., 1986). Polar and steric interactions were considered in aligning the substrate in the binding pocket. Counterions were positioned near free charged surface residues of the protein (thirteen chlorine ions and three sodium ions). In addition to the counterions, the model contained the internal calcium ion identified in the crystal structure. The overall charge of the system was -1.

It was necessary to refine the coordinates prior to extracting the catalytically important residues for the quantum mechanical calculations. The molecular mechanics program AMBER version 3.0 (Singh et al., 1986) was used for the calculations described below. Standard all-atom parameters were used for the catalytically important residues (His 40, Asp 84, Ser 177, and the capped, amidated Lys of the substrate) (Weiner et al., 1986). Standard united atom parameters were used for all other residues (Weiner et al., 1984). The energy of the system was minimized to a root-mean-square (r.m.s.) energy gradient of 0.1 kcal/mole-$Å^2$ to remove bad contacts. Then, a sphere of TIP3P water molecules (Jorgensen et al., 1983) extending 20 Å in any direction from the hydroxyl oxygen of Ser 177 was generated; this resulted in the introduction of 384 water

molecules. Following the addition of the water molecules the energy of the system was minimized. For this minimization, only those residues within 15 Å of Ser 177 were allowed to move. A 12 Å nonbonded cut-off was used. Minimization was carried out using the steepest descent method for 200 steps, followed by conjugate gradient minimization to a r.m.s. energy gradient of 0.1 kcal/mole-$Å^2$.

Construction of the Active Site Model

After refining the coordinates of trypsin, the cartesian coordinates for His 40, Asp 84, Gly 175, Ser 177 and N-methylated Lys of the substrate were extracted from the rest of the system for the quantum mechanical calculations. Abbreviated forms of these residues were used (Figure 7.2). To make the conversion to the truncated residues, some of the initial hydrogen positions were taken from the coordinates of the appropriate carbon or nitrogen, thus terminating the chain (the circled atoms in Figure 7.2 were replaced by hydrogens or oxygens). Instead of including the mainchain atoms that make up the oxyanion hole (residues 175 and 177), we positioned two water molecules in their positions to decrease the size of the system.

The calculations reported below were carried out using either the AM1 (Dewar et al., 1985) or PM3 (Stewart, 1989) molecular models as implemented within a modified version of the AMPAC program (Merz and Besler, 1989). All of the reactions were followed by using the reaction coordinate method (Dewar and Kirschner, 1971) with an internuclear distance or dihedral angle specified as the reaction coordinate. The active site model described above, and shown schematically in Figure 7.2, contained a number of long bonds because of the atom type swaps made to truncate the system. Therefore, the positions of the swapped atoms were optimized by minimizing their energy while "freezing" the geometry of all other groups (ie. by turning off the optimization flag). To further refine the model prior to reaction coordinate calculations, the distance between the hydroxyl oxygen of the Ser 177 mimic (MeOH) and the carbonyl carbon of the substrate was optimized while fixing all other parameters. Then, the entire system was

optimized followed by further high precision optimization (ie. by use of the PRECISE flag). The resulting geometries (ES of Figure 7.3) were used for the model calculations and reaction coordinate studies described below.

Model Calculations

We began this study by calculating heats of formation, proton affinities, and hydrogen bond strengths using the AM1 molecular model to assess its ability to reproduce various hydrogen bonding properties of relevant models for the trypsin-catalyzed reaction. The geometries for the various models were taken from the fully optimized structure of the active site residues, ES. Because AM1 did not satisfactorily reproduce the structural features of the hydrogen bonds in the model structures, we investigated these same model systems using the new PM3 model, which was designed to improve hydrogen bond properties. For all of the calculations, the geometries of the models were fully optimized to determine the heats of formation. PM3 best represented the hydrogen bonding structures in the trypsin active site and was employed for all further calculations.

Reaction Coordinate Calculations

To simulate the reaction shown schematically in Figure 7.3, slightly different procedures were used for the different steps. For the first step, the reaction coordinate method (Dewar and Kirschner, 1971) was used with the proton transfer distance between the hydroxyl proton of Ser 177 to the free nitrogen of His 40 specified as the reaction coordinate (decreasing from 1.77 to 1.00 Å). Full geometry optimization was accomplished at each point along the reaction path, with the exception of the distance between the hydroxyl oxygen of Ser 177 and the carbonyl carbon of the substrate which was frozen (not optimized) at four different distances (3.3, 2.7, 2.1, and 1.5 Å). So, the proton transfer reaction was carried out at each of the C-O distances. The grid search approach was used to address both the proton transfer step and serine's attack on the substrate to avoid forcing the order of the steps.

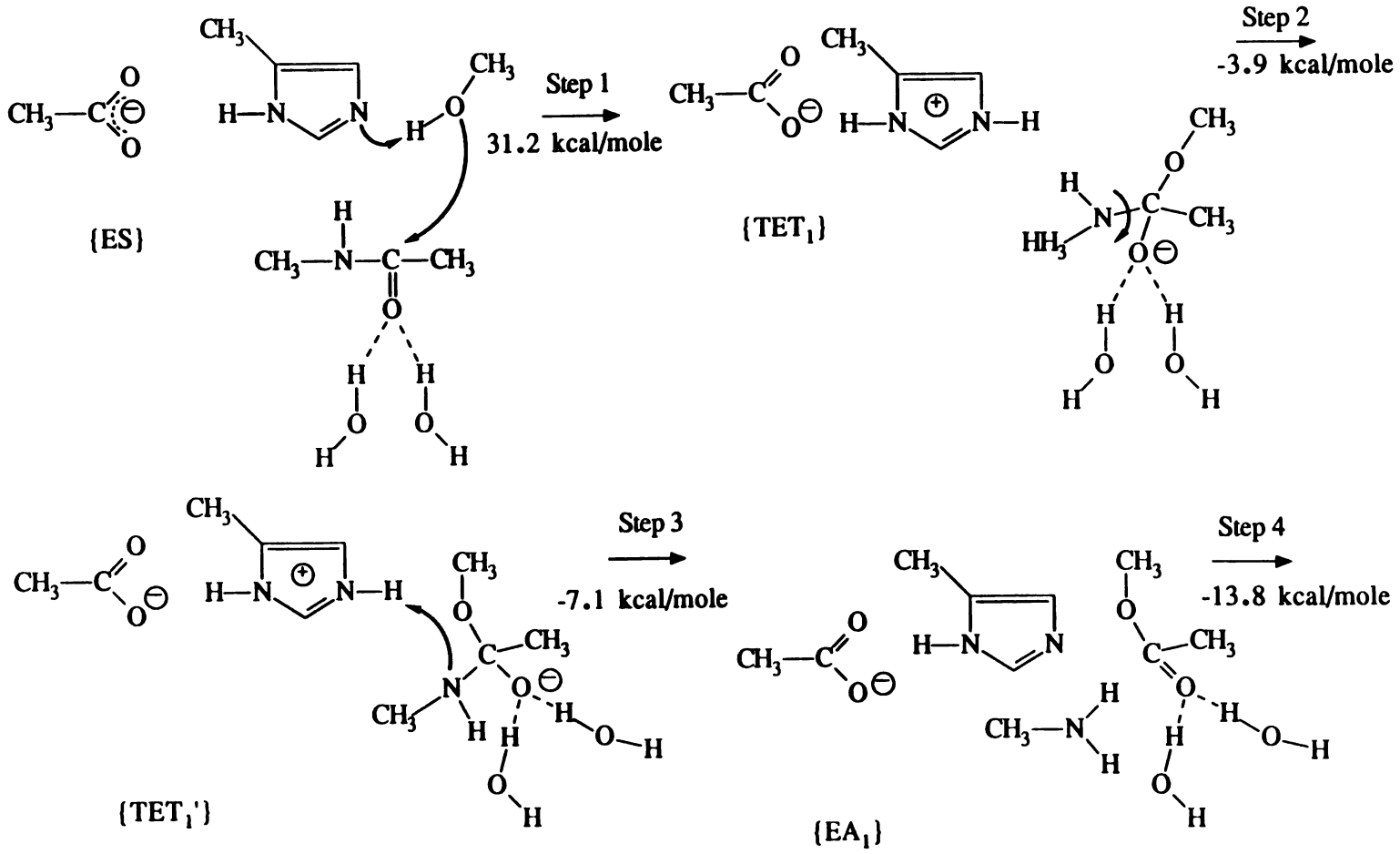
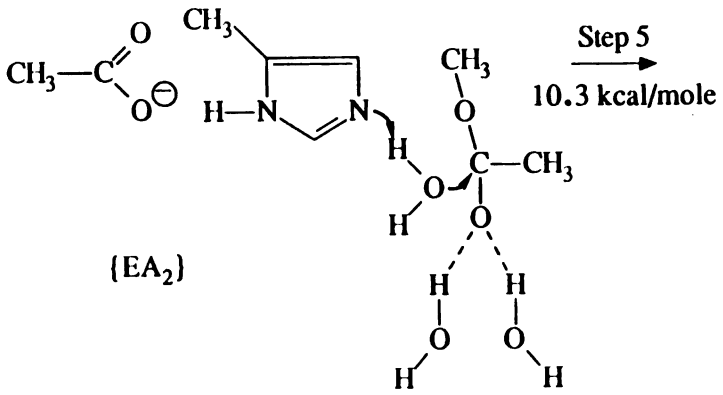
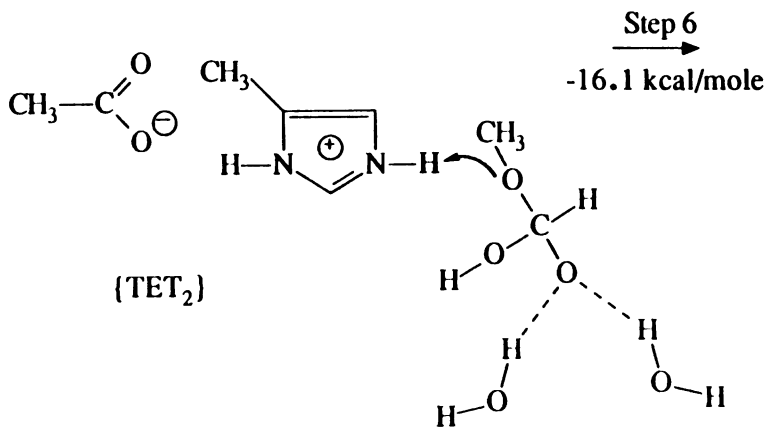


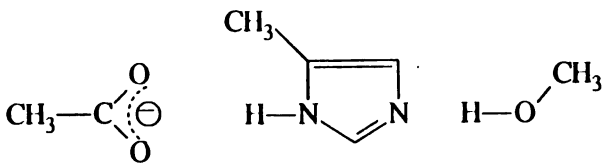
Figure 7.3: Steps along the reaction pathway of trypsin-catalyzed peptide bond cleavage. The differences in energy given for each step are between energies for the fully optimized systems.

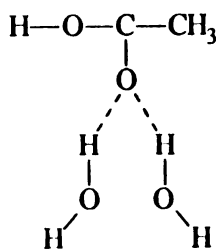


{EA_2}



{TET_2}





{EP}

Figure 7.3: continued

The second step of the reaction in Figure 7.3 with the amide substrate involved inversion of the substrate's nitrogen. The inversion was accomplished by using the appropriate dihedral angle (carbon of the methyl group on N, N, C, carbonyl oxygen) as the reaction coordinate and rotating the angle from 40 ° to 0 ° and finally to -40 °. Step 3 involved formation of the acylenzyme intermediate. The distance between the nitrogen of the substrate and the hydrogen on N_ε of His 40 was the reaction coordinate, decreasing from 2.67, 1.85 to 1.00 Å. During the proton transfer, cleavage of the C-N bond (or C-O with the ester) of the substrate occurred, leaving the free amine and the acylenzyme intermediate.

The next step in the reaction was to replace the amine or alcohol by a water molecule (step 4). The water molecule was superimposed on the NH_2 portion of the methyl-amine group or the hydroxyl group of the methanol to get its initial position, followed by full geometry optimization. Step 5 represents attack of water on the acylenzyme, leading to formation of the second tetrahedral intermediate. The grid search approach was used as described in step 1. In this case, the distance between the water's oxygen and the carbonyl carbon were fixed at 3.13, 2.30 and 1.50 Å. At these fixed positions the distance between N_ε and one of the water's hydrogens was decreased from 1.77, 1.40 to 1.00 Å. The final step, regeneration of the enzyme with bound product, was accomplished by transferring the proton from N_ε of the His mimic to the methoxy group representing the serine (the distance began at 1.81 and was decreased to 1.40 and then 1.00 Å, step 6). The bond between the serine and the product broke as a result of the transfer. All of the final structures resulting from the reactions described above were fully optimized with respect to all degrees of freedom. To determine the importance of the Asp and the oxyanion hole in stabilizing the intermediates along the reaction pathway, the energy of the optimized structures was reevaluated leaving out the Asp mimic and the two water molecules representing the oxyanion hole.

RESULTS

We have performed semiempirical molecular orbital calculations of trypsin-catalysed amide and ester hydrolysis. First, we present the results of the calculations on the model systems relevant to the reaction pathway and then the results on the enzymatic reactions.

Model Calculations

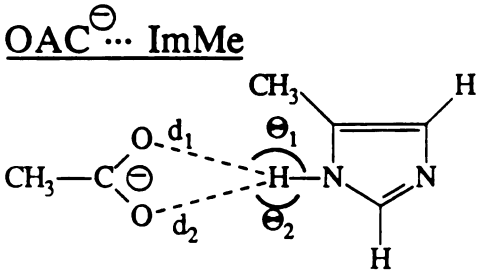
We had two main objectives in performing the model compound calculations: to determine how well the models reproduce hydrogen bond geometries of complexes relevant to trypsin catalyzed amide hydrolysis and to determine which molecular model to employ (AM1 or PM3). First, we calculated proton affinities for methyl-imidazole (the histidine mimic), methanol (representing the serine), and acetic acid (for the Asp). Our calculated energies and the corresponding experimental values are given in Table 7.1. The experimental heat of formation for protons (367.2 kcal/mole; Stull and Prophet, 1971) was used in determining the proton affinities. The structures for the molecules were taken from the minimized trypsin structure as described above. The calculated proton affinities of methoxide and acetate using AM1 agree well with the earlier values calculated by Dewar and Dieter (1986) (Table 7.1). The proton affinities calculated with PM3 were close to those found using AM1, but PM3 better represented the experimental data overall.

We then evaluated geometric properties of hydrogen bonds between pairs of molecules relevant to the catalytic process in order to compare the two quantum mechanical. We examined methyl-imidazole complexes with acetate with the imidazole both neutral and positively charged (Figure 7.4). With the negatively charged complex (OAc^-...Im-Me), PM3 gave shorter hydrogen bond distances and closer to linear hydrogen bond angles compared to AM1 (d_1 and θ_1 of Figure 7.4). In fact, d_1 with PM3 was very close to the value in minimized trypsin (1.76 Å). In the neutral complex (OAc^-...Im-Me^+), PM3 again produced an hydrogen bond distance of nearly the same length as in minimized trypsin (1.60 versus 1.65 Å) and a close to linear hydrogen bond

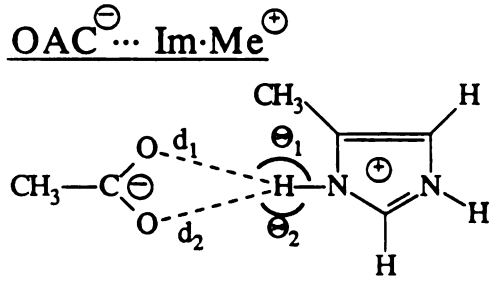
TABLE 7.1: Proton Affinities for Model Compounds Using AM1 and PM3

Reaction	AM1[c]	AM1	PM3	Expt
Im.Me + H^+ ---> Im.Me$^+$	--	224.5	221.8	220.0[a]
MeO^- + H^+ ---> MeOH	385.4	385.7	381.1	376.8[b] 379.2[c] 381.4[c]
OAc^- + H^+ ---> HOAc	354.2	354.8	349.6	342.0[d] 348.5[c] 345.2[c]

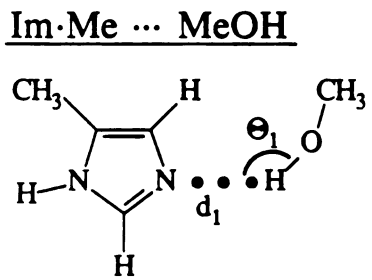
[a]Kollman and Hayes (1981).
[b]McIver and Miller (1974).
[c]Dewar and Dieter (1986).
[d]Yamdagni and Kebarle (1973).



AM1	d_1 = 1.97 Å	Θ_1 = 165.8
	d_2 = 2.20 Å	Θ_2 = 128.7
PM3	d_1 = 1.72 Å	Θ_1 = 174.2
	d_2 = 2.58 Å	Θ_2 = 128.8



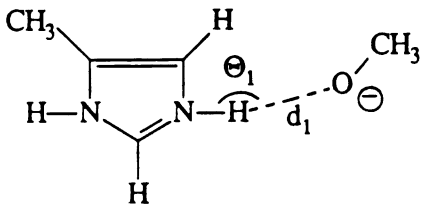
AM1	d_1 = 0.97	Θ_1 = 146.4
	d_2 = 2.27	Θ_2 = 97.3
PM3	d_1 = 1.60	Θ_1 = 174.3
	d_2 = 2.40	Θ_2 = 112.9



AM1	d_1 = 2.66	Θ_1 = 157.7
PM3	d_1 = 2.49	Θ_1 = 155.2

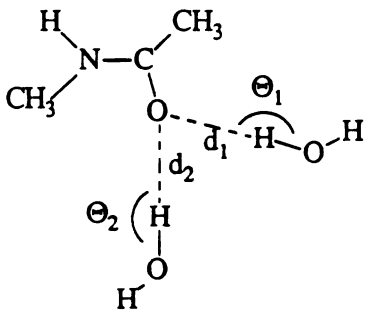
Figure 7.4: Geometric properties of hydrogen bonds in optimized model complexes using the molecular models AM1 and PM3.

Im·Me$^{\oplus}$... MeO$^{\ominus}$



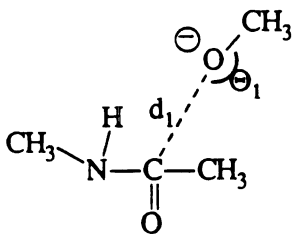
AM1	$d_1 = 3.21$	$\Theta_1 = 97.0$
PM3	$d_1 = 1.69$	$\Theta_1 = 120.1$

NMA ... $2H_2O$



AM1	$d_1 = 2.30$	$\Theta_2 = 135.4$
	$d_2 = 2.19$	$\Theta_2 = 104.0$
PM3	$d_1 = 1.81$	$\Theta_1 = 178.1$
	$d_2 = 1.81$	$\Theta_2 = 166.2$

NMA ... MeO$^{\ominus}$



AM1	$d_1 = 3.17$	$\Theta_1 = 123.0$
PM3	$d_1 = 3.52$	$\Theta_1 = 116.8$

Figure 7.4: continued

angle (θ_1, Figure 7.4). With AM1, the acetate molecule abstracted the proton from the imidazole group and a stable charged complex could not be found.

We then investigated methyl-imidazole complexes with methanol. Use of both AM1 and PM3 models produced long hydrogen bonds that were kinked in the complex with both groups neutral (Im-Me...MeOH) (Figure 7.4). When starting with both groups ionized ($Im\text{-}Me^+...MeO^-$), AM1 gave an hydrogen bond angle of 97° and PM3 gave an improved but small angle of 120°. The hydrogen bond distance was much improved using PM3, however.

Both models gave similar results for the substrate-methoxide ($NMA...MeO^-$) complex, with AM1 doing a slightly better job. The PM3 model was much better than AM1 in mimicking the oxyanion hole hydrogen bond distances ($NMA...2H_2O$, Figure 7.4). In fact, PM3 reproduced the hydrogen bond distances found in minimized trypsin almost exactly (d_1=1.84 Å and d_2=1.90 Å in trypsin).

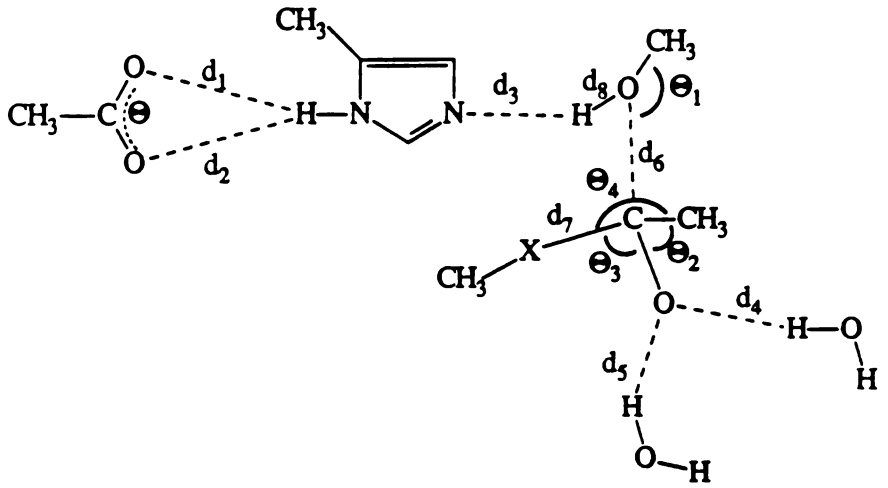
Given the preference of PM3 over AM1 with respect to the proton affinities and clear advantage regarding the hydrogen bond geometries, all further studies were carried out using the PM3 model.

Reaction Pathway

We calculated the steps along the reaction pathway shown in Figure 7.3, with the amide substrate shown. The starting structure (ES) was extracted from the active site of minimized trypsin and then optimized using the PM3 model. The relative orientations of the active site residues were fairly well-maintained in the truncated system (compare MIN and ES in Table 7.2), which was the starting structure for the calculations along the reaction pathway. The reaction coordinate method was used for the steps along the pathway where the reaction coordinate for each step is specified in Figure 7.3 by an arrow. The energy differences given in the figure are between the energies of the optimized structures and also represent activation energies except where noted below.

TABLE 7.2

Distances between Active Site Residues in Minimized Trypsin (MIN) and between corresponding Models Optimized Structures in Acylation Pathway.



		AMIDE SUBSTRATE X = NH				ESTER SUBSTRATE X = O		
Property	MIN	ES	TET_1	TET_1'	EA_1	ES	TET_1	EA
d_1	1.76	1.70	1.65	1.66	1.69	1.70	1.65	1.71
d_2	2.15	2.15	2.42	2.44	2.49	2.52	2.43	2.57
d_3	1.96	1.76	1.00	1.02	1.76	1.77	1.02	1.79
d_4	1.84	1.81	1.77	1.75	1.80	1.82	1.75	1.81
d_5	1.90	2.53	1.75	1.74	1.77	2.55	2.45	2.63
d_6	2.91	3.30	1.50	1.48	1.41	3.32	1.48	1.36
d_7	1.34	1.42	1.53	1.54	1.78	1.36	1.44	5.55
d_8	0.97	0.98	1.75	1.75	1.54	0.98	1.74	5.83
Θ_1	104	154	118	118	119	153	117	120
Θ_2	121	123	116	116	120	127	117	128
Θ_3	122	120	113	112	----	120	116	----
Θ_4	116	117	108	111	----	114	107	----

* MIN represents the minimized X -ray coordinates of trypsin with a peptide substrate as described in the text.

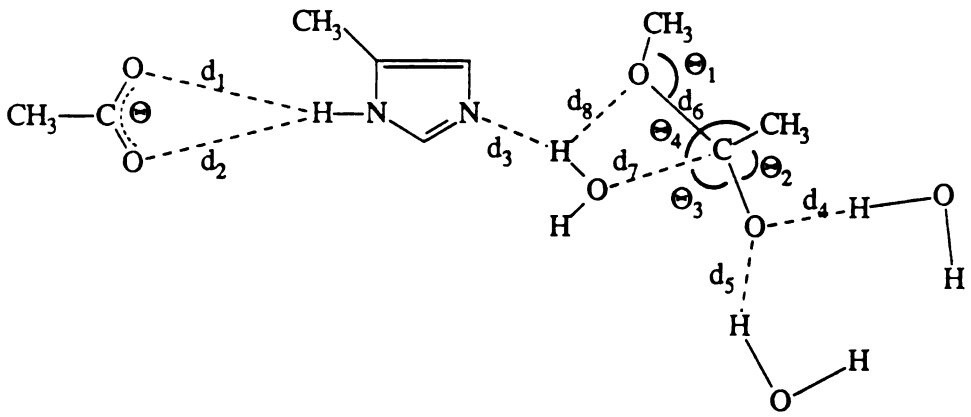
The first step in the trypsin pathway represents proton transfer from the serine to the histidine and attack of the carbonyl carbon by the serine to form the tetrahedral intermediate. Since there were two reaction coordinates for this first step, we used a grid approach with proton transfer (d_1 of Tables 7.4 and 7.5) occurring at fixed, but different, distances between the serine oxygen and the carbonyl carbon (d_2 of Table 7.4 and 7.5). This type of approach allowed us to investigate the energetics of different possible modes of forming the tetrahedral intermediates. The lowest energy path is underlined in both tables. The most favorable path for both substrates was for approach of the serine, while maintaining the Ser...His hydrogen bond. When the O-C bond distance was 2.10 Å, partial transfer of the proton was favored, and from that point simultaneous transfer of the proton and formation of the C-O bond were favored. This progression yielded an activation energy of 31.5 kcal/mole for the amide and 22.3 kcal/mole with the ester. A slight approach of Ser to the substrate (3.3 $\rightarrow$ 2.7 Å), followed by concerted attack and proton transfer (moving along the diagonal in Table 7.3) gave the same activation energy in both cases. When these two reaction coordinates were considered separately, complete proton transfer followed by methoxide attack on the substrate, the calculated activation energies were 35.0 and 31.5 kcal/mole for the amide and ester, respectively.

Step 1 yielded the tetrahedral intermediate for acylation (TET_1, optimized structure shown in Figure 7.5). As can be seen, the substrate carbon did indeed became tetrahedral (compare structures ES and TET_1 of Figure 7.5). In amide hydrolysis, the hydrogen bonds to the water molecules making up the oxyanion hole improved in forming the oxyanion, while other interactions remained roughly the same (Table 7.2). The hydrogen bonds to the oxyanion hole did not improve upon forming the tetrahedral intermediate with the ester substrate, however (Table 7.2).

The next step in the reaction pathway for the amide substrate involved inversion of the substrate's nitrogen. This process was favorable, by -3.9 kcal/mole, and without a barrier to rotation. Rotation of this bond left the nitrogen of the substrate in a better posi-

TABLE 7.3

Distances between active Site Residues in Minimized Trypsin (MIN) & between Corresponding Models in Optimized Structures in the deacylation pathway.

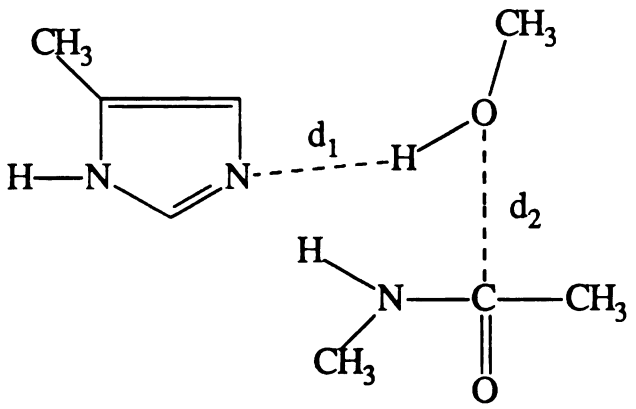


Property	MIN	EA_2	TET_2	E.P
d_1	1.76	1.71	1.66	1.70
d_2	2.15	2.52	2.39	2.51
d_3	1.96	1.78	1.05	1.76
d_4	1.84	1.84	1.77	1.83
d_5	1.90	1.81	1.75	1.82
d_6	2.91	1.35	1.46	3.66
d_7	1.34	3.90	1.42	1.34
d_8	0.97	3.80	1.81	0.98
Θ_1	104	121	117	128
Θ_2	121	127	119	127
Θ_3	122	----	104	117
Θ_4	116	----	113	116

MIN represints the minimized X - ray coordinates of trypsin w/ a peptide substrate as described in the text.

TABLE 7.4

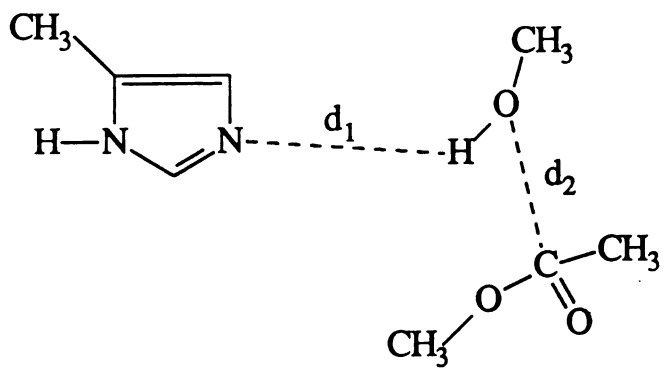
Energies for Formation of First Tetrahedral Intermediate with an Amide Substrate (kcal/mole)*



Fixed d_2 (Å)	d_1 (Å) 1.77	1.40	1.00
3.30	-357.8	-345.4	-331.3
2.70	-355.5	-343.5	-328.6
2.10	-344.1	-333.2	-322.8
1.50	-315.0	-310.8	-326.3

*Step 1 of Figure 7.3. The low energy path is underlined.

TABLE 7.5
Energies for Formation of First Tetrahedral Intermediate with an Ester Substrate (kcal/mole)*



Fixed d_2 (Å)	d_1 (Å) 1.77	1.40	1.00
3.30	-396.3	-384.9	-369.5
2.70	-394.7	-382.8	-367.6
2.10	-382.0	-374.0	-364.8
1.50	-360.6	-357.3	-374.2

*Step 1 of Figure 7.3. The low energy path is underlined.

tion to abstract the proton on the histidine (3.10 Å in structure TET_1 versus 2.84 Å in structure TET_1'). All other interactions shown in Table 7.2 were maintained.

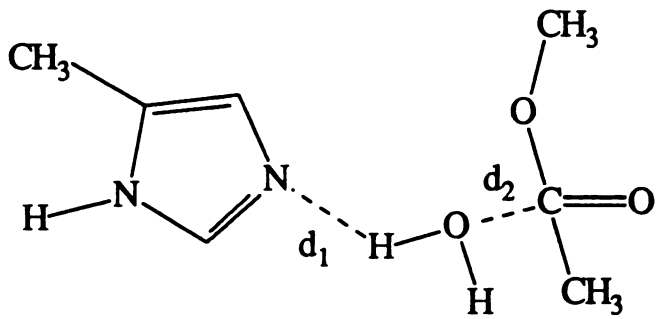
The next step in the pathway involved transfer of a proton from the positively charged histidine to the nitrogen of the substrate. As a result of this proton transfer the C-N bond of the amide or C-O bond of the ester was cleaved, leaving the acylenzyme intermediate (EA_1) and methylamine. This process was favorable and again without an activation barrier. After cleavage, the histidine remained hydrogen bonded to the methylamine or methanol (Table 7.2). Step 4 of Figure 7.3 represents not a chemical step but instead the exchange of methylamine or methanol for water (EA_2).

After postioning of the water molecule, there was a proton transfer from the water to the histidine and attack of the carbonyl carbon of the acylated serine by water, yielding the second tetrahedral intermediate, TET_2 (step 5, Figure 7.3). Here again we used a grid approach to examine the energetics of different possible pathways for this process (Table 7.6). The lowest energy pathway is underlined in Table 7.6 and had an activation energy of 15.6 kcal/mole. The concerted process with simultaneous proton transfer and attack also had an activation energy of 15.6 kcal/mole (moving along the diagonal in Table 7.6). The other possible path had a much higher activation energy. To first transfer the proton to give attack by an hydroxyl group yielded a 39.1 kcal/mole barrier, which was due entirely to abstracting the proton (Table 7.6). This second tetrahedral intermediate (TET_2) was very similar to the first tetrahedral intermediate (TET_1) (Figure 7.5 and Tables 7.2 and 7.3).

The final step in this process was transfer of a proton from the histidine to the serine and regeneration of the resting state of the active site residues with bound product (EP in Figure 7.5). Breakdown to product occurred as a result of this proton transfer. The hydrogen bonds to the oxyanion hole became slightly longer with loss of the oxyanion (structures TET_2 and EP of Figure 7.5 and Table 7.3). Although this appeared to be a favorable process, this step had an activation barrier of 14.2 kcal/mole.

TABLE 7.6

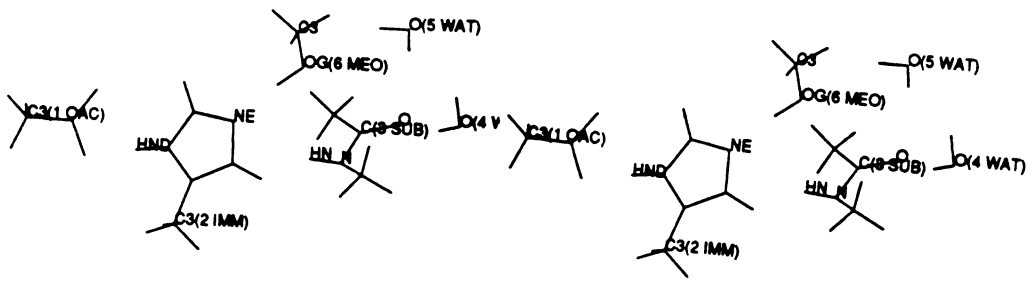
Energies for Formation of Second Tetrahedral Intermediate (kcal/mole)*



	d_1 (Å)		
Fixed d_2 (Å)	1.77	1.40	1.00
3.13	-396.7	-384.4	-357.5
2.30	-393.2	-381.1	-359.3
1.50	-379.1	-375.8	-384.9

*Step 5 of Figure 7.3. The low energy path is underlined.

[ES]



[TET_1]

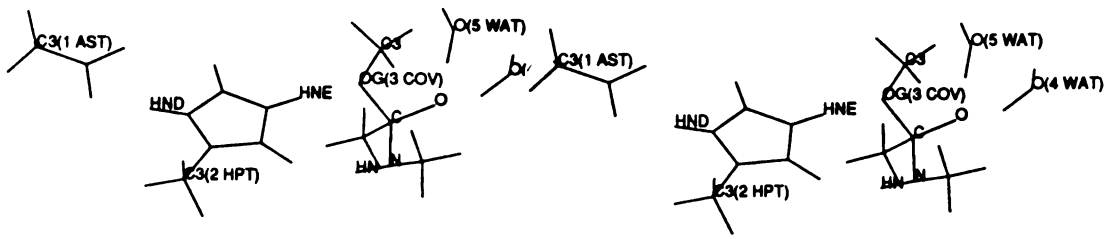
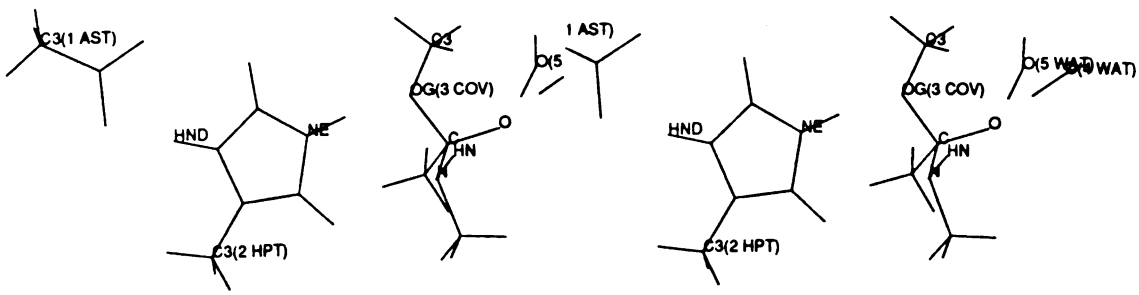
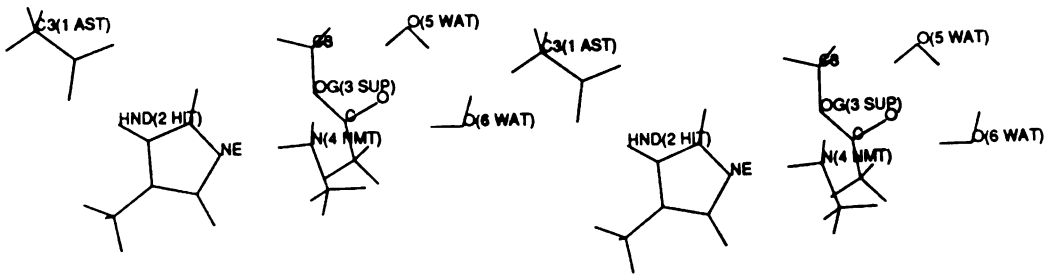


Figure 7.5: Stereoviews of optimized structures along the reaction pathway.

$[TET_1']$



$[EA_1]$



$[EA_2]$

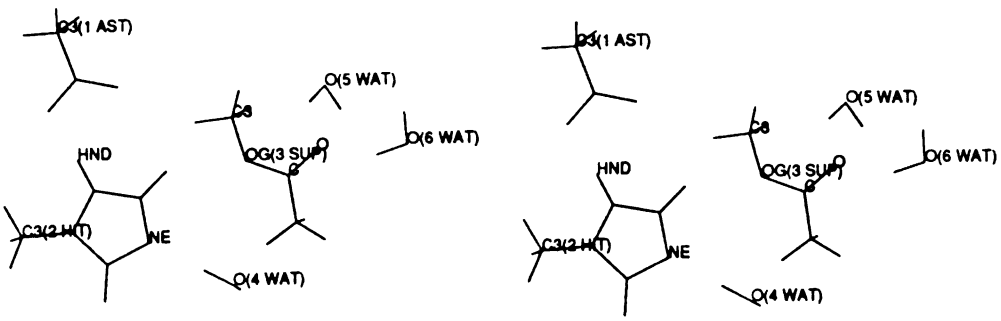
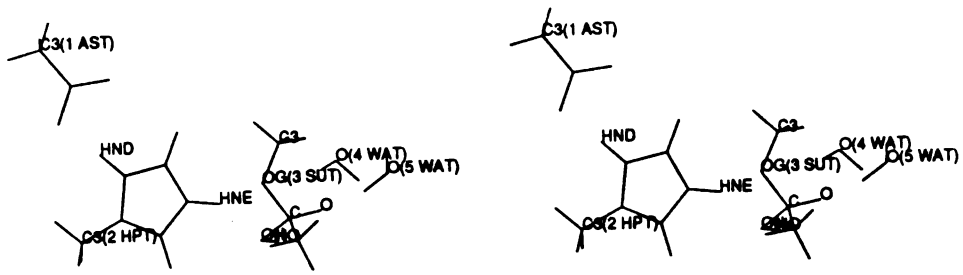


Figure 7.5: continued

[TET_2]



[EP]

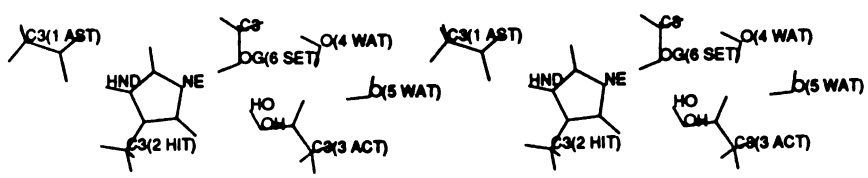


Figure 7.5: continued

After investigating the trypsin pathway with the residues of the catalytic triad and oxyanion hole mimics, we examined the importance of the Asp and the oxyanion hole to this reaction with the amide substrate. To this end, we took the optimized structures (Figure 7.5) along the pathway and evaluated their energies without the groups in question. The results are shown schematically in Figure 7.6 with the values for the reaction with the ester given for comparison. As can be seen, removing the Asp mimic and oxyanion hole oxygens drastically destabilized the three tetrahedral intermediates along the pathway (structures TET_1, TET_1', and TET_2 of Figure 7.6). In each of these cases the Asp contributed more to stabilization of the tetrahedral species than did the water molecules representing the oxyanion hole. These groups were not as important in the stabilization of the other structures, except for approximately 6 kcal/mole stabilization of the acylenzyme intermediate provided by the oxyanion hole (EA_1). Asp was actually destabilizing for the acylenzyme intermediates and the enzyme-product complex (structures EA_1, EA_2, and EP).

DISCUSSION

We began doing model calculations relevant to trypsin catalysis using the AM1 molecular model. We used AM1 because it was reported that AM1 reproduces experimental hydrogen bond energetics (Dewar et al., 1985; Dannenberg, 1988) and proton affinities (Dewar and Dieter, 1986). However, it was also noted that the structural features of hydrogen bonds are not well reproduced by AM1 (Dannenberg, 1988; Williams, 1987). We also found some bad geometries for hydrogen bonded complexes using AM1 and tested the PM3 model when it became available. We calculated similar proton affinities using AM1 and PM3 although PM3 was slightly better. PM3 produced much better hydrogen bond geometries than AM1, however. A major problem with using AM1 to study trypsin catalysis is that the complex between the positively charged histidine mimic and acetate (for Asp) was not stable. Instead of maintaining their charged states, the acetate molecule abstracted one of the imidazole's protons. Protonated Asp 84 is not

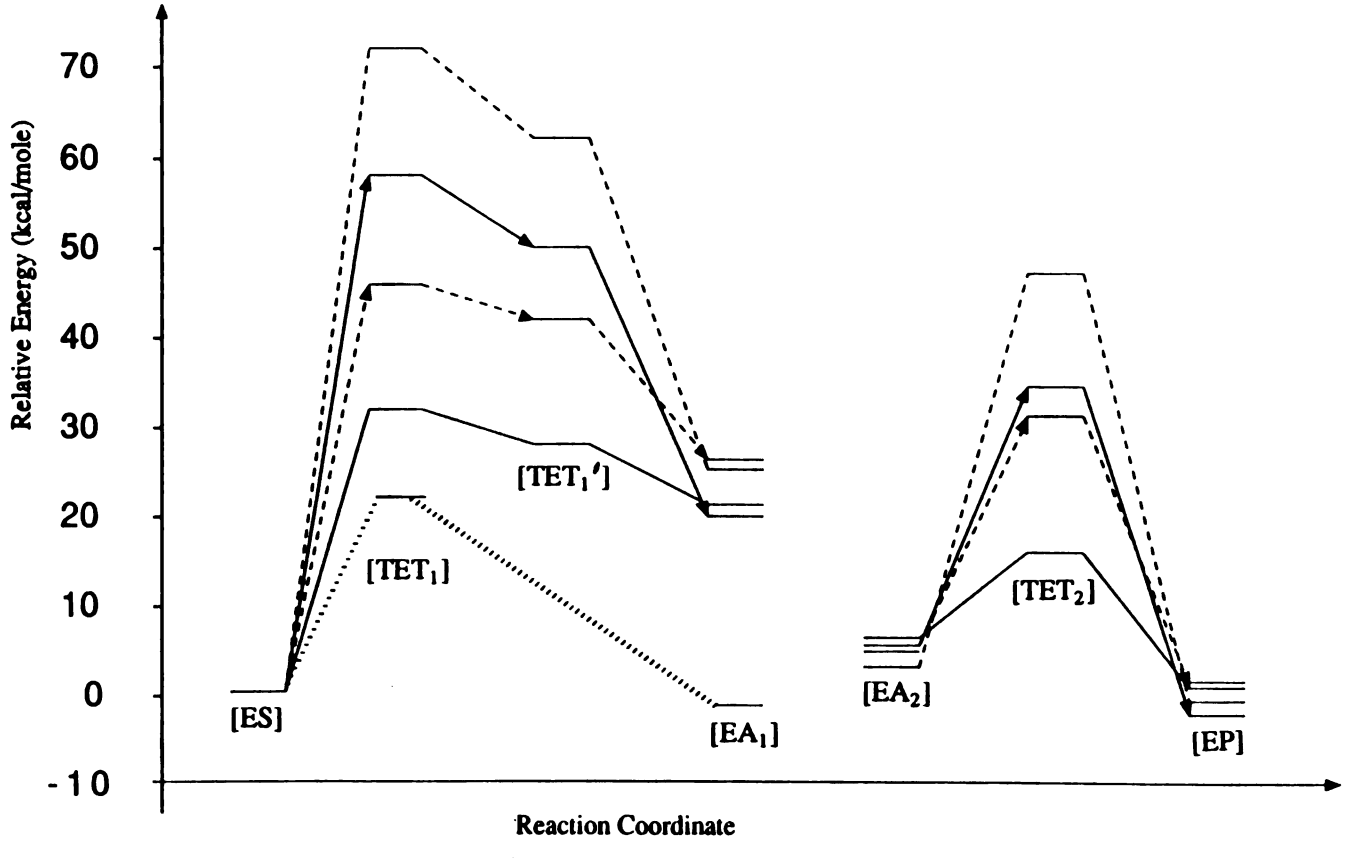


Figure 7.6: Relative energies of structures along the reaction path as a function of environment. Within each curve, the energies are relative to the energy of the appropriate starting structure. The solid line is for the full system as shown in figure 7.3; the solid line with an arrow is for the system without the Asp mimic; the dashed line with an arrow is for the system without the waters representing the oxyanion hole; the dashed line is for the system without Asp and without the oxyanion hole; and the final line extending only as far as [EA] is with the ester substrate.

observed experimentally (Kossiakoff and Spencer, 1981). Hence, PM3 was the best molecular model for this system where hydrogen bonding is clearly important and where the process of interest is occurring in the enzyme active site as opposed to the gas phase. For this reason, all further studies were carried out with the PM3 model.

For the reaction pathway for serine proteases shown in Figure 7.3, the rate limiting step was formation of the first tetrahedral intermediate with both the amide and ester substrates. Experimentally, the rate limiting step for amide hydrolysis is acylation and is deacylation for ester hydrolysis (Fersht, 1985). And, based on pre-steady state kinetic studies, formation of the tetrahedral intermediate preceding the breakdown to the acylenzyme appears to be the actual rate-determining step (Hirohara et al., 1974, 1977). We did not observe differences in the rate-limiting step with the different substrates although the activation energy for acylation was lower with the ester than the amide. If arguments regarding motion are invoked, deacylation becomes the rate-determining step with the ester, though (discussed below).

Although many investigators have examined the mechanism of formation of the tetrahedral intermediate, the question has not been definitively resolved. After the catalytic triad was first observed (Blow et al., 1969), it was suggested that the nucleophilicity of the serine was increased by a concerted transfer of two protons, Ser $\rightarrow$ His $\rightarrow$ Asp (Hunkapillar et al., 1973). Although this mechanism was generally accepted for a number of years, more recent NMR (Bachouchin and Roberts, 1978) and neutron diffraction studies (Kossiakoff and Spencer, 1981) indicate that the triple ion (Asp^-...His^+...Ser^-) is favored over the double proton transfer generated structure (Asp-H...His...Ser^-) in the tetrahedral intermediate. This finding suggests, then, that the Asp will remain ionized in the transition state as well (Warshel and Russell, 1986), but a proton transfer at the transition state cannot be ruled out.

Many quantum mechanical calculations employing a system composed entirely of the catalytic triad (or reduced representations of the triad) without the surrounding pro-

tein support the concerted double proton transfer mechanism (Dewar and Storch, 1985; Scheiner et al., 1975; Umeyama et al., 1973). Studies by Kollman and Hayes using a limited system, however, concluded that the triple ion ($Asp^-...His^+...Ser^-$) is lower in energy than (Asp-H...His...Ser^-) (Hayes and Kollman, 1979; Kollman and Hayes, 1981). Calculations including the protein environment in some way have also contradicted the concerted double proton transfer mechanism (Umeyama et al., 1981; Warshel and Russell, 1986). Calculations presented here on our simple model also are consistent with the single proton transfer mechanism; we never observed proton transfer from the histidine to the aspartate (Table 7.2). Nor did we ever observe a bifurcated hydrogen bond to the Asp.

It is probably misleading to speak of triple ions in the reaction pathway of the serine proteases, because it is unlikely that a bare methoxide (MeO^-) is produced. Kollman and Hayes (1981) found the Ser → His proton transfer accompanied by Ser attack on the substrate with an upper bound of 27 kcal/mole. Other workers have found that proton transfer is approximately complete prior to nucleophilic attack (Scheiner and Lipscomb, 1976; Warshel and Russell, 1986). Our work here shows almost the opposite proton transfer behavior. We found that the lowest energy path for formation of the tetrahedral intermediate was for serine first to approach the substrate. When serine's oxygen was within approximately 2 Å of the substrate carbon, proton transfer began. Finally, from this position, Ser attack and proton transfer occurred simultaneously. This mechanism involving coupled heavy atom movement and proton transfer suggests that there is protonic bridging in the transition state, which is consistent with isotope-effect experiments (Schowen, 1988). Our estimate of the activation barrier is probably an upper bound for the formation of the tetrahedral intermediate, as dynamical fluctuations bringing O_γ and the carbonyl carbon of the substrate near in space will facilitate catalysis. A fluctuation of 1.2 Å could decrease the activation energy of amide hydrolysis to 18 kcal/mole (Table 7.4) and the activation energy for the hydrolysis of the ester would decrease to 8.9

kcal/mole (Table 7.5). If this same argument is applied to the deacylation step, thermal motion bringing the water closer to the acylated serine would lower the activation energy for this process to 12 kcal/mole. A fluctuation of this magnitude (~1.2 Å) is possible although probably relatively rare, like catalysis itself. In any case, if thermal motion facilitates heavy atom motion, acylation would still have the higher activation barrier with amides but the barrier to deacylation would be greater than for acylation with esters. Then, the final step, that of $TET_2 \rightarrow EP$, would be rate limiting for the ester with an activation barrier of 14.2 kcal/mole. If we again attempt to account for motion and in this case assume that protein motion could bring the proton on the histidine 0.4 Å closer to O_γ of the serine, the reaction occurs without an activation barrier. Thus, allowing that the motion can be facilitated in this way, these results are consistent with the experimental findings.

Komiyama and Bender have proposed another model for serine protease catalysis (1979). They suggest that a tetrahedral intermediate is not formed at all. Instead, they propose that a proton is donated directly to the nitrogen without ever forming the C-O bond between serine and the substrate. There is experimental evidence, that inhibitors with tetrahedral geometry bind more tightly than substrates, which, albeit circumstantial, is in opposition to Komiyama and Bender's proposal. Also, there is an entropic advantage to use only one proton transfer rather than to engineer precise alignment of all three residues for concerted proton transfer (Scheiner and Hillebrand, 1985).

Schowen and co-workers have found that for normal substrates only one proton transfer is indicated in the transition state using solvent isotope effects and the proton inventory technique (Stein et al., 1983; Elrod et al., 1980). Further, they show that two protons appear to be transferred with very good substrates, suggesting interactions between the substrate and the enzyme at sites distant from the active site. The first proton transfer is attributed to the Ser → His movement, as mentioned above. The second transferred proton has been attributed to His → Asp; however based on all the experi-

mental and theoretical evidence, this is unlikely. Alternatively, "compressive forces" leading to a bridging hydrogen between His and Asp have been proposed to be responsible (Schowen, 1988). Other possible reasons are the formation or strengthening of a number of hydrogen bonds in the tetrahedral intermediate (Fink, 1987) or that better hydrogen bonds are formed with the oxyanion hole in the tetrahedral intermediate than in the Michaelis complex. Our results support this last hypothesis. In our calculations, the hydrogen bonds to the oxyanion improved greatly in forming the tetrahedral intermediate (1.81 $\rightarrow$ 1.77 Å and 2.53 $\rightarrow$ 1.75 Å). The hydrogen bonds to Asp only improved slightly (< 0.1 Å), which is in opposition to the hypothesis regarding "compression" of the Asp and His (Schowen, 1988). We did not observe substantial differences in other hydrogen bonds. Hence, our results suggest that the improved hydrogen bonds to the oxyanion may be responsible for the apparent second proton transfer in good substrates. But, given our reduced system, the other alternatives cannot be ruled out.

In the case of the deacylation step (formation of the second tetrahedral intermediate), we also found that the low energy path was for approach of the water, partial transfer of the proton, and then simultaneous attack and transfer (Table 7.6). Concerted proton transfer and attack had the same activation energy, although it was not the low energy path considering the individual steps. Based on our results, if the enzyme had to deprotonate the water completely, prior to attack, this step would have the rate-determining activation barrier. By allowing attack to occur with deprotonation, the activation barrier was lowered by approximately 24 kcal/mole. Warshel and co-workers (1989) have also pointed out how enzymes cannot use bare water molecules because of the high energy required to strip off a proton (~22 kcal/mole) and how metals and protein residues can aid in this process to lower the barrier.

So far we have only discussed proton transfer and nucleophilic attack without considering the importance of conformation. Enzymes act by lowering the activation free energies of the reactions they catalyze. This can be accomplished in a variety of ways:

(1) active sites with electrostatic complementarity to the transition state, (2) inducing strain in the substrate, and (3) with substrate binding that induces a change in the enzyme such that it becomes more reactive. Warshel and co-workers have argued persuasively for the importance of the first mechanism for serine proteases (Warshel and Russell, 1986; Warshel et al., 1989); this possibility is discussed further below. The second mechanism has gained support from the fact that the binding site is complementary to the structure of the tetrahedral intermediate (Henderson et al, 1971; Poulos et al., 1976). There is NMR evidence, however, that the carbonyl carbon of the substrate remains relatively planar in the Michaelis complex and hence that there is little to no distortion of the substrate on binding (Baillurgeon et al., 1980; Richarz et al, 1980). In our models, the carbon did indeed remain planar in the presence of the active site residues and was not induced to adopt the tetrahedral geometry. On work with transition state analogs it has been suggested that the major driving force for catalysis is the distortion of the scissile bond from planarity by the enzyme, thereby eliminating the stabilization energy of the peptide bond (Delbaere and Brayer, 1985). In our models we found comparable values for the ω dihedral angle in the substrate, substrate complex with water molecules, and in the structure with the active site residues (ω=161, 163, 155° respectively). Thus, our results are inconsistent with this hypothesis, but given our limited representation of the active site, we cannot rule out the strain hypotheses. Warshel and co-workers have pointed out that even a change in geometry of the carbonyl carbon from sp^2 to sp^3 results in less than a 0.3 Å displacement of the surrounding atoms, which can easily be accommodated by the flexible enzyme (1989).

The third possible manner for lowering the activation energy of a process is for the substrate to induce a change in the enzyme that favors catalysis. This appeared to be the case with serine proteases. For many years it was thought that the Ser and His of the catalytic triad did not form an hydrogen bond (Brayer, 1979). It has been shown more recently that an hydrogen bond is present in most free serine proteases (Tsukada and

Blow, 1985; Bachovchin, 1986). These residues also form a strong hydrogen bond in the presence of substrate (Huber and Bode, 1978). Instead, this mechanism for lowering the activation energy may be due merely to burying of the charge relay system when substrate or inhibitor is bound, enhancing the polarizing influence of the system (Schowen, 1988). In our simple system we observed a slightly long but acceptable hydrogen bond between the Ser and His mimics in the model complex (2.49 Å) that decreased in the presence of the substrate (1.76 Å).

The theory of stereoelectronic control has been applied to serine proteases (Bizzozero and Dutler, 1981; Taira and Gorenstein, 1987). Molecular orbital calculations provide support for Deslongchamps' work (1983) demonstrating selective cleavage of bonds that are trans-antiperiplanar (app) to lone pairs on directly bonded oxygen or nitrogen atoms on compounds with tetrahedral carbons (Gorenstein and Taira, 1984). The stereoelectronic effect is presumed to be due to interactions between a σ^* antibonding C-X_1 orbital with the app lone pair on X_2, where X=oxygen or nitrogen (Gorenstein et al., 1980). Or, in other words, this effect has a double bond-no bond resonance contribution, thereby making the particular bond easier to make or break (Taira and Gorenstein, 1987). Furthermore, Gorenstein and co-workers suggest that the stereoelectronic effect is more important in the transition states than in the ground states or intermediates (1980).

For our purposes, the application of this theory, as presented by Taira and Gorenstein (1987), to serine proteases stresses two factors for amide hydrolysis: that C-O bond rotation of the serine occur and that the nitrogen of the substrate is inverted after formation of the tetrahedral intermediate. For nucleophilic attack by the serine the most stereoelectronically favored transition state possesses a nitrogen lone pair app to the nascent O-C bond while not containing a lone pair on the serine oxygen that is app to the substrate's N-C bond. Then, to facilitate N-C cleavage, the lone pair on the serine oxygen must be app to the scissile bond and the nitrogen should not have a lone pair app to the serine substrate bond. This conformation is achieved by N-inversion and a conforma-

tional change in the serine, causing kinking of the C_β-O_γ-C angle. The N-inversion stabilizes the intermediate by 3.1 kcal/mole (Taira and Gorenstein, 1987). It has also been argued that the nitrogen inversion prevents reversion to the Michaelis complex, thus acting as a switch controlling whether the C-O bond or the C-N bond is broken (Bizzozero and Dutler, 1981).

In our calculations we found kinking of the serine side chain as proposed by Taira and Gorenstein; however, we observed the conformational change in the serine during the formation of the first tetrahedral intermediate, and the resulting conformation was maintained in the N-inverted form of the intermediate. But, the lone pairs, as required by the stereoelectronic effect, may have been altered with rotation of the nitrogen while maintaining θ_1 of Table 7.2. We found that inversion of the nitrogen stabilized the tetrahedral intermediate by 3.9 kcal/mole, which is close to the value reported by Taira and Gorenstein (1987). This rotation put the nitrogen in a much better position to abstract a proton from the histidine mimic. Our studies cannot distinguish between stereoelectronic control in determining the relative energies of the two forms of the first tetrahedral intermediate (structures TET_1 and TET_1' of Figure 7.5) and the importance of other interactions. The improvement we saw in inverting the nitrogen could be due to release of H-H repulsion with nitrogen inversion that leads to favorable N...H interactions. It has been pointed out by Asboth and Polgar that hydrogen bonding or any other significant interaction with the protein environment, could override stereoelectronic control (1983). There are also a number of other problems with stereoelectronic theory as applied to serine proteases and they have been reviewed by Fink (1987).

What is the role of environmental effects on the catalytic process? Asp and the oxyanion hole are critical components of the active site and, in fact, have been seen in all serine proteases to date (Kraut, 1977). One of the hydrogen bonds to the oxyanion hole is long in the Michaelis complex (Fersht, 1985) but strong hydrogen bonds are formed with each as the reaction proceeds from C=O $\rightarrow$ C-O$^-$ (Fersht et al., 1973). The oxyan-

ion hole forms strong hydrogen bonds with the oxyanion of transition state analog inhibitors (Kossiakoff and Spencer, 1981). Thus, the oxyanion hole appears to not only be suited for binding the transition state or tetrahedral intermediate but to stabilize the developing negative charge on the oxygen. We also found a long hydrogen bond initially in our models that decreased as the tetrahedral intermediate was formed (Table 7.2). Removing the oxyanion hole waters from our models for amide hydrolysis had a drastic effect on the stability of the tetrahedral intermediates, destabilizing them by 13-14 kcal/mole. This can be related to the lower enzymatic activity (10^{-4}-10^{-7}) of serine protease zymogens compared to active enzymes (Steitz and Shulman, 1982). The major difference in the zymogens appears to be the lack of the oxyanion hole.

Aspartate may be an invariant member of the active site of serine proteases because of its role in stabilizing the transition state through electrostatic interactions (Warshel et al., 1989; Warshel and Russell, 1986; Nakagawa and Umeyama, 1984). This is important because the formation of the transition state involves a drastic change in the charge distribution, which is akin to the formation of an ion pair from a neutral ground state (Warshel et al., 1989). Another possibility is that the Asp is there to raise the pKa of the histidine and confine its location (Fersht and Sperling, 1973; Sprang et al., 1987). All of these mechanisms are probably in effect.

We found that the Asp mimic in our models for amide hydrolysis stabilized the first tetrahedral intermediate and its N-inverted form by 26.6 and 22.2 kcal/mole, respectively. Asp was destabilizing for the other intermediate states (besides the second tetrahedral intermediate); the largest effect was on the enzyme-product complex. The destabilizing effect of the Asp in this case may be to keep the enzyme-product complex from being too stable so that the enzyme can be regenerated for further hydrolysis reactions. We found that the oxyanion hole hydrogen bonds stabilized the tetrahedral intermediates by 13.5-14.5 kcal/mole. Overall, we found the Asp to be more important in stabilizing the tetrahedral intermediates, and therefore presumably the transition states, than

the oxyanion hole. Our values for the stabilization afforded by the Asp and oxyanion hole are greatly exaggerated, however, compared to experiment because of the limited representation of the environment. Experimentally, the replacement of Asp by a neutral residue destabilizes the transition state by approximately 4 kcal/mole (Craik et al., 1987). Experimentally, one hydrogen bond of the oxyanion hole contributes about 5 kcal/mole to stabilization of the tetrahedral intermediate (Wells et al., 1986).

We do not mean to suggest with the use of such a reduced representation of the trypsin active site that the protein is unimportant in catalysis, especially in light of evidence indicating that specific side chain interactions between substrate and enzyme sites remote from the active site can alter both the rate of catalysis and the mechanism (Schowen, 1988). Due to system size limitations, though, one is forced to employ truncated systems even with semiempirical methods. We find it gratifying that such a simple model can reproduce many aspects of trypsin catalysis. Inclusion of the protein environment and surrounding solvent will, no doubt, be necessary for quantitative agreement with experiment, but we find these initial results encouraging and informative. Our studies suggest modes of proton transfer and nucleophilic attack that are consistent with the experimental results and, to our knowledge, novel. In the future we hope to pursue studies combining both quantum mechanical and classical methods (e.g. molecular mechanics and dynamics) to more fully account for environmental effects.

REFERENCES: For Chapters 5-7

1. Alagona, G., Desmeules, P., Ghio, C., and P.A. Kollman (1984). J. Am. Chem. Soc. 106: 3623.
2. Alagona, G., Ghio, C., and P.A. Kollman (1986). J. Mol. Biol. 191: 23.
3. Alber, T.C., Davenport, Jr., R.C., Giammona, D.A., Lolis, E., Petsko, G.A., and D. Ringe (1987). Cold Spring Harbor Symp. Quan. Biol. LII: 603.
4. Albery, J.W., and J. R. Knowles (1976a). Biochemistry 15: 5627.
5. Albery, J.W., and J.R. Knowles (1976b). Biochemistry 15: 5631.
6. Asboth, B. and L. Polgar (1983). Biochemistry 22: 117.
7. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., and C.I. Pogson (1971). Cold Spring Harbor Symp. Quant. Biol. 36: 151.
8. Banner, D. W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offort, R.E., Priddle, J.D., and S.G. Waley (1975). Nature 255: 609.
9. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., and I.A. Wilson (1976). Biochem. Biophys. Res. Commun. 72: 146.
10. Bash, P.A., Singh, U.C., Langridge, R., and P.A. Kollman (1987). Science 236: 564.
11. Bash, P.A., Singh, U.C., Brown, F.K., Langridge, R., and P.A. Kollman (1987). Science 235: 574.
12. Bachovchin, W.W. and J.D. Roberts (1978). J. Am. Chem. Soc. 100: 8041-8047
13. Bachovchin, W.W. (1986). Biochemistry 25: 7751.
14. Baillargeon, M.W., Laskowski, M. Jr., Neves, D.E., Porubcan, M.A., Santini, R.E., and J.L. Markley (1980). Biochemistry 19: 5703.
15. Belasco, J.G., and J.R. Knowles (1980). Biochemistry 19: 472.

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

17. Bizzozero, S.A. and H. Dutler (1981). Bioorg. Chem. 10: 46.

18. Blacklow, S.C., Raines, R.T., Lim, W.A., Zamore, P.D., and J.R. Knowles (1988). Biochemistry 27: 1158.

19. Blacklow, S.C. and J.R. Knowles (1990). Biochemistry, submitted for publication.

20. Blow, D.M., Birktoft, J.J., and B.S. Hartley (1969). Nature 221: 337.

21. Brown, F.K. and P.A. Kollman (1987). J. Mol. Biol. 198: 533.

22. Campbell, I.D., Jones, R.B., Kiener, P.A., Richards, E., Waley, S.G., and R. Wolfenden (1978). Biochem. Biophys. Res. Commun. 83: 347.

23. Campbell, I.D., Jones, R.B., Kiener, P.A., and S.G. Waley (1979). Biochem. J. 179: 607.

24. Chambers, J. and R. Stroud (1977). Acta Crystallogr., Sect. B 33: 1824.

25. Collins, K.D. (1974). J. Biol. Chem. 249: 136.

26. Craik, C.S., Rozniak, S., Largeman, C. and W.J. Rutter (1987). Science 237: 909.

27. Dannenberg, J.J. (1988) J. Phys. Chem. 92: 6869.

28. Delbaere, L.T.J. and G.D. Brayer (1985). J. Mol. Biol. 183: 39.

29. Deslongchamps, P. (1983). Stereoelectronic Effects in Organic Chemistry, Pergamon Press, Oxford.

30. Dewar, M.J.S. and S.J. Kirschner (1971). J. Am. Chem. Soc. 93: 4290.

31. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., and J.J.P. Stewart (1985). J. Am. Chem. Soc. 107: 3902.

32. Dewar, M.J.S. and D.M. Storch (1985). Proc. Natl. Acad. Sci. USA 82: 2225.

33. Dewar, M.J.S. and K.M. Dieter (1986). J. Am. Chem. Soc. 108: 8075.

34. Elrod, J.P., Hogg, J.L., Quinn, D.M., Venkatasubban, K.S., and R.L. Schowen (1980). J. Am. Chem. Soc. 102: 3917.

35. Fersht, A.R., Blow, D.M. and J. Fastrez (1973). Biochemistry 12: 2035.

36. Fersht, A.R. and J. Sperling (1978). J. Mol. Biol., 74: 137.

37. Fersht, A.R. (1985). Enzyme Structure and Mechanism, W.H. Freeman & Co., New York.

38. Fink, A. (1987). In Enzyme Meechanisms, M.I. Page and A. Williams, Editors, Roy. Soc. Chemistry, London, 159.

39. Gandour, R.D. (1981). Bioorg. Chem. 10: 169.

40. Gorenstein, D.G., Luxon, B.A., and E.M. Goldfield (1980). J. Am. Chem. Soc. 102: 1757.

41. Gorenstein, D.G. and K. Taira (1984). Biophys. J. 46: 749.

42. Harvey, S.C. (1989). Proteins 5: 78.

43. Hayes, D.M. and P. Kollman (1979). In Catalysis in Chemistry and Biochemistry: Theory and Experiment, Pullman, B. Ed., D. Reidel Publishing Company, Boston, MA.

44. Henderson, R., Wright, C.S., Hess, G.P., and D.M. Blow, Cold Spring Harbor Symp. Quant. Biol. 36: 63-70 (1971).

45. Hermes, J.D., Blacklow, S.C., and J.R. Knowles (1987). Cold Spring Harbor Symp. on Quant. Biol. LIIA 597.

46. Hirohara, H., Bender, M.L., and R.S. Stark (1974). Proc. Natl. Acad. Sci. USA 71: 1643-1647.

47. Hirohara, H., Philipp, M. and M.L. Bender (1977). Biochemistry 16: 1573.

48. Hunkapiller, M.W., Smallcombe, S.H., Whitaker, D.R., and J.H. Richards (1973). Biochemistry 12: 4732.

49. Iyengar, R. and I. Rose (1981). Biochemistry 20: 1223.

50. Jarvis, L., Huang, C., Ferrin, T. and R. Langridge (1985), UCSF MIDAS, University of California, San Francisco.

51. Jorgensen, W.L., Chandrasekhar, J., Madura, J. Impey, R.W., and M.L. Klein (1983). J. Chem. Phys. 79: 926.

52. Knowles, J.R. and W.J. Albery (1977). Acc. Chem. Res. 10: 105.

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

54. Kollman, P.A. and D.M. Hayes (1981). J. Am. Chem. Soc. 103: 2955.

55. Komiyama, M. and M. Bender (1979). Proc. Natl. Acad. Sci. USA 76: 557.

56. Kossiakoff, A.A. and S.A. Spencer (1981). Biochemistry 20: 6462.

57. Kraut, J. (1977). Ann. Rev. Biochem. 46: 331.

58. Maister, S.G., Pett, C.P., Albery, W.J., and J.R. Knowles (1976). Biochemistry 15: 5607.

59. McIver, R.T.,Jr. and J.S. Miller (1974). J. Am. Chem. Soc. 96: 4323.

60. Merz, K.M., Jr. and B.H. Besler (1990). QCPE Bulletin, 10, 15, QCPE program #589.

61. Nakagawa, S. and H. Umeyama (1984). J. Mol. Biol. 179: 103.

62. Nickbarg, E.B., Davenport, R.C., Petsko, G.A., and J.R. Knowles (1988). Biochemistry 27: 5948.

63. Pethig, R. (1979). Dielectric and electronic properties of biological materials. John Wiley & Sons, Chichester.

64. Phillips, D.C., Rivers, P.S., Sternberg, M.J.E., Thornton, J.M. and I.A. Wilson (1977). Biochem. Soc. Tran. 5: 642.

65. Poulos, T.L., Alden, R.A., Freer, S.T., Birktoft, J.J., and J. Kraut (1976). J. Biol. Chem. 251: 1097.

66. Pozsgay, M., Szabo, G., Bajusz, S., Simonsson, R., Caspar, R., and P. Elodi (1981). Eur. J. Biochem. 115: 497.

67. Raines, R.T., Sutton, E.L., Straus, D.R., Gilbert, W. and J.R. Knowles (1986). Biochemistry 25: 7142.

68. Rao, S.N., Singh, U.C., Bash, P.A. and P.A. Kollman (1987). Nature 328: 551.

69. Rees, D.C. (1980). J. Mol. Biol. 141: 323.

70. Richard, J.P. (1984). J. Am. Chem. Soc. 106: 4926.

71. Richarz, R., Tschesche, H., and K. Wuthrich (1980). Biochemistry 19: 5711.

72. Rose, I.A. (1981). Phil. Trans. R. Soc. Lond. B 293: 131.

73. Russell, A.J., and A.R. Fersht (1987). Nature 328: 496.

74. Russell, A.J., Thomas, P.G., and A.R. Fersht (1987). J. Mol. Biol. 193: 803.

75. Ryckaert, J., Ciccotti, G. and H.J.C. Berendsen (1977). J. Comp. Phys. 23: 327.

76. Scheiner, S., Kleier, D.A. and W.N. Lipscomb (1975). Proc. Natl. Acad. Sci. USA 72: 2606.

77. Scheiner, S. and W. Lipscomb (1976). Proc. Natl. Acad. Sci. USA 73: 432.

78. Scheiner, S. and E. Hillebrand (1985). Proc. Natl. Acad. Sci. USA 82: 2741.

79. Schowen, R.L. (1988). In Principles of Enzyme Activity vol. 9, Molecular Structure and Energetics, Liebman, J.F. and A. Greenberg, editors, VCH Publishers, Weinheim, FRG.

80. Singh, U.C. and P.A. Kollman (1984). J. Comp. Chem. 5: 129.

81. Singh, U.C., Pattabiramin, N. and P.A. Kollman (1986a). Proc. Natl. Acad. Sci. USA 83: 6402.

82. Singh, U.C., Weiner, P.K., Caldwell, J., and P.A. Kollman (1986b). AMBER 3.0, University of California, San Francisco. AMBER version 3.1 was used for the work described in chapter 6; it is an in-house revision of version 3.0 by Jim Caldwell.

83. Singh, U.C., Brown, F.K., Bash, P.A., and P.A. Kollman (1987). J. Am. Chem. Soc. 109: 1607.

84. Sprang, S., Standing, T., Fletterick, R.J, Stroud, R.M., Finer-Moore, J., Xuong, N.-H., Hamlin, R., Rutter, W.J., and C.S. Craik (1987). Science 238: 905.

85. Stein, R.L., Elrod, J.P. and R.L. Schowen (1983). J. Am. Chem. Soc. 105: 2446.

86. Steitz, T.A. and R.G. Shulman (1982). Ann. Rev. Biophys. Bioeng. 11: 419.

87. Stewart, J.P. (1989). J. Comp. Chem. 10: 209.

88. Straus, D., Raines, R.T., Kawashima, E., Knowles, J.R. and W. Gilbert (1985). Proc. Natl. Acad. Sci. USA 82: 2272.

89. Stull, D.R. and J. Prophet (1971). JANAF Thermochemical Tables, NSRDS-NBS37.

90. Taira, K. and D.G. Gorenstein (1987). Bull. Chem. Soc. Jpn. 60: 3625.

91. Tsukada, H. and D.M. Blow (1985). J. Mol. Biol. 184: 703.

92. Umeyama, H., Imamura, A., Nagato, C. and M. Hanano (1973). J. Theor. Biol. 41: 485.

93. Umeyama, H., Nakagawa, S. and T. Kudo (1981). J. Mol. Biol. 150: 409.

94. Warshel, A. and S. Russell (1986). J. Am. Chem. Soc. 108: 6569.

95. Warshel, A., Naray-Szabo, G., Sussman, F. and J.-K. Hwang (1989). Biochemistry 28: 3629.

96. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, Jr., S., and P. Weiner (1984). J. Am. Chem. Soc. 106: 765.

97. Weiner, S.J., Kollman, P.A., Nguyen, D.T., and D.A. Case (1986a). J. Comp. Chem. 7: 230.

98. Weiner, S.J., Seibel, G.L. and P.A. Kollman (1986b). Proc. Natl. Acad. Sci. USA 83: 649.

99. Wells, J.A., Cunningham, B.C., Graycar, T.P., and D.A. Estell (1986). Philos. Trans. R. Soc. London, A. 317: 415.

100. Williams, I.H., J. Am. Chem. Soc. 109: 6299 (1987).

101. Wipff, G., Blaney, J., Weiner, P., Dearing, A. and P.A. Kollman (1983). 105: 997.

102. Wolfenden, R. (1969). Nature 223: 704.

103. Yamdagni, R. and P. Kebarle (1973). J. Am. Chem. Soc. 95: 4050.

APPENDIX I

Appendix I contains the source code for the sigmoidal dielectric function described in Chapter 3 and for the program written to analyze the peptide trajectories (ANALMD).

The background material and calibration of the sigmoidal dielectric function are described in Chapter 3. This function and its derivative, multiplied by R, are given below, where R is the distance between charges and S is a constant.

$$\varepsilon(R) = 78 - 38.5\, e^{-RS}\,[(RS)^2 + 2RS + 2]$$

$$\frac{d(\varepsilon(R).R)}{dr} = \frac{R\,(-78 + 38.5\, e^{-RS}[(RS)^2 - (RS)^3 + 2RS + 2])}{[78R - 38.5\, e^{-RS}(R^3S^2 + 2SR^2 + 2R)]^2}$$

The implementation of this function within AMBER was first tested by evaluating the potential energy between two charges separated by different distances. The energy as a function of distance and dielectric model is given below.

Dielectric Model	Potential Energy as a Function of Distance Between Charges (Å)		
	5	10	30
ε = R	13.3	3.3	0.4
ε = sig	4.2	0.7	0.1
ε = 80	0.8	0.4	0.1

As can be seen in the table above, electrostatic interactions are quickly attenuated with increasing distance for the sigmoidal dielectric function, while retaining the strength of the close interactions. The sigmoidal dielectric function has another advantage over the linear function when simulating small charged structures, long nonbonded cutoffs can be used. For example, in a test case of the dynamics of polyalanine (20 residues) with

charged termini, we found that the ends came together after approximately 10 ps with a 99 Å nonbonded cutoff. Thus, one is forced to use short cutoffs with ε=R for systems like this one. With the sigmoidal function, however, long cutoffs can be employed, which can be important since electrostatic interactions are long range.

The sigmoidal function was implemented differently on the VAX 8650 and the FPS 264 machines. On the VAX, it was found that a look-up table was the most efficient way of implementing this function (code follows). Initially a grid of 0.1 Å was used in generating the look-up table but it was found that 0.05 Å gave more accurate values without increasing the computation time. When the dielectric function was added in-line to the nonbonded portion of the program or as a subroutine, the computation time increased by a factor of 2.

For the FPS, the most efficient way of implementing the sigmoidal function was to calculate both the dielectric constant and its derivative in-line. This approach was approximately 50% faster than to use a look-up table. The code for this implementation follows and includes some original AMBER code to illustrate where the function is calculated. On both machines, with the most efficient manners of implementation, the linear function was twice as fast as the sigmoidal function.

```
c
c
c The code for implementation of the sigmoidal function
c on the VAX 8650 with look-up tables for the
c dielectric constant and its derivative
c follows.
c
c
      subroutine table
c
      real r,dr,rmax
c
c     ----- generate table of dielectric constants -----
c
c
c set up diel array for the range dr to rmax angstroms with spacing
c between intervals corresponding to dr
c
        common /dielarray/ diel(1500)
        common /derarray/ der(1500)
        data dr / 0.05 /
c
      rmax = 50.0
      imax = int(rmax/dr)
c
      do 1 ir = 1, imax
c
c get the distance corresponding to the index
c
         r = (ir)*dr
c
c calculate the corresponding dielectric for the current r
c
         call sigmoid (r,dielec)
         diel(ir) = dielec
1     continue

c
c     ----- generate table of derivatives -----
c
c
c set up diel array for the range dr to rmax angstroms with spacing
c between intervals corresponding to dr
c
c
      rmax = 50.0
      imax = int(rmax/dr)
c
      do 2 ir = 1, imax
c
c get the distance corresponding to the index
c
         r = (ir)*dr
c
c calculate the corresponding derivative for the current r
c
         call sigder (r,deriv)
         der(ir) = deriv
2     continue
      return
      end
c
c
C     sigmoidal function for epsilon
C
      SUBROUTINE sigmoid (R,dielec)
```

```
      DOUBLE PRECISION S,ep,dielec,dnen
C
      S = 0.3d0
      dnen = exp(-(R*S))
      ep = (78.d0-38.5d0*dnen*((R*S)**2+(2.d0*R*S)+2))
      dielec = EP * R
      RETURN
      END

C
C     sigmoidal function for epsilon to get derivative
C
      SUBROUTINE sigder (R,deriv)
      DOUBLE PRECISION S,dnum,dden,dnen,deriv
C
      S = 0.3d0
      dnen = exp(-(R*S))
      dnum = R*(-78.d0+38.5d0*dnen*((R*S)**2-(R*S)**3+2.d0*R*S+2))
      dden = (78.d0*R-38.5d0*dnen*((R**3*S**2)+2.d0*S*R**2+2.d0*R))**2
      deriv = dnum / dden
      RETURN
      END

c
c the following two functions are added in the main part of the
c nonbonded subroutine to get the dielectric constant and
c derivative from the existing arrays
c
c
      function getdiel(r)
c
      common /dielarray/ diel(1500)
      data dr /0.05/
c
c do a linear interpolation between the two closest distances
c
      index_lo = int(r/dr)
      index_hi = index_lo + 1
      r_lo = (index_lo) * dr
c
      frac = r - r_lo
      diel_lo = diel(index_lo)
      diel_hi = diel(index_hi)
      diel_range = diel_hi - diel_lo
c
      getdiel = diel_lo + (diel_range * frac/dr)
c
      return
      end
c
c
c
c
c
      function getder(r)
c
      common /derarray/ der(1500)
      data dr /0.05/
c
c do a linear interpolation between the two closest distances
c
      index_lo = int(r/dr)
      index_hi = index_lo + 1
      r_lo = (index_lo) * dr
c
      frac = r - r_lo
```

```
      der_lo = der(index_lo)
      der_hi = der(index_hi)
      der_range = der_hi - der_lo
c
      getder = der_lo + (der_range * frac/dr)
c
      return
      end
c
```

```
c
c
c  The code for implementation of the sigmoidal function
c  on the FPS follows. The entire subroutine for calculating
c  nonbonded interactions within AMBER is given. My
c  modifications are delineated by astericks.
c
c
      SUBROUTINE NNBOND(NATOM,NPAIR,IAR1,FAR2,IAC,ICO,X,F,CN1,CN2,ASOL,
     +                  BSOL,HBCUT,CG,XCHRG,ENB,EHB,EEL,XWIJ,RW,JPW,XRC,
     +                  FXW,FYW,FZW,VIR,NTYPES,IPTATM)
      DOUBLE PRECISION R2,R6,R12,F1,F2,G
*     DOUBLE PRECISION EP,DNEN,EPR,DNUM,DDEN,DER,S
      DOUBLE PRECISION ENB,EHB,EEL,E,ENBT,EHBT,EELT
      LOGICAL VIRIAL
C
C     ----- ROUTINE TO EVALUATE THE ENERGY AND GRADIENT DUE TO
C           THE NON-BONDED INTERACTIONS (DIST. DEP. DIELECTRIC) -----
C
      COMMON/SETBOX/BOX(3),BETA,BOXH(3),BOXHM,BOXHM2,BOXOH,BOXOQ,
     +              COSB,COSB2,NTM,NTB
C
      DIMENSION IAR1(2),FAR2(2),IAC(2),X(3,2),F(3,2),CG(2)
      DIMENSION XCHRG(2),CN1(2),CN2(2),ASOL(2),BSOL(2),HBCUT(2)
      DIMENSION XWIJ(3,2),RW(2),ICO(2),JPW(2),XRC(3,2),VIR(2),R(20000)
      DIMENSION FXW(2),FYW(2),FZW(2)
      DIMENSION VIRT(3)
C
      LPAIR = 0
      LPACK = 1
      LIM = NATOM-1
      INDHB = NATOM
      ENBT = 0.0E0
      EELT = 0.0E0
      EHBT = 0.0E0
      VIRIAL = IABS(NTB).GE.2
      VIRT(1) = 0.0E0
      VIRT(2) = 0.0E0
      VIRT(3) = 0.0E0
C
C     ----- TRANSFORM CARTESIAN TO OBLIQUE IF NECESSARY -----
C
      CALL VCLR(JPW,1,INDHB)
      IF(NTM.NE.0) CALL TRACO(NATOM,0,X,BETA,1)
C
C     ----- MAIN LOOP OVER THE ATOMS -----
C
      DO 900 I = 1,LIM
C
      LPR = IAR1(I)
      LHB = IAR1(I+INDHB)
      NPR = LPR+LHB
      IACI = NTYPES*(IAC(I)-1)
      CGI = CG(I)
      NPACK = NPR/4+1
C
C     ----- IF NPR.EQ.0 THEN SKIP TO NEXT ATOM -----
C
      IF(NPR.EQ.0) GO TO 880
C
C     ----- UNPACK THE STUFF -----
C
      CALL VUP16(FAR2(LPACK),1,RW,1,NPACK)
      CALL VIFIX(RW,1,JPW,1,NPR,0)
C
C     ----- GENERATE THE IJ VECTOR FOR NPR PAIRS -----
```

```
C
      DO 100 JN = 1,NPR
      J = JPW(JN)
      XWIJ(1,JN) = X(1,I)-X(1,J)
      XWIJ(2,JN) = X(2,I)-X(2,J)
      XWIJ(3,JN) = X(3,I)-X(3,J)
  100 CONTINUE
C
C     ----- IF BOUNDARY CONDITION IS PRESENT PUT IT -----
C
      IF(NTB.EQ.0) GO TO 140
C
      DO 230 JN = 1,NPR
      IF(JPW(JN).LE.IPTATM) GO TO 230
      DO 220 M = 1,3
      IF(XWIJ(M,JN) .LT. BOXH(M)) GO TO 200
      XWIJ(M,JN) = XWIJ(M,JN)-BOX(M)
      GO TO 220
  200 IF(XWIJ(M,JN) .GE. -BOXH(M)) GO TO 220
      XWIJ(M,JN) = XWIJ(M,JN)+BOX(M)
  220 CONTINUE
  230 CONTINUE
      DO 120 JN = 1,NPR
      RW(JN) = XWIJ(1,JN)**2+XWIJ(2,JN)**2+XWIJ(3,JN)**2
  120 CONTINUE
C
      IF(NTB.GT.0) GO TO 250
      DO 240 JN = 1,NPR
      RW(JN) = RW(JN)+AMIN1(0.0E0,BOXOQ-ABS(XWIJ(1,JN))-ABS(XWIJ(2,JN))
     +          -ABS(XWIJ(3,JN)))*BOX(1)
  240 CONTINUE
      GO TO 270
C
  250 CONTINUE
      IF(NTM.EQ.0) GO TO 270
      DO 260 JN = 1,NPR
      RW(JN) = RW(JN)+COSB2*XWIJ(1,JN)*XWIJ(3,JN)
  260 CONTINUE
C
  270 CONTINUE
      CALL VRECIP(RW,1,RW,1,NPR)
C
C     ----- TRANSFORM TO CARTESIAN IF NEEDED -----
C
      IF(NTM.NE.0) CALL TRACO(NPR,0,XWIJ,BETA,-1)
      GO TO 180
C
C     ----- NO PERIODICITY -----
C
  140 CONTINUE
      DO 160 JN = 1,NPR
      RW(JN) = XWIJ(1,JN)**2+XWIJ(2,JN)**2+XWIJ(3,JN)**2
 *    R(JN) = SQRT(RW(JN))
  160 CONTINUE
      CALL VRECIP(RW,1,RW,1,NPR)
  180 CONTINUE
C
C     ----- NOW CALCULATE THE ENERGY CONTRIBUTIONS -----
C           with the sigmoidal dielectric function
C           in-line.
C
      IF(LPR.LE.0) GO TO 320
      DO 300 JN = 1,LPR
          J = JPW(JN)
          IC = ICO(IACI+IAC(J))
  *       S = 0.3d0
```

```
*       dnen = exp(-(R(JN)*S))
*       ep = (78.d0-38.5d0*dnen*((R(JN)*S)**2+(2.d0*R(JN)*S)+2))
*       epr = EP * R(JN)
*       G = CGI*CG(J)/ epr
        EELT = EELT+G
*       dnum = R(JN)*(-78.d0+38.5d0*dnen*((R(JN)*S)**2-(R(JN)*S)**3+
*  +           2.d0*R(JN)*S+2))
*       dden = (78.d0*R(JN)-38.5d0*dnen*((R(JN)**3*S**2)+
*  +           2.d0*S*R(JN)**2+2.d0*R(JN)))**2
*       der = dnum / dden
        DF2 = CGI*CG(J)*DER
        R6 = RW(JN)*RW(JN)*RW(JN)
        R12 = R6*R6
        F1 = CN1(IC)*R12
        F2 = CN2(IC)*R6
        ENBT = ENBT + (F1-F2)
        DF1 = (-12.*F1 + 6.*F2)
        DF = (DF1+DF2)*RW(JN)
C
C     ----- CALCULATE THE GRADIENT -----
C
        FXW(JN) = XWIJ(1,JN)*DF
        FYW(JN) = XWIJ(2,JN)*DF
        FZW(JN) = XWIJ(3,JN)*DF
  300 CONTINUE
  320 CONTINUE
C
C     ----- H-BOND PAIRS 10-12 POTENTIAL -----
c     with the sigmoidal dielectric function in-line.
C
      IF(LHB.LE.0) GO TO 420
      DO 400 JN = LPR+1,NPR
        J = JPW(JN)
        IC = -ICO(IACI+IAC(J))
*       S = 0.3d0
*       dnen = exp(-(R(JN)*S))
*       ep = (78.d0-38.5d0*dnen*((R(JN)*S)**2+(2.d0*R(JN)*S)+2))
*       epr = EP * R(JN)
*       G = CGI*CG(J)/ epr
        EELT = EELT+G
*       dnum = R(JN)*(-78.d0+38.5d0*dnen*((R(JN)*S)**2-(R(JN)*S)**3+
*  +           2.d0*R(JN)*S+2))
*       dden = (78.d0*R(JN)-38.5d0*dnen*((R(JN)**3*S**2)+
*  +           2.d0*S*R(JN)**2+2.d0*R(JN)))**2
*       der = dnum / dden
        DF2 = CGI*CG(J)*DER
        R10 = RW(JN)**5
        F1 = ASOL(IC)*R10*RW(JN)
        F2 = BSOL(IC)*R10
        EHBT = EHBT+F1-F2
        DF1 = (-12.0E0*F1+10.0E0*F2)
        DF = (DF1+DF2)*RW(JN)
C
C     ----- CALCULATE THE GRADIENT -----
C
        FXW(JN) = XWIJ(1,JN)*DF
        FYW(JN) = XWIJ(2,JN)*DF
        FZW(JN) = XWIJ(3,JN)*DF
  400 CONTINUE
  420 CONTINUE
C
C     ----- SUMUP THE GRADIENT -----
C
c      CALL SVE(FXW,1,DUMX,NPR)
c      CALL SVE(FYW,1,DUMY,NPR)
c      CALL SVE(FZW,1,DUMZ,NPR)
```

```
      dumx = 0.
      dumy = 0.
      dumz = 0.
      do 430 icnt = 1,npr
         dumx = dumx + fxw(icnt)
         dumy = dumy + fyw(icnt)
         dumz = dumz + fzw(icnt)
 430  continue
      F(1,I) = F(1,I)-DUMX
      F(2,I) = F(2,I)-DUMY
      F(3,I) = F(3,I)-DUMZ
      DO 480 JN = 1,NPR
      J = JPW(JN)
      F(1,J) = F(1,J)+FXW(JN)
      F(2,J) = F(2,J)+FYW(JN)
      F(3,J) = F(3,J)+FZW(JN)
  480 CONTINUE
C
C     ----- CALCULATE THE VIRIAL IF NTB.EQ.2 -----
C
      IF(.NOT.VIRIAL) GO TO 560
      DO 540 JN = 1,NPR
      J = JPW(JN)
      VIRT(1) = VIRT(1)+FXW(JN)*(XWIJ(1,JN)-XRC(1,I)+XRC(1,J))
      VIRT(2) = VIRT(2)+FYW(JN)*(XWIJ(2,JN)-XRC(2,I)+XRC(2,J))
      VIRT(3) = VIRT(3)+FZW(JN)*(XWIJ(3,JN)-XRC(3,I)+XRC(3,J))
  540 CONTINUE
  560 CONTINUE
C
C     ----- END OF PAIRS INVOLVING ATOM I -----
C
  880 CONTINUE
      LPACK = LPACK+NPACK
  900 CONTINUE
C
      ENB = ENBT
      EEL = EELT
      EHB = EHBT
      VIR(1) = VIRT(1)*0.50E0
      VIR(2) = VIRT(2)*0.50E0
      VIR(3) = VIRT(3)*0.50E0
C
C     ----- TRANSFORM THE OBLIQUE COORDINATES TO CARTESIAN IF NEEDED
C
      IF(NTM.NE.0) CALL TRACO(NATOM,0,X,BETA,-1)
      RETURN
      END
```

```
      program analmd
c
c     this program reads in coordinates from md generated input
c      files (e.g. 10f8.3 format). Then distances, angles, dihedrals
c      are calculated as a function of time if desired.
c      There is also a call to use subroutine conf from the
c      dihedral portion of the program. if this subroutine is
c      called then you will get output about probabilities
c      of helix coil transitions, helix length, and percentages of
c      time the structure spends in different regions of conforma-
c      tional space. if the distance, angle or dihedral routines
c      are called then statistical information like averages, rms
c      fluctuations etc are given in the output file. this
c      program also looks at hydrogen bonds and will print out
c      information about when h-bonds break during the simulation
c      and how long they remain broken, among other things. this
c      option geneerates a lot of output if the hydrogen bonds are
c      broken a good deal of of the simulation time. also,
c      there is an option to look at hydrogen bonding to aromatic
c      rings. to compare to nmr NOESY crosspeaks you can also
c      print out r-6 weighted nonbonded distances. also, cross
c      correlation coefficients for pairs of dihedrals can be
c      calculated. plot files can be printed for any of the
c      options mentioned above.
c
c      the dihedral subroutine is based on that of allison howard, with
c      some minor changes and the dihavg subroutine using
c      circular statistics was written by david spellmeyer,
c      again minor changes were made. the angular variance is essentially
c      equal to the rms fluctuation (the differ on order of a
c      tenth of a degree).
c
c
c
      integer nbond,dopt,npoints
      integer nhbond,hbopt,coropt,ncor
      integer cntopt,ncntr
      integer nmropt,nnmr,sub
      integer nang,aopt,bopt,ndih
      integer natoms,ntime,maxtim,maxat
      integer at1,at2,at3,at4
      integer at5,at6,at7,at8
      integer angat1,angat2,cent1,cent2
      integer maxres,copt,breg
      integer ncf,nt2,nt3,nt4
c
      parameter(maxtim=4000,maxat=120,maxres=20)
c
      real c(maxtim,maxat*3)
      real val(maxtim),dis(maxtim),val2(maxtim)
      real theta(maxres*2,maxtim)
      real dt,colow,cohigh,cdlow,cdhigh,sumco
c
      common/ xcoord/x(maxtim,maxat)
      common/ ycoord/y(maxtim,maxat)
      common/ zcoord/z(maxtim,maxat)
c
c
c     open the necessary files. unit 2=input file. unit 3=coord file.
c
c
      open(unit=2,status='old')
      open(unit=3,status='old')
c
c     open(unit=2,file='analin',status='old')
```

```
c     open(unit=3,file='md7.crd',status='old')
c     open(unit=6,file='anal.out',status='new')
c     open(unit=5,file='DIST.PLT',status='new')
c     open(unit=4,file='ANG.PLT',status='new')
c     open(unit=7,file='DIH.PLT',status='new')
c     open(unit=8,file='HIST.PLT',status='new')
c     open(unit=9,file='FRAC.PLT',status='new')
c     open(unit=10,file='HB.PLT',status='new')
c     open(unit=11,file='CANG.PLT',status='new')
c     open(unit=12,file='CDIS.PLT',status='new')
c     open(unit=13,file='NOE.PLT',status='new')
c     open(unit=14,file='CORR.PLT',status='new')
c     open(unit=15,file='md2.crd',status='old')
c     open(unit=16,file='md3.crd',status='old')
c     open(unit=17,file='md4.crd',status='old')
c
c
c     Read first two lines of input file.
c
      read(2,'(a80)')dummy
      read(2,'(2i6,f5.2,4i6)')natoms,ntime,dt,ncf,nt2,nt3,nt4
c
c     Next read in the coordinates.
c
      read(3,'(a80)')dummy
c
c
      npoints = 1
      do 15 k=1,ntime
         read(3,'(10F8.3)',end=16) (c(k,j),j=1,3*natoms)
         npoints = npoints +  1
15    continue
16    continue
c
      if (ncf .ge. 2) then
        open(unit=15,status='old')
        read(15,'(a80)')dummy
        do 17 k=ntime+1,nt2
          read(15,'(10f8.3)',end=18)(c(k,j),j=1,3*natoms)
          npoints = npoints + 1
17      continue
18      continue
      endif
c
      if (ncf .ge. 3) then
        open(unit=16,status='old')
        read(16,'(a80)')dummy
        do 19 k=nt2+1,nt3
          read(16,'(10f8.3)',end=20)(c(k,j),j=1,3*natoms)
          npoints = npoints + 1
19      continue
20      continue
      endif
c
      if (ncf .ge. 4) then
        open(unit=17,status='old')
        read(17,'(a80)')dummy
        do 21 k=nt3+1,nt4
          read(17,'(10f8.3)',end=22)(c(k,j),j=1,3*natoms)
          npoints = npoints + 1
21      continue
22      continue
      endif
c
         npoints = npoints - 1
         write (6,*) 'coord sets read',npoints
```

```
c
      do 25 k = 1, npoints
        j = 1
          do 23 i=1,natoms
            x(k,i) = c(k,j)
            y(k,i) = c(k,j + 1)
            z(k,i) = c(k,j + 2)
            j = j + 3
23         continue
25     continue
c
c
c-----DISTANCES-------------
c
c
      read(2,'(a8,3i5,2f5.2)')dummy,bopt,nbond,breg,colow,cohigh
c
        if (bopt .eq. 0) goto 41
c
          do 40 j=1,nbond
            read(2,'(2i5)')at1,at2
            call distance(npoints,at1,at2,val,k)
c
c     output statistics to output file, meanfl subroutine
c
              write(6,*)'statistics for distances between the
     & following atoms:'
              write(6,*)at1,at2
              call meanfl(npoints,val,k)
              write(6,*)
c
c     determine % of time in region specified by colow and
c      cohigh if breg=1
c
              sumco = 0
              if (breg .eq. 0) goto 35
                do 34 k=1,npoints
                  if (val(k) .ge. colow .and. val(k) .le. cohigh) then
                    sumco = sumco + 1
                  endif
34                continue
                  write(6,*)'% of time within cutoff',colow,cohigh,
     & sumco*100/npoints
                  write(6,*)
35    continue
c
c     open plot file
c
              if (bopt .eq. 2) then
                write(5,*)'distances for ',at1,at2
                do 36 k=1,npoints
                  write(5,'(f7.2,f8.3)')k*dt,val(k)
36                continue
                  write(5,*)
              endif
c
40    continue
41    continue
c
c
c-----ANGLES-------------
c
c     start by reading in information from input file
c
      read(2,'(a8,3i5,2f7.2)')dummy,aopt,nang,breg,colow,cohigh
c
```

```
      if (aopt .eq. 0) goto 51
c
      do 50 j=1,nang
        read(2,'(3i5)')at1,at2,at3
        call angle(npoints,at1,at2,at3,val,k)
c
c     output statistics to output file, meanfl subroutine
c
        write(6,*)'statistics for angles between the
     & following atoms:'
        write(6,*)at1,at2,at3
        call meanfl(npoints,val,k)
        write(6,*)
c
c     determine % of time in region specified by colow and
c      cohigh if breg=1
c
              sumco = 0
              if (breg .eq. 0) goto 45
                do 44 k=1,npoints
                  if (val(k) .ge. colow .and. val(k) .le. cohigh) then
                    sumco = sumco + 1
                  endif
44              continue
                  write(6,*)'% of time within cutoff',colow,cohigh,
     & sumco*100/npoints
                  write(6,*)
45    continue
c
c     open plot file
c
        if (aopt .eq. 2) then
          write(4,*)'angles for ',at1,at2,at3
          do 46 k=1,npoints
            write(4,'(f7.2,f8.3)')k*dt,val(k)
46        continue
          write(4,*)
        endif
c
50    continue
51    continue
c
c
c
c-----DIHEDRALS---------------
c
c     start by reading in information from input file
c
      read(2,'(a8,3i5)')dummy,dopt,ndih,copt
c
      if (dopt .eq. 0) goto 62
c
      do 60 j=1,ndih
        read(2,'(4i5)')at1,at2,at3,at4
        call dihed(npoints,at1,at2,at3,at4,val,k)
c
c     open plot file
c
        if (dopt .eq. 2) then
          write(7,*)'dihedral angle for ',at1,at2,at3,at4
          do 58 k=1,npoints
            write(7,'(f7.2,x,f8.3)')k*dt,val(k)
58        continue
          write(7,*)
        endif
c
```

```
c     put dihedral angles into 1-360 degree convention and
c     set up array to be used for conformational analysis
c     based on phi and psi values. copt = 1 if this
c     option is in effect.
c
        do 59 k=1,npoints
          if (val(k) .lt. 0.0) val(k) = val(k) + 360.0d0
          if (copt .eq. 1) theta(j,k) = val(k)
59      continue
c
c
c     output statistics to output file, dihavg subroutine
c
        write(6,*)'statistics for dihedral angle between the
     &   following atoms:'
        write(6,*)at1,at2,at3,at4
        call dihavg(npoints,val,k)
c
60    continue
c
c     now call subroutine for phi and psi conformational analysis
c       passover the call statement if copt not wanted (.ne. 1)
c
        if (copt .eq. 1) then
          call conf(ndih,npoints,dt,theta,j,k)
        endif
62    continue
c
c
c-----HYDROGEN BONDS--------------------------------
c
c
c     read in information from input files.
c
      read(2,'(a8,2i5)')dummy,hbopt,nhbond
c
      if (hbopt .eq. 0) goto 71
c
      do 70 j=1,nhbond
            read(2,'(2i5)')at1,at2
            call distance(npoints,at1,at2,val,k)
c
c     output statistics to meanfl subroutine
c
              write(6,*)'statistics for distances between the
     & following hydrogen bonding atoms:'
              write(6,*)at1,at2
              call meanfl(npoints,val,k)
              write(6,*)
c
c     open plot file
c
            if (hbopt .eq. 2) then
              write(10,*)'H-B distances for ',at1,at2
              do 69 k=1,npoints
                write(10,'(f7.2,f8.3)')k*dt,val(k)
69            continue
              write(10,*)
            endif
c
c     now call subroutine hbond to determine statistics
c     for hydrogen bonds.
c
            write(6,*)'H-B statistics for',at1,at2
            call hbond(npoints,dt,val,k)
            write(6,*)
```

```
c
70    continue
71    continue
c
c
c-------CENTER OF RING CALCULATION----------
c
c
c     read in information from input files.
c
      read(2,'(a8,3i5,2f7.2,2f5.2)')dummy,cntopt,ncntr,breg,
     & colow,cohigh,cdlow,cdhigh
c
      if (cntopt .eq. 0) goto 81
c
      do 80 j=1,ncntr
            read(2,'(4i5)')angat1,angat2,cent1,cent2
            call center(npoints,angat1,angat2,cent1,cent2,val,
     &           dis,k)
c
c     output statistics to meanfl subroutine
c
            write(6,*)'statistics for angles between the
     & following hydrogen bonding'
            write(6,*)'atoms and center of ring:'
            write(6,*)angat1,angat2,cent1,cent2
            call meanfl(npoints,val,k)
            write(6,*)
c
c     determine % of time in region specified by colow and
c      cohigh if breg=1
c
            sumco = 0
            if (breg .eq. 0) goto 75
              do 72 k=1,npoints
                if (val(k) .ge. colow .and. val(k) .le. cohigh) then
                  sumco = sumco + 1
                    endif
72              continue
                write(6,*)'% of time within cutoff',colow,cohigh,
     & 100*sumco/npoints
                write(6,*)
75    continue
c
c     open plot file
c
          if (cntopt .eq. 2) then
            write(11,*)'angles for atoms and ring',angat1,
     & angat2,cent1,cent2
            do 73 k=1,npoints
              write(11,'(f7.2,f8.3)')k*dt,val(k)
73          continue
            write(11,*)
          endif
c
c      reassign val(k) to send to meanfluc subroutine
c
            do 74 k=1,npoints
                val(k) = dis(k)
74          continue
c
            write(6,*)'statistics for distances between the
     & hydrogen bonding atoms and center of ring:'
            write(6,*)angat1,cent1,cent2
            call meanfl(npoints,val,k)
            write(6,*)
```

```
c
c     determine % of time in region specified by colow and
c      cohigh if breg=1
c
              sumco = 0
              if (breg .eq. 0) goto 77
                do 76 k=1,npoints
                  if (val(k) .ge. cdlow .and. val(k) .le. cdhigh) then
                    sumco = sumco + 1
                      endif
76                continue
                  write(6,*)'% of time within cutoff',cdlow,cdhigh,
     & sumco/npoints
                  write(6,*)
77    continue
c
c     open plot file
c
            if (cntopt .eq. 2) then
              write(12,*)'distances for h-bonding atom and ring',
     & angat1,cent1,cent2
              do 78 k=1,npoints
                write(12,'(f7.2,f8.3)')k*dt,val(k)
78            continue
              write(12,*)
            endif
c
c
80    continue
81    continue
c
c
c-----WEIGHTED DISTANCES FOR NMR COMPARISON-------------
c
c
      read(2,'(a8,2i5)')dummy,nmropt,nnmr
c
        if (nmropt .eq. 0) goto 91
c
          do 90 j=1,nnmr
            read(2,'(3i5)')at1,at2,sub
            call distance(npoints,at1,at2,val,k)
c
c     open plot file
c
              if (nmropt .eq. 2) then
                write(13,*)'weighted distances for ',at1,at2
                do 86 k=1,npoints
                  write(13,'(f7.2,f8.3)')k*dt,val(k)
86              continue
                write(13,*)
              endif
c
c     output mean and weighted distance statistics from
c       r3 subroutine
c
              write(6,*)'statistics for distances between the
     & following atoms'
              write(6,*)'corresponding to NOESY crosspeaks:'
              write(6,*)at1,at2
              call r3(npoints,sub,val,k)
              write(6,*)
c
90    continue
91    continue
c
```

```
c
c-----CROSS CORRELATION COEFFICIENTS OF DIHEDRALS-------------
c
c
      read(2,'(a8,2i5)')dummy,coropt,ncor
c
        if (coropt .eq. 0) goto 101
c
          do 100 j=1,ncor/2
            read(2,'(4i5)')at1,at2,at3,at4
            read(2,'(4i5)')at5,at6,at7,at8
c
            call dihed(npoints,at1,at2,at3,at4,val,k)
c
            do 95 k=1,npoints
               val2(k) = val(k)
95          continue
c
       write(6,*)'Cross correlation coefficent and mean values for:'
       write(6,*)'a is',at5,at6,at7,at8
       write(6,*)'b is',at1,at2,at3,at4
c
            at1 = at5
            at2 = at6
            at3 = at7
            at4 = at8
c
            call dihed(npoints,at1,at2,at3,at4,val,k)
c
            call corr(npoints,coropt,dt,val,val2,k)
c
c
100    continue
101    continue
c
c
      stop
      end
c
c-------------------------------------------------------
      subroutine distance(npoints,at1,at2,val,k)
c-------------------------------------------------------
c
c
      integer npoints,at1,at2,maxat,maxtim
c
      parameter(maxtim=4000,maxat=120)
c
      real xat1,xat2,yat1,yat2,zat1,zat2
      real val(maxtim)
c
      common/ xcoord/x(maxtim,maxat)
      common/ ycoord/y(maxtim,maxat)
      common/ zcoord/z(maxtim,maxat)
c
c
      do 10 k=1,npoints
c
c     i=atom number 1, at1
c
        xat1 = x(k,at1)
        yat1 = y(k,at1)
        zat1 = z(k,at1)
c
c     i=atom number 2, at2
c
```

```
        xat2 = x(k,at2)
        yat2 = y(k,at2)
        zat2 = z(k,at2)
c
c     Calculate the distance
c
        dist = (xat2-xat1)**2 + (yat2-yat1)**2 +
     &            (zat2-zat1)**2
        val(k) = sqrt(dist)
10    continue
c
      return
      end
c
c
c-----------------------------------------------------
      subroutine angle(npoints,at1,at2,at3,val,k)
c-----------------------------------------------------
c
c
      integer k,npoints,at1,at2,at3,maxtim,maxat
c
      parameter(maxtim=4000,maxat=120)
c
      real xat1,xat2,yat1,yat2,zat1,zat2
      real xat3,yat3,zat3
      real bot,d12,d13,d23,ang
      real val(maxtim)
c
      common/ xcoord/x(maxtim,maxat)
      common/ ycoord/y(maxtim,maxat)
      common/ zcoord/z(maxtim,maxat)
c
c
      do 10 k=1,npoints
c
c     i=atom number 1, at1
c
        xat1 = x(k,at1)
        yat1 = y(k,at1)
        zat1 = z(k,at1)
c
c     i=atom number 2, at2
c
        xat2 = x(k,at2)
        yat2 = y(k,at2)
        zat2 = z(k,at2)
c
c     i=atom number 3, at3
c
        xat3 = x(k,at3)
        yat3 = y(k,at3)
        zat3 = z(k,at3)
c
c     Calculate the distances
c
        d12 = sqrt((xat2-xat1)**2 + (yat2-yat1)**2 +
     &            (zat2-zat1)**2)
c
        d23 = sqrt((xat3-xat2)**2 + (yat3-yat2)**2 +
     &            (zat3-zat2)**2)
c
        d13 = sqrt((xat3-xat1)**2 + (yat3-yat1)**2 +
     &            (zat3-zat1)**2)
c
c
```

```
c      now calculate the angle
c
         bot = 2* d12 * d23
         ang = d12**2 + d23**2
         ang = (ang - d13**2) / bot
         val(k) = (acos(ang))*180.0 / 3.141592654
c
10     continue
c
       return
       end
c
c
c-------------------------------------------------------------------------
        subroutine meanfl(npoints,val,k)
c-------------------------------------------------------------------------
c
c
       integer npoints,maxtim,maxat
c
       parameter(maxtim=4000,maxat=120)
c
       real val(maxtim),sum,flmean,mean,flsum
       real fl2mean,fl2sq,fl2sum,min,max
       real abfl(maxtim),fluc(maxtim),fluc2(maxtim)
       real maxdev, mindev, maxx,dummy,valf
c
       sum = 0
       max = 0
       min = 360
       do 50 k=1,npoints
          if (val(k) .gt. max) then
            max = val(k)
                elseif (val(k) .lt. min) then
                 min = val(k)
                   endif
          sum = sum + val(k)
50     continue
        mean = sum / npoints
        flsum = 0
        fl2sum = 0
       do 55 k=1,npoints
           fluc(k) = val(k) - mean
            if (fluc(k) .ge. 0.0) then
               abfl(k) = fluc(k)
                 else
                  abfl(k) = - fluc(k)
                    endif
           fluc2(k)=abfl(k)**2
           flsum = flsum + abfl(k)
           fl2sum = fl2sum + fluc2(k)
55      continue
            flmean = flsum / npoints
            fl2mean = fl2sum / npoints
            fl2sq = sqrt(fl2mean)
c
c
        maxdev = max - mean
        mindev = min - mean
        if (maxdev .lt. 0) then
          maxdev = -maxdev
        else
         maxdev = maxdev
        endif
        if (mindev .lt. 0) then
         mindev = -mindev
```

```
        else
          mindev = mindev
        endif
c
c
        if (maxdev .gt. mindev) then
          maxx=maxdev
        else
          maxx=mindev
        endif
c
        k = npoints
        valf = val(k)
c
c
c         write (6,*)'npoints y',npoints
          write (6,*)'mean y',mean
          write (6,*)'final value of y',valf
          write (6,*)'mean fluctuation y',flmean
          write (6,*)'rms fluctuation y',fl2sq
          write (6,*)'max value',max
          write (6,*)'max deviation',maxdev
          write (6,*)'min value', min
          write (6,*)'min deviation', mindev
          write (6,*)'max dev from mean', maxx
c
      return
      end
c
c-----------------------------------------------------------------------
      subroutine dihed(npoints,at1,at2,at3,at4,val,k)
c-----------------------------------------------------------------------
c
c
      integer       j, k, at1, at2, at3, at4, npoints
      integer       maxtim,maxat
c
      parameter(maxtim=4000,maxat=120)
c
      real          a1, a2, a3, b1, b2, b3, c1, c2, c3, d1, d2, d3
      real          e1, e2, e3, f1, f2, f3
      real          rnorm1, rnorm2, adotb, costhe, t, test
      real          pi, conver, val(maxtim), th
      real          x1,x2,x3,x4,y1,y2,y3,y4,z1,z2,z3,z4
c
c
      common/ xcoord/x(maxtim,maxat)
      common/ ycoord/y(maxtim,maxat)
      common/ zcoord/z(maxtim,maxat)
c
      pi     = 6.0 * asin(0.5d0)
      conver = 360.0d0 / (2.0d0 * pi)
c
c
      do 10 k=1,npoints
c
c     ---- Generate vectors ----
c
      x1 = x(k,at1)
      y1 = y(k,at1)
      z1 = z(k,at1)
c
      x2 = x(k,at2)
      y2 = y(k,at2)
      z2 = z(k,at2)
c
```

```
      x3 = x(k,at3)
      y3 = y(k,at3)
      z3 = z(k,at3)
c
      x4 = x(k,at4)
      y4 = y(k,at4)
      z4 = z(k,at4)
c
      a1 = x2 - x1
      a2 = y2 - y1
      a3 = z2 - z1
c
      b1 = x2 - x3
      b2 = y2 - y3
      b3 = z2 - z3
c
      c1 = x3 - x2
      c2 = y3 - y2
      c3 = z3 - z2
c
      d1 = x3 - x4
      d2 = y3 - y4
      d3 = z3 - z4
c
c     ---- Evaluate the cross product ----
c
      e1 = a2 * b3 - a3 * b2
      e2 = a3 * b1 - a1 * b3
      e3 = a1 * b2 - a2 * b1
c
      f1 = c2 * d3 - c3 * d2
      f2 = c3 * d1 - c1 * d3
      f3 = c1 * d2 - c2 * d1
c
c     ---- Calculate norms ----
c
      rnorm1 = sqrt(e1**2 + e2**2 + e3**2)
      rnorm2 = sqrt(f1**2 + f2**2 + f3**2)
c
c     ---- Calculate the dot product ----
c
      adotb = e1 * f1 + e2 * f2 + e3 * f3
c
c     ---- Calculate the dihedral ----
c
      costhe = adotb / (rnorm1 * rnorm2)
c
c     test---as have problems with invalid
c      argument to math library for one
c      dihedral.
c
      if (costhe .gt. 1.0) costhe = 1.0
      if (costhe .lt. -1.0) costhe = -1.0
c
      th  = acos(costhe)
      t     = conver * th
      test  = a1 * f1 + a2 * f2 + a3 * f3
      if (test .gt. 0.0d0) t = 0.0d0 - t
      val(k) = t
c
c
c
10    continue
c
      return
      end
```

```
c
c-----------------------------------------------------------------
      subroutine dihavg(npoints,val,k)
c-----------------------------------------------------------------
c
c
      integer maxtim
c
      parameter(maxtim=4000)
c
      real val(maxtim),dihd(72),angle(72)
c
c        initialize variables
c
      degtorad=3.1415927/180.0
c
c        read in the dihedral angles
c        in 0 to 360 convention and
c        convert to radians
c
      do 10 k=1,npoints
        val(k) = val(k)*degtorad
10    continue
c
c
c
c       nbins: the number of bins to use for the dihedral
c              angle conformation population
c       amean : the mean of the angles in the array phi
c       r     : the mean vector length of the array phi
c       s     : the mean variance of the array phi
c       angle : the midpoint of the bin for counting
c       dihd  : the
c
c the mean angle of the sample is described by the following equation
c   amean  =       arctan(ybar/xbar) if xbar > 0
c            180 + arctan(ybar/xbar) if xbar < 0
c   where xbar = sum(cosine(phi[i]))/number points
c   where ybar = sum(  sine(phi[i]))/number points
c
c exceptions occur when:
c         xbar = 0 and ybar > 0  ==> phi = 90 degrees
c         xbar = 0 and ybar < 0  ==> phi =270 degrees
c         xbar and ybar = 0 ==> phi is undetermined
c
c the mean vector length "r" is determined from
c         r = ( [sum(cosine(phi(i)))]**2 + [sum(sine(phi(i)))]**2 )/numpoints
c
c the angular deviation is determined from
c         s = sqrt[2*(1-r)]
c
c this program is compiled to handle MAXPHI data points
c  and 72 "bins" for the determination of angular distributions
c
c
c initialize some variables
c
      degtorad=3.1415927/180.0
      amean = 0.0
      ybar = 0.0
      xbar = 0.0
      cossqu = 0.0
      sinsqu = 0.0
      nbins = 72
      xinc = 360.0/float(nbins)
      xmid = xinc / 2.0
```

```
      inc = 360/nbins
c
c        the angles are collected in bins
c
      do 15  i = 1,nbins
      dihd(i)  = 0.0
      angle(i) = xmid+(i-1)*xinc
   15 continue
c
      do 20 k=1,npoints
      intd = int(val(k)/degtorad)/inc + 1
      dihd(intd) = dihd(intd) + 1.0
      phirad  = val(k)
      cosphi  = cos(phirad)
      sinephi = sin(phirad)
c
c        sum all of the sines and cosines
c
      ybar = ybar + sinephi
      xbar = xbar + cosphi
   20 continue
c
c        divide the running sums by the total number of points
c
      ybar = ybar / float(npoints)
      xbar = xbar / float(npoints)
c
c        calculate the mean angle and test for some special cases
c
      if(xbar .eq. 0.0 .and. ybar .eq. 0.0) then
          amean = 9999.99999
      else if( (xbar .eq. 0.0) .and. (ybar .lt. 0.0)) then
          amean = 270.0
      else if( (xbar .eq. 0.0) .and. (ybar .gt. 0.0)) then
          amean = 90.0
      else if( xbar .gt. 0.0) then
          amean = atan(ybar/xbar) / degtorad
      else
          amean = 180.0 + (atan(ybar/xbar) / degtorad)
      endif
c
c        calculate the mean vector length
c
      r = sqrt(xbar*xbar + ybar*ybar)
c
c        calculate the mean angular variation
c
      s = sqrt(2.0*(1-r))/degtorad
c
c        normalize the distributions
c
      do 40 i=1,nbins
      dihd(i)=dihd(i)/float(npoints)
   40 continue
c
c        write out the results
c
      if (amean .gt. 180.0) then
         amean = amean - 360.0
      endif
c
c
      write(6,30) npoints,amean,r,s
   30 format(' For ',i6,' points'/
     :' the mean angle is :',f10.4,' degrees.'/
     :' the mean vector length is ',f8.5/
```

```
     :' the angular variance is ',f10.4,' degrees'//)
c
c
      return
      end
c
c
c-------------------------------------------------
      subroutine conf(ndih,npoints,dt,theta,j,k)
c-------------------------------------------------
c
c
c     This subroutine is to determine if
c       phi and psi are in a helical,
c       beta, extended, or unspecified other
c       regions of conformational space and
c       prints out marker for structure type
c       for peptide as a function of time.
c       then, determines average length of
c       helix and generates a histogram of
c       lengths, among other things.
c
c
c
      integer npoints,maxres,maxtim,dihpr,maxdih
      integer tran,sumtran,ch1,ch2
      integer ch3,ndih
c
      real totbeta,tothel,totext,totother,numst
      real hel,sumhel,numhel,max,tot,nummax
      real phisum,psisum,phiavg,psiavg,totavg
      real helf,helb,nonf,nonb,h,num,nucf,nucb,nucc
      real time,sumfrac
      real hctim,chtim,hhtim,hcavg,chavg,hhavg
      real hhsum,hcsum,chsum
      real hhsumi,hcsumi,chsumi
c
      parameter(maxtim=4000,maxdih=150,maxres=20)
c
      real frac(maxtim)
      real theta(maxres*2,maxtim)
      integer ang(maxres,maxtim)
c
c
c
      totavg = 0
      tothel = 0
      totbeta = 0
      totext = 0
      totother = 0
c
c     first, based on phi and psi, each residue is assigned to a
c     region of conformational space---reading in theta(res-angle,
c     time) and writing out in array---ang(res,time). in ang
c     1 denotes helix, 2 beta, 3 extended, and 0 other.
c
c
      i = 1
      do 20 j=1,ndih-1,2
        phisum = 0
        psisum = 0
      do 10 k=1,npoints
c
      if (theta(j,k) .ge. 288.0 .and. theta(j,k) .le. 318.0) then
         phisum = phisum + 1
      endif
```

```
      if (theta(j+1,k) .ge. 298.0 .and. theta(j+1,k) .le. 328.0) then
         psisum = psisum + 1
      endif
c
         if (theta(j,k) .ge. 283.0 .and. theta(j,k) .le. 323.0
     +        .and. theta(j+1,k) .ge.
     +       293.0 .and. theta(j+1,k) .le. 333.0) then
           ang(i,k) = 1
           tothel = tothel + 1
         elseif (theta(j,k) .ge. 201.0 .and. theta(j,k)
     +        .le. 261.0 .and. theta(j+1,k)
     +        .ge. 93.0 .and. theta(j+1,k) .le. 155.0) then
           ang(i,k) = 2
           totbeta = totbeta + 1
         elseif (theta(j,k) .ge. 160.0 .and. theta(j,k)
     +        .le. 200.0 .and. theta(j+1,k)
     +        .ge. 160.0 .and. theta(j+1,k) .le. 200.0) then
           ang(i,k) = 3
           totext = totext + 1
         else
           ang(i,k) = 0
           totother = totother + 1
         endif
c
10    continue
c
       phiavg = (phisum * 100) / npoints
       psiavg = (psisum * 100) / npoints
       totavg = totavg + ((phiavg+psiavg)/2)
c
       write(6,*)'residue number', (j+1)/2
       write(6,*)'% helix based on phi (15 deg)',phiavg
       write(6,*)'% helix based on psi (15 deg)',psiavg
       write(6,*)'average of phi and psi values',(phiavg+psiavg)/2
       write(6,*)
c
         i = i + 1
c
20    continue
c
      totavg = (totavg*2)/ndih
      write(6,*)'overall avg % helix based on avg phi,psi
     & values',totavg
      write(6,*)
c
c
c     this next part defines a helical region as having at least
c     3 residues in helical phi,psi space and then computes
c     different structural properties of the peptide with time.
c
c
      write(8,*)-99
      write(9,*)-99
c
      frac(k) = 0
      sumfrac = 0
      sumhel = 0
      hel = 0
      numhel = 0
      max = 3
      tot = 0
      nummax = 0
      dihpr = ndih/2
      do 40 k=1,npoints
         num = 0
         do 35 i=1,dihpr
```

```
          if (i .eq. dihpr) goto 33
             if (ang(i,k) .eq. 1) then
                hel = hel + 1
                elseif (ang(i,k) .ne. 1 .and. hel .lt. 3) then
                 hel = 0
                   elseif (ang(i,k) .ne. 1 .and. hel .ge. 3) then
                      sumhel = sumhel + hel
                      numhel = numhel + 1
                      num = num + 1
                      frac(k) = frac(k) + hel
                      write(8,*)hel,0.0
                      if (hel .gt. max) then
                       max = hel
                       time = k
                         endif
                      hel = 0
                endif
                goto 35
33      if (ang(i,k) .ne. 1 .and. hel .ge. 3) then
          sumhel = sumhel + hel
          numhel = numhel + 1
          num = num + 1
          frac(k) = frac(k) + hel
          write(8,*)hel,0.0
           if (hel .gt. max) then
             max = hel
             time = k
               endif
          hel = 0
            elseif (ang(i,k) .eq. 1 .and. hel .ge. 2) then
              sumhel = sumhel + hel + 1
              numhel = numhel + 1
              num = num + 1
              frac(k) = frac(k) + hel + 1
              write(8,*)hel + 1,0.0
              if (hel .gt. max) then
                max = hel
                time = k
                  endif
              hel = 0
                 elseif (ang(i,k) .ne. 1 .and. hel .lt. 3) then
                   hel = 0
                     elseif (ang(i,k) .eq. 1 .and. hel .lt. 2) then
                       hel = 0
         endif
35      continue
              frac(k) = frac(k)/dihpr
              write(9,*)k*dt,frac(k)
              sumfrac = sumfrac + frac(k)
              if (num .ge. 1.0) then
                  numst = numst + 1
                     endif
              if (num .gt. 1) then
                 tot = tot + 1
                   endif
              if (num .gt. nummax) then
                  nummax = num
                    endif
40      continue
c
c
c
      write(6,*)'Conformational analysis based on phi and'
      write(6,*)'psi values, repeating structure taken into account'
      write(6,*)
      write(6,*)'total # of time points',npoints
```

```
      write(6,*)'total # of helical residues in series is', sumhel
      write(6,*)'total # of helices is', numhel
      write(6,*)'avg length of helix',sumhel/numhel
      write(6,*)'avg number of hel. per peptide',numhel/npoints
      write(6,*)'total # of structures w/ over 1 helix',tot
      write(6,*)'% structures w/ over 1 helix',100*tot/npoints
      write(6,*)'max # of helices per peptide',nummax
      write(6,*)'maximum length of helix', max
      write(6,*)'time pt of max helix', time*dt
      write(6,*)'number of structures w/ at least 1 helix',numst
      write(6,*)'% time at least 1 helix',100*numst/npoints
c     write(6,*)'totbet',totbeta
c     write(6,*)'tothel',tothel
c     write(6,*)'totext',totext
c     write(6,*)'totother',totother
      write(6,*)'npts times # of res', npoints* dihpr
      write(6,*)'% of total pts in beta region',
     +         100*totbeta/(npoints*dihpr)
      write(6,*)'% of total pts in helical region',
     +         100*tothel/(npoints*dihpr)
      write(6,*)'% of tot pts in extended region',
     +         100*totext/(npoints*dihpr)
      write(6,*)'% of total pts in other region',
     +         100*totother/(npoints*dihpr)
      write(6,*)
c
c
c
c
c     now, calculate forward and backward frequency that adjacent
c     residues will be helical if the particular residue that
c     you are at is helical or nonhelical. also, as a measure
c     of nucleation the probability of finding a helical
c     residue with 2 coil residues on each side (e.g. cchcc)
c     is determined.
c
c
         totavg = 0
c
      write(6,*)'occurrences of res in dif conf spaces, to look
     & at zippering etc'
      write(6,*)
c
      do 85 i=1,dihpr
         helf = 0
         helb = 0
         nonf = 0
         nonb = 0
         nucf = 0
         nucb = 0
         nucc = 0
         h = 0
         do 80 k=1,npoints
           if (ang(i,k) .eq. 1) then
             h = h + 1
               endif
           if (i .eq. 1) goto 73
           if (i .eq. dihpr) goto 74
           if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1) then
             helf = helf + 1
              elseif (ang(i,k) .ne. 1 .and. ang(i+1,k) .eq. 1) then
                nonf = nonf + 1
                  endif
              if (ang(i,k) .eq. 1 .and. ang(i-1,k) .eq. 1) then
                helb = helb + 1
                  elseif (ang(i,k) .ne. 1 .and. ang(i-1,k) .eq. 1) then
```

```
                    nonb = nonb + 1
                      endif
            goto 75
73        if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1) then
            helf = helf + 1
              elseif (ang(i,k) .ne. 1 .and. ang(i+1,k) .eq. 1) then
                nonf = nonf + 1
                  endif
          go to 75
74        if (ang(i,k) .eq. 1 .and. ang(i-1,k) .eq. 1) then
            helb = helb + 1
              elseif (ang(i,k) .ne. 1 .and. ang(i-1,k) .eq. 1) then
                nonb = nonb + 1
                  endif
          go to 75
75       if (i .eq. 1 .or. i .eq. 2) goto 76
         if (i .eq. dihpr .or. i .eq. dihpr - 1) goto 77
         if (ang(i,k) .eq. 1 .and. ang(i-1,k) .ne. 1 .and.
        &    ang(i-2,k) .ne. 1) then
             nucf = nucf + 1
           elseif (ang(i,k) .eq. 1 .and. ang(i+1,k) .ne. 1 .and.
        &           ang(i+2,k) .ne. 1) then
                    nucb = nucb + 1
         endif
         goto 79
76       if (ang(i,k) .eq. 1 .and. ang(i+1,k) .ne. 1 .and.
        &    ang(i+2,k) .ne. 1) then
             nucb = nucb + 1
           endif
           goto 80
77       if (ang(i,k) .eq. 1 .and. ang(i-1,k) .ne. 1 .and.
        &    ang(i-2,k) .ne. 1) then
             nucf = nucf + 1
           endif
           goto 80
79       if (ang(i,k) .eq. 1 .and. ang(i-1,k) .ne. 1 .and.
        &    ang(i-2,k) .ne. 1 .and. ang(i+1,k) .ne. 1 .and.
        &    ang(i+2,k) .ne. 1) then
             nucc = nucc + 1
         endif
         goto 80
80       continue
c
         totavg = totavg + ((h * 100) / npoints)
c
            write(6,*)'res # is',i
            write(6,*)'percentage time helical',(h*100)/(npoints)
            write(6,*)'prob of hel to hel(i)',(helb*100)/(npoints)
            write(6,*)'prob of hel(i) to hel',(helf*100)/(npoints)
            write(6,*)'prob of hel to non(i)',(nonb*100)/(npoints)
            write(6,*)'prob of non(i) to hel',(nonf*100)/(npoints)
            write(6,*)'prob non,non to hel(i)',(nucf*100)/(npoints)
            write(6,*)'prob hel(i) to non,non',(nucb*100)/(npoints)
            write(6,*)'non,non,hel(i),non,non',(nucc*100)/(npoints)
            write(6,*)
85       continue
c
c
c
           write(6,*)'overall avg % helix (phi and psi)',
        & totavg / dihpr
           write(6,*)
c
c        now, calculate transition times for each residue in going
c        from hel to coil and back to helix. first, it is on a
c        per residue basis, sliding through time, h-c-h.
```

```
c
c
c
      hhsum = 0
      hcsum = 0
      chsum = 0
      tran = 0
      sumtran = 0
      k = 0
      i = 1
105   continue
      if (k .eq. npoints .and. i .eq. dihpr) goto 140
c
      if (k .eq. npoints .and. i .ne. dihpr) then
        write(6,*) 'residue #', i
        write(6,*) '# of transitions', tran
        write(6,*) 'avg hel to coil time interval', hcsumi/tran
        write(6,*) 'avg coil to hel time interval', chsumi/tran
        write(6,*) 'avg hel to hel time interval', hhsumi/tran
        write(6,*) 'equil const, c-h',hcsumi/chsumi
        write(6,*)
c
        i = i + 1
        k = 0
        hhtim = 0
        hctim = 0
        chtim = 0
        tran = 0
        hhsumi = 0
        chsumi = 0
        hcsumi = 0
      endif
c
      k = k + 1
c
      if (ang(i,k) .ne. 1) goto 105
c
      if (ang(i,k) .eq. 1) then
        ch1 = k
      endif
      if (k .eq. npoints) goto 105
c
      do 115 k=ch1+1,npoints
        if (ang(i,k) .ne. 1 ) then
          ch2 = k
          goto 120
        endif
        if (k .eq. npoints) goto 105
115   continue
c
120   if (k .eq. npoints) goto 105
c
      do 125 k=ch2+1,npoints
         if (ang(i,k) .eq. 1 ) then
            ch3 = k
            goto 130
         endif
         if (k .eq. npoints) goto 105
125   continue
c
130   continue
c
c     calculate average transition times for each residue
c     and also add up for overall average over all time
c     and all residues. first sum for each residue and
c     then set to zero when move on to the next residue.
```

```
c
c
      tran = tran + 1
      hctim = (ch2 - ch1) * dt
      chtim = (ch3 - ch2) * dt
      hhtim = (ch3 - ch1) * dt
      hcsumi = hcsumi + hctim
      chsumi = chsumi + chtim
      hhsumi = hhsumi + hhtim
c
c     now  sum for all residues
c
      chsum = chsum + chtim
      hcsum = hcsum + hctim
      hhsum = hhsum + hhtim
      sumtran = sumtran + 1
c
c     write out this information for each residue
c     after going through all time points, done
c     below at line 140 for the last time point
c     of the last residue and above otherwise at
c     line 105.
c
c
      goto 105
c
c
140   continue
      write(6,*) 'residue #', i
      write(6,*) '# of transitions', tran
      write(6,*) 'avg hel to coil time interval', hcsumi/tran
      write(6,*) 'avg coil to hel time interval', chsumi/tran
      write(6,*) 'avg hel to hel time interval', hhsumi/tran
      write(6,*) 'equil const, c-h',hcsumi/chsumi
      write(6,*)
c
      hcavg = hcsum / sumtran
      chavg = chsum / sumtran
      hhavg = hhsum / sumtran
c
      write(6,*) 'overall averages---all residues,all time'
      write(6,*) 'simulation time',npoints*dt
      write(6,*) 'average # of transitions per res',sumtran/dihpr
      write(6,*) 'average interval of time from hel to coil',hcavg
      write(6,*) 'average interval of time from coil to hel',chavg
      write(6,*) 'average interval of time from hel to hel',hhavg
      write(6,*) 'average equil const, c-h',hcavg/chavg
      write(6,*)
c
c     now, calculate transition times for two residues going
c     from hel to coil and back to helix.
c
c
      hhtim = 0
      hctim = 0
      chtim = 0
      hhsumi = 0
      chsumi = 0
      hcsumi = 0
      hhsum = 0
      hcsum = 0
      chsum = 0
      tran = 0
      sumtran = 0
      k = 0
      i = 1
```

```
205   continue
      if ( i .eq. dihpr) goto 250
c
      if (k .eq. npoints .and. i .ne. dihpr) then
        write(6,*) 'residues', i,i+1
        write(6,*) '# of transitions', tran
        write(6,*) 'avg hh to cc time interval', hcsumi/tran
        write(6,*) 'avg cc to hh time interval', chsumi/tran
        write(6,*) 'avg hh to hh time interval', hhsumi/tran
        write(6,*) 'equil const, cc-hh',hcsumi/chsumi
        write(6,*)
        i = i + 1
        k = 0
        hhtim = 0
        hctim = 0
        chtim = 0
        tran = 0
        hhsumi = 0
        chsumi = 0
        hcsumi = 0
      endif
c
      k = k + 1
c
      if (ang(i,k) .ne. 1) goto 205
c
      if (ang(i,k) .eq. 1 .and. ang(i+1,k) .ne. 1) goto 205
c
      if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1) then
        ch1 = k
      endif
      if (k .eq. npoints) goto 205
c
      do 215 k=ch1+1,npoints
        if (ang(i,k) .ne. 1 .and. ang(i+1,k) .ne. 1 ) then
          ch2 = k
          goto 220
        endif
      if (k .eq. npoints) goto 205
215   continue
c
220   if (k .eq. npoints) goto 205
c
      do 225 k=ch2+1,npoints
         if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1 ) then
            ch3 = k
            goto 230
         endif
         if (k .eq. npoints) goto 205
225   continue
c
230   continue
c
c     calculate average transition times for each pair of residues
c     and also add up for overall average over all time
c     and all residues. first sum for each pair of residues and
c     then set to zero when move on to the next residue.
c
c
      tran = tran + 1
      hctim = (ch2 - ch1) * dt
      chtim = (ch3 - ch2) * dt
      hhtim = (ch3 - ch1) * dt
      hcsumi = hcsumi + hctim
      chsumi = chsumi + chtim
      hhsumi = hhsumi + hhtim
```

```
c
c     now  sum for all residues
c
      chsum = chsum + chtim
      hcsum = hcsum + hctim
      hhsum = hhsum + hhtim
      sumtran = sumtran + 1
c
c     write out this information for each residue
c     after going through all time points, done
c     below at line 240 for the last time point
c     of the last residue and above otherwise at
c     line 205.
c
c
      goto 205
c
c
240   continue
      write(6,*) 'residues', i,i+1
      write(6,*) '# of transitions', tran
      write(6,*) 'avg hh to cc time interval', hcsumi/tran
      write(6,*) 'avg cc to hh time interval', chsumi/tran
      write(6,*) 'avg hh to hh time interval', hhsumi/tran
      write(6,*) 'equil const, cc-hh',hcsumi/chsumi
      write(6,*)
c
250   continue
c
      hcavg = hcsum / sumtran
      chavg = chsum / sumtran
      hhavg = hhsum / sumtran
c
      write(6,*) 'overall averages---all residues,all time'
      write(6,*) 'simulation time',npoints*dt
      write(6,*) 'average # of transitions per 2 res',sumtran/(dihpr-1)
      write(6,*) 'average interval of time from hh to cc',hcavg
      write(6,*) 'average interval of time from cc to hh',chavg
      write(6,*) 'average interval of time from hh to hh',hhavg
      write(6,*) 'average equil const, cc-hh',hcavg/chavg
      write(6,*)
c
c     now, calculate transition times for three residues going
c     from hel to coil and back to helix.
c
c
      hhsum = 0
      hcsum = 0
      chsum = 0
      hhtim = 0
      hctim = 0
      chtim = 0
      hhsumi = 0
      chsumi = 0
      hcsumi = 0
      tran = 0
      sumtran = 0
      k = 0
      i = 1
305   continue
      if ( i .eq. dihpr .or. i .eq. dihpr - 1) goto 350
c
      if (k .eq. npoints .and. i .lt. dihpr -1) goto 306
      goto 307
c
c
```

```
306   if (tran .eq. 0) then
        tran = 1
        hcsumi = 1
        chsumi = 1
        hhsumi = 1
        endif
c
        write(6,*) 'residues', i,i+1,i+2
        write(6,*) '# of transitions', tran
        write(6,*) 'avg hhh to ccc time interval', hcsumi/tran
        write(6,*) 'avg ccc to hhh time interval', chsumi/tran
        write(6,*) 'avg hhh to hhh time interval', hhsumi/tran
        write(6,*) 'equil const, ccc-hhh',hcsumi/chsumi
        write(6,*)
        i = i + 1
        k = 0
        hhtim = 0
        hctim = 0
        chtim = 0
        tran = 0
        hhsumi = 0
        chsumi = 0
        hcsumi = 0
c
307   k = k + 1
c
      if (ang(i,k) .ne. 1) goto 305
c
      if (ang(i,k) .eq. 1 .and. ang(i+1,k) .ne. 1) goto 305
c
      if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1 .and.
     +    ang(i+2,k) .ne. 1) goto 305
c
      if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1 .and.
     +    ang(i+2,k) .eq. 1) then
        ch1 = k
      endif
      if (k .eq. npoints) goto 305
c
      do 315 k=ch1+1,npoints
        if (ang(i,k) .ne. 1 .and. ang(i+1,k) .ne. 1 .and.
     +      ang(i+2,k) .ne. 1) then
          ch2 = k
          goto 320
        endif
        if (k .eq. npoints) goto 305
315   continue
c
320   if (k .eq. npoints) goto 305
c
      do 325 k=ch2+1,npoints
         if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1 .and.
     +       ang(i+2,k) .eq. 1 ) then
            ch3 = k
            goto 330
         endif
         if (k .eq. npoints) goto 305
325   continue
c
330   continue
c
c     calculate average transition times for each set of residues
c     and also add up for overall average over all time
c     and all residues. first sum for each set of 3 residues and
c     then set to zero when move on to the next residue.
c
```

```
c
      tran = tran + 1
      hctim = (ch2 - ch1) * dt
      chtim = (ch3 - ch2) * dt
      hhtim = (ch3 - ch1) * dt
      hcsumi = hcsumi + hctim
      chsumi = chsumi + chtim
      hhsumi = hhsumi + hhtim
c
c     now  sum for all residues
c
      chsum = chsum + chtim
      hcsum = hcsum + hctim
      hhsum = hhsum + hhtim
      sumtran = sumtran + 1
c
c     write out this information for each residue
c     after going through all time points, done
c     below at line 340 for the last time point
c     of the last residue and above otherwise at
c     line 305.
c
c
      goto 305
c
c
340   continue
c
      if (tran .eq. 0) then
         tran = 1
         hcsumi = 1
         chsumi = 1
         hhsumi = 1
         endif
c
      write(6,*) 'residues', i,i+1,i+2
      write(6,*) '# of transitions', tran
      write(6,*) 'avg hhh to ccc time interval', hcsumi/tran
      write(6,*) 'avg ccc to hhh time interval', chsumi/tran
      write(6,*) 'avg hhh to hhh time interval', hhsumi/tran
      write(6,*) 'equil const, ccc-hhh',hcsumi/chsumi
      write(6,*)
c
350   continue
c
      hcavg = hcsum / sumtran
      chavg = chsum / sumtran
      hhavg = hhsum / sumtran
c
      write(6,*) 'overall averages---all residues,all time'
      write(6,*) 'simulation time',npoints*dt
      write(6,*) 'average # of transitions per 3 res',sumtran/(dihpr-2)
      write(6,*) 'average interval of time from hhh to ccc',hcavg
      write(6,*) 'average interval of time from ccc to hhh',chavg
      write(6,*) 'average interval of time from hhh to hhh',hhavg
      write(6,*) 'average equil const, ccc-hhh',hcavg/chavg
      write(6,*)
c
c
c
c     now, calculate transition times for three residues going
c     from hel to hhc and back to hhh. this is to get the free
c     energy for adding a residue to an existing helical segment.
c
c
      hhsum = 0
```

```
      hcsum = 0
      chsum = 0
      hhtim = 0
      hctim = 0
      chtim = 0
      hhsumi = 0
      chsumi = 0
      hcsumi = 0
      tran = 0
      sumtran = 0
      k = 0
      i = 1
405   continue
      if ( i .eq. dihpr .or. i .eq. dihpr - 1) goto 450
c
      if (k .eq. npoints .and. i .lt. dihpr -1) then
        write(6,*) 'residues', i,i+1,i+2
        write(6,*) '# of transitions', tran
        write(6,*) 'avg hhh to hhc time interval', hcsumi/tran
        write(6,*) 'avg hhc to hhh time interval', chsumi/tran
        write(6,*) 'avg hhc to hhc time interval', hhsumi/tran
        write(6,*) 'equil const, hhc-hhh',hcsumi/chsumi
        write(6,*)
        i = i + 1
        k = 0
        hhtim = 0
        hctim = 0
        chtim = 0
        tran = 0
        hhsumi = 0
        chsumi = 0
        hcsumi = 0
      endif
c
      k = k + 1
c
      if (ang(i,k) .ne. 1) goto 405
c
      if (ang(i,k) .eq. 1 .and. ang(i+1,k) .ne. 1) goto 405
c
      if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1 .and.
     +    ang(i+2,k) .eq. 1) goto 405
c
      if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1 .and.
     +    ang(i+2,k) .ne. 1) then
        ch1 = k
      endif
      if (k .eq. npoints) goto 405
c
      do 415 k=ch1+1,npoints
        if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1 .and.
     +      ang(i+2,k) .eq. 1) then
          ch2 = k
          goto 420
        endif
        if (k .eq. npoints) goto 405
415   continue
c
420   if (k .eq. npoints) goto 405
c
      do 425 k=ch2+1,npoints
         if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1 .and.
     +       ang(i+2,k) .ne. 1 ) then
            ch3 = k
            goto 430
         endif
```

```
         if (k .eq. npoints) goto 405
425   continue
c
430   continue
c
c     calculate average transition times for each set of residues
c     and also add up for overall average over all time
c     and all residues. first sum for each set of 3 residues and
c     then set to zero when move on to the next residue.
c
c
      tran = tran + 1
      chtim = (ch2 - ch1) * dt
      hctim = (ch3 - ch2) * dt
      hhtim = (ch3 - ch1) * dt
      hcsumi = hcsumi + hctim
      chsumi = chsumi + chtim
      hhsumi = hhsumi + hhtim
c
c     now  sum for all residues
c
      chsum = chsum + chtim
      hcsum = hcsum + hctim
      hhsum = hhsum + hhtim
      sumtran = sumtran + 1
c
c     write out this information for each residue
c     after going through all time points, done
c     below at line 440 for the last time point
c     of the last residue and above otherwise at
c     line 405.
c
c
      goto 405
c
c
440   continue
      write(6,*) 'residues', i,i+1,i+2
      write(6,*) '# of transitions', tran
      write(6,*) 'avg hhc to hhh time interval', chsumi/tran
      write(6,*) 'avg hhh to hhc time interval', hcsumi/tran
      write(6,*) 'avg hhc to hhc time interval', hhsumi/tran
      write(6,*) 'equil const, hhc-hhh',hcsumi/chsumi
      write(6,*)
c
450   continue
c
      hcavg = hcsum / sumtran
      chavg = chsum / sumtran
      hhavg = hhsum / sumtran
c
      write(6,*) 'overall averages---all residues,all time'
      write(6,*) 'simulation time',npoints*dt
      write(6,*) 'average # of transitions per 3 res',sumtran/(dihpr-2)
      write(6,*) 'average interval of time from hhh to hhc',hcavg
      write(6,*) 'average interval of time from hhc to hhh',chavg
      write(6,*) 'average interval of time from hhc to hhc',hhavg
      write(6,*) 'average equil const, hhc-hhh',hcavg/chavg
      write(6,*)
c
c
c
c
c     now, calculate transition times for three residues going
c     from ccc to chc and back to ccc. this is to get the free
c     energy for nucleation and the nucleation parameter.
```

```
c
c
      hhsum = 0
      hcsum = 0
      chsum = 0
      hhtim = 0
      hctim = 0
      chtim = 0
      hhsumi = 0
      chsumi = 0
      hcsumi = 0
      tran = 0
      sumtran = 0
      k = 0
      i = 1
505   continue
      if ( i .eq. dihpr .or. i .eq. dihpr - 1) goto 550
c
      if (k .eq. npoints .and. i .lt. dihpr -1) then
        write(6,*) 'residues', i,i+1,i+2
        write(6,*) '# of transitions', tran
        write(6,*) 'avg ccc to chc time interval', chsumi/tran
        write(6,*) 'avg chc to ccc time interval', hcsumi/tran
        write(6,*) 'avg ccc to ccc time interval', hhsumi/tran
        write(6,*) 'equil const, ccc-chc',hcsumi/chsumi
        write(6,*)
        i = i + 1
        k = 0
        hhtim = 0
        hctim = 0
        chtim = 0
        tran = 0
        hhsumi = 0
        chsumi = 0
        hcsumi = 0
      endif
c
      k = k + 1
c
      if (ang(i,k) .eq. 1) goto 505
c
      if (ang(i,k) .ne. 1 .and. ang(i+1,k) .eq. 1) goto 505
c
      if (ang(i,k) .ne. 1 .and. ang(i+1,k) .ne. 1 .and.
     +    ang(i+2,k) .eq. 1) goto 505
c
      if (ang(i,k) .ne. 1 .and. ang(i+1,k) .ne. 1 .and.
     +    ang(i+2,k) .ne. 1) then
        ch1 = k
      endif
      if (k .eq. npoints) goto 505
c
      do 515 k=ch1+1,npoints
        if (ang(i,k) .ne. 1 .and. ang(i+1,k) .eq. 1 .and.
     +      ang(i+2,k) .ne. 1) then
          ch2 = k
          goto 520
        endif
        if (k .eq. npoints) goto 505
515   continue
c
520   if (k .eq. npoints) goto 505
c
      do 525 k=ch2+1,npoints
         if (ang(i,k) .ne. 1 .and. ang(i+1,k) .ne. 1 .and.
     +       ang(i+2,k) .ne. 1 ) then
```

```
            ch3 = k
            goto 530
         endif
         if (k .eq. npoints) goto 505
525   continue
c
530   continue
c
c     calculate average transition times for each set of residues
c     and also add up for overall average over all time
c     and all residues. first sum for each set of 3 residues and
c     then set to zero when move on to the next residue.
c
c
      tran = tran + 1
      chtim = (ch2 - ch1) * dt
      hctim = (ch3 - ch2) * dt
      hhtim = (ch3 - ch1) * dt
      hcsumi = hcsumi + hctim
      chsumi = chsumi + chtim
      hhsumi = hhsumi + hhtim
c
c     now  sum for all residues
c
      chsum = chsum + chtim
      hcsum = hcsum + hctim
      hhsum = hhsum + hhtim
      sumtran = sumtran + 1
c
c     write out this information for each residue
c     after going through all time points, done
c     below at line 540 for the last time point
c     of the last residue and above otherwise at
c     line 505.
c
c
      goto 505
c
c
540   continue
      write(6,*) 'residues', i,i+1,i+2
      write(6,*) '# of transitions', tran
      write(6,*) 'avg ccc to chc time interval', chsumi/tran
      write(6,*) 'avg chc to ccc time interval', hcsumi/tran
      write(6,*) 'avg ccc to ccc time interval', hhsumi/tran
      write(6,*) 'equil const, ccc-chc',hcsumi/chsumi
      write(6,*)
c
550   continue
c
      hcavg = hcsum / sumtran
      chavg = chsum / sumtran
      hhavg = hhsum / sumtran
c
      write(6,*) 'overall averages---all residues,all time'
      write(6,*) 'simulation time',npoints*dt
      write(6,*) 'average # of transitions per 3 res',sumtran/(dihpr-2)
      write(6,*) 'average interval of time from ccc to chc',chavg
      write(6,*) 'average interval of time from chc to ccc',hcavg
      write(6,*) 'average interval of time from ccc to ccc',hhavg
      write(6,*) 'average equil const, ccc-chc',hcavg/chavg
      write(6,*)
c
c
c
c
```

```
c     now, calculate transition times for three residues going
c     from hcc to hhc and back to hcc. this is to get the free
c     energy for propagation of the helix.
c
c
      hhsum = 0
      hcsum = 0
      chsum = 0
      hhtim = 0
      hctim = 0
      chtim = 0
      hhsumi = 0
      chsumi = 0
      hcsumi = 0
      tran = 0
      sumtran = 0
      k = 0
      i = 1
605   continue
      if ( i .eq. dihpr .or. i .eq. dihpr - 1) goto 650
c
      if (k .eq. npoints .and. i .lt. dihpr -1) then
        write(6,*) 'residues', i,i+1,i+2
        write(6,*) '# of transitions', tran
        write(6,*) 'avg hcc to hhc time interval', chsumi/tran
        write(6,*) 'avg hhc to hcc time interval', hcsumi/tran
        write(6,*) 'avg hcc to hcc time interval', hhsumi/tran
        write(6,*) 'equil const, hcc-hhc',hcsumi/chsumi
        write(6,*)
        i = i + 1
        k = 0
        hhtim = 0
        hctim = 0
        chtim = 0
        tran = 0
        hhsumi = 0
        chsumi = 0
        hcsumi = 0
      endif
c
      k = k + 1
c
      if (ang(i,k) .ne. 1) goto 605
c
      if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1) goto 605
c
      if (ang(i,k) .eq. 1 .and. ang(i+1,k) .ne. 1 .and.
     +    ang(i+2,k) .eq. 1) goto 605
c
      if (ang(i,k) .eq. 1 .and. ang(i+1,k) .ne. 1 .and.
     +    ang(i+2,k) .ne. 1) then
        ch1 = k
      endif
      if (k .eq. npoints) goto 605
c
      do 615 k=ch1+1,npoints
        if (ang(i,k) .eq. 1 .and. ang(i+1,k) .eq. 1 .and.
     +      ang(i+2,k) .ne. 1) then
          ch2 = k
          goto 620
        endif
        if (k .eq. npoints) goto 605
615   continue
c
620   if (k .eq. npoints) goto 605
c
```

```
      do 625 k=ch2+1,npoints
         if (ang(i,k) .eq. 1 .and. ang(i+1,k) .ne. 1 .and.
     +       ang(i+2,k) .ne. 1 ) then
            ch3 = k
            goto 630
         endif
         if (k .eq. npoints) goto 605
625   continue
c
630   continue
c
c     calculate average transition times for each set of residues
c     and also add up for overall average over all time
c     and all residues. first sum for each set of 3 residues and
c     then set to zero when move on to the next residue.
c
c
      tran = tran + 1
      chtim = (ch2 - ch1) * dt
      hctim = (ch3 - ch2) * dt
      hhtim = (ch3 - ch1) * dt
      hcsumi = hcsumi + hctim
      chsumi = chsumi + chtim
      hhsumi = hhsumi + hhtim
c
c     now  sum for all residues
c
      chsum = chsum + chtim
      hcsum = hcsum + hctim
      hhsum = hhsum + hhtim
      sumtran = sumtran + 1
c
c     write out this information for each residue
c     after going through all time points, done
c     below at line 640 for the last time point
c     of the last residue and above otherwise at
c     line 605.
c
c
      goto 605
c
c
640   continue
      write(6,*) 'residues', i,i+1,i+2
      write(6,*) '# of transitions', tran
      write(6,*) 'avg hcc to hhc time interval', chsumi/tran
      write(6,*) 'avg hhc to hcc time interval', hcsumi/tran
      write(6,*) 'avg hcc to hcc time interval', hhsumi/tran
      write(6,*) 'equil const, hcc-hhc',hcsumi/chsumi
      write(6,*)
c
650   continue
c
      hcavg = hcsum / sumtran
      chavg = chsum / sumtran
      hhavg = hhsum / sumtran
c
      write(6,*) 'overall averages---all residues,all time'
      write(6,*) 'simulation time',npoints*dt
      write(6,*) 'average # of transitions per 3 res',sumtran/(dihpr-2)
      write(6,*) 'average interval of time from hcc to hhc',chavg
      write(6,*) 'average interval of time from hhc to hcc',hcavg
      write(6,*) 'average interval of time from hcc to hcc',hhavg
      write(6,*) 'average equil const, hcc-hhc',hcavg/chavg
      write(6,*)
c
```

```
c
c
      return
      end
c
c
c-------------------------------------------------
      subroutine hbond(npoints,dt,val,k)
c-------------------------------------------------
c
c     this routine calculates the times of hbond
c     breaks and reformation based on hbond
c     distances being < 3 A. this info is printed
c     out and from it the average time the hbonds
c     are broken etc is determined.
c
      integer npoints,numbet,numbr,maxtim,i,k
c
      parameter(maxtim=4000)
c
      real end,sttim,betw
      real broke,totbr,totbet,val(maxtim)
c
c
      k=0
      numbet = 0
      numbr = 0
      totbet = 0
      totbr = 0
      i = 1
      end = 0
27    continue
      k=k+1
      if (k .eq. npoints) goto 62
      if (val(k) .le. 3.0) goto 27
        sttim = k
        write(6,*)'starting time of H-bond break',sttim*dt
        betw = sttim - end
        totbet = totbet + betw
        numbet = numbet + 1
      do 50 i=k,npoints - 1
        if (i .eq. npoints - 1 .and. val(i) .gt. 3.0) then
          k = i
          goto 27
          endif
        if (val(i) .gt. 3.0) goto 50
          end = i
          broke = end - sttim
          write(6,*)'length of time broken',broke*dt
          totbr = totbr + broke
          numbr = numbr + 1
          k = i
          goto 60
50    continue
60    continue
      goto 27
62    continue
      if (val(k) .gt. 3.0 .and. val(k-1) .gt. 3.0) then
        end = k
        write(6,*)'end point is', end*dt
        broke = end - sttim
        write(6,*)'length of time broken',broke*dt
        totbr = totbr + broke
        numbr = numbr + 1
      elseif (val(k) .gt. 3.0 .and. val(k-1) .le. 3.0) then
        sttim = k
```

```
          write(6,*)'end point is', sttim*dt
          betw = sttim - end
          write(6,*)'time between breaks',betw*dt
          totbet = totbet + betw
          numbet = numbet + 1
        elseif (val(k) .le. 3.0 .and. val(k-1) .le. 3.0) then
          sttim = k
          write(6,*)'end point is',sttim*dt
          betw = sttim -end
          write(6,*)'time between breaks',betw*dt
          totbet = totbet + betw
          numbet = numbet + 1
        elseif (val(k) .le. 3.0 .and. val(k-1) .gt. 3.0) then
          end = k
          write(6,*)'end point is', end*dt
          broke = end - sttim
          write(6,*)'length of time broken',broke*dt
          totbr = totbr + broke
          numbr = numbr + 1
        endif
c
c
c       check to make sure do not have divide by zero
c       which happens when the hydrogen bond never breaks.
c
        if (numbr .eq. 0.0) numbr = 1
        if (numbet .eq. 0.0) numbet = 1
c
c
        write(6,*)
        write(6,*)'total time broken (psec)= ',totbr*dt
        write(6,*)'total time between (psec)= ',totbet*dt
        write(6,*)'number of times broken= ',numbr
        write(6,*)'average time broken= ',(totbr*dt)/numbr
        write(6,*)'avg time interval betw breaks',(totbet*dt)/numbet
        write(6,*)'percentage of time broken',100*totbr/npoints
c
        return
        end
c
c
c-----------------------------------------------------------------------
        subroutine center(npoints,angat1,angat2,cent1,cent2,val,dis,k)
c-----------------------------------------------------------------------
c
c       this routine calculates properties for hydrogen
c       bonding to an aromatic ring.
c
c
        integer npoints,i,k,maxtim,maxat
        integer angat1,angat2,cent1,cent2
c
        parameter(maxtim=4000,maxat=120)
c
        real rx,ry,rz
        real bot,angle,d12,d23,d13,x1,x2
        real y1,y2,z1,z2,val(maxtim),dis(maxtim)
        real xat1,xat2,yat1,yat2,zat1,zat2
c
        common/ xcoord/x(maxtim,maxat)
        common/ ycoord/y(maxtim,maxat)
        common/ zcoord/z(maxtim,maxat)
c
        do 10 k=1,npoints
c
c         i = cent1  ----- 1st atom in ring
```

```
        x1 = x(k,cent1)
        y1 = y(k,cent1)
        z1 = z(k,cent1)
c
c       i = cent2  ----- 2nd atom in ring
        x2 = x(k,cent2)
        y2 = y(k,cent2)
        z2 = z(k,cent2)
c
        rx = (x1 + x2) / 2
        ry = (y1 + y2) / 2
        rz = (z1 + z2) / 2
c
c       read in other two atoms for angle calculation where
c        center of ring point is one of the angle atoms.
c
c
c       i = angat1  ---- 1st atom of angle
        xat1 = x(k,angat1)
        yat1 = y(k,angat1)
        zat1 = z(k,angat1)
c
c       i = angat2  ---- 2nd atom of angle
        xat2 = x(k,angat2)
        yat2 = y(k,angat2)
        zat2 = z(k,angat2)
c
c       calculate distance between two input atoms and center
c       of ring
c
c
        d12 = (xat2-xat1)**2 + (yat2-yat1)**2 + (zat2-zat1)**2
        d12 = sqrt(d12)
c
        d23 = (rx-xat2)**2 + (ry-yat2)**2 + (rz-zat2)**2
        d23 = sqrt(d23)
c
        d13 = (rx-xat1)**2 + (ry-yat1)**2 + (rz-zat1)**2
        d13 = sqrt(d13)
        dis(k) = d13
c
c
        bot = 2*d12*d23
        angle = d12**2 + d23**2
        angle = (angle - d13**2)/bot
        val(k) = (acos(angle))*180.0 / 3.141592654
c
c
10      continue
c
c
c
      return
      end
c
c
c-------------------------------------------------------
      subroutine r3(npoints,sub,val,k)
c-------------------------------------------------------
c
c     This subroutine converts distances
c        calculated in the distance subroutine and
c        subtracts the necessary number of c-h
c        bond lengths specified in the input file
c        and then converts each distance to
c        1/d*3 and then averages these
```

```
c        new distances. Then the average
c        is inverted and converted back
c        to d, giving a weighted distance
c        corresponding to a noe.
c
      integer npoints,sub
      integer maxtim
c
      parameter(maxtim=4000)
c
      real sum,min,max,avg,mean
      real fluc(maxtim),fl2(maxtim),fl2mean,fl2sum,fl2sq
      real flsum,flmean,sumst,flucst(maxtim),maxdev
      real mindev, maxx, fl6stsum, fl6stavg,val(maxtim)
c
c
      sumst = 0
      sum = 0
      max = 0.000001
      min = 0.000001
      do 50 k=1,npoints
         if (sub .eq. 0) then
            val(k) = val(k)
             elseif (sub .eq. 1) then
               val(k) = val(k) - 1.09
                 elseif (sub .eq. 2) then
                   val(k) = val(k) - 2.18
         endif
         sumst = sumst + val(k)
50    continue
          avg = sumst / npoints
      do 55 k=1,npoints
         flucst(k) = val(k) - avg
          if (flucst(k) .lt. 0) then
            flucst(k) = - flucst(k)
             else
              flucst(k) = flucst(k)
               endif
        flsum = flsum + flucst(k)
        val(k) = 1 / (val(k)**3)
        sum = sum + val(k)
55    continue
      flmean = flsum / npoints
      mean = sum / npoints
      fl2sum = 0
      fl6stsum =0
      do 60 k=1,npoints
        fluc(k) = val(k) - mean
        if (fluc(k) .lt. 0) then
           fluc(k) = -fluc(k)
        else
           fluc(k) = fluc(k)
        endif
c
        fl6stsum = fl6stsum + fluc(k)
        fl2(k) = fluc(k)**2
        fl2sum = fl2sum + fl2(k)
        if(val(k) .gt. min) then
          min = val(k)
         endif
        if(val(k) .lt. max) then
          max = val(k)
         endif
60    continue
      fl2mean = fl2sum / npoints
      fl6stavg = fl6stsum / npoints
```

```
      fl2mean = 1/(fl2mean**0.3333333)
      fl6stavg = 1/(fl6stavg**0.3333333)
      fl2sq = sqrt (fl2mean)
c
      mindev = (min - mean)
      if (mindev .lt. 0.0) mindev = - mindev
      maxdev = (max - mean)
      if (maxdev .lt. 0.0) maxdev = - maxdev
      mindev = 1/((mindev)**0.3333333)
      maxdev = 1/((maxdev)**0.3333333)
c
      if (mindev .gt. maxdev) then
        maxx = mindev
      else
        maxx = maxdev
      endif
c
c
         write(6,*)'avg r is', avg
         write(6,*)'avg weighted r is', 1/(mean**0.3333333)
         write(6,*)'mean fluc is', flmean
         write(6,*)'max deviation', maxx
         write(6,*)'rms fluc is (r3 base)', fl2sq
         write(6,*)'mean fluc (r3 base)',fl6stavg
c
c
      return
      end
c
c-------------------------------------------------------------
      subroutine corr(npoints,coropt,dt,val,val2,k)
c-------------------------------------------------------------
c
c     this routine calculates cross correlation
c     coefficients between different dihedral pairs.
c     the dihedrals must be read in sequence in the
c     input file, as pairs of dihedrals are fed into
c     this routine.
c
c
      integer npoints,maxtim
c
      parameter(maxtim=4000)
c
      real val(maxtim),val2(maxtim),valr,val2r
      real fa,fb
      real cosa, sina,cosb,sinb
      real fa2,fb2
      real avgfa,avgfb,bot
      real sumfa,sumfb,top,coravg
      real sumfa2,sumfb2,sqavgfa,sqavgfb
      real amean,bmean,smsina,smcosa,smsinb,smcosb
c
c
      smsina = 0
      smcosa = 0
      smsinb = 0
      smcosb = 0
      sumfa = 0
      sumfb = 0
      sumfa2 = 0
      sumfb2 = 0
c
c        convert to 0-360 convention and radians
c
      degtorad = 3.1415927 / 180.0
```

```
c
      do 10 k=1,npoints
         if (val(k) .lt. 0.0) then
           val(k) = val(k) + 360.0
             endif
         if (val2(k) .lt. 0.0) then
             val2(k) = val2(k) + 360.0
               endif
         valr  = val(k)*degtorad
         val2r  = val2(k)*degtorad
         cosa  = cos(valr)
         sina = sin(valr)
         cosb  = cos(val2r)
         sinb = sin(val2r)
c
c        sum all of the sines and cosines
c
         smsina = smsina + sina
         smcosa = smcosa + cosa
         smsinb = smsinb + sinb
         smcosb = smcosb + cosb
10    continue
c
c        divide the running sums by the total number of points
c
      smsina = smsina / npoints
      smcosa = smcosa / npoints
      smsinb = smsinb / npoints
      smcosb = smcosb / npoints
c
c
c     calculate the mean angle and test for some special cases for a
c
      if(smcosa .eq. 0.0 .and. smsina .eq. 0.0) then
          amean = 9999.99999
      else if( (smcosa .eq. 0.0) .and. (smsina .lt. 0.0)) then
          amean = 270.0
      else if( (smcosa .eq. 0.0) .and. (smsina .gt. 0.0)) then
          amean = 90.0
      else if( smcosa .gt. 0.0) then
          amean = atan(smsina/smcosa) / degtorad
      else
          amean = 180.0 + (atan(smsina/smcosa) / degtorad)
      endif
c
      if (amean .gt. 180.0) amean = amean - 360.0
c
c     calculate the mean angle and test for some special cases for b
c
      if(smcosb .eq. 0.0 .and. smsinb .eq. 0.0) then
          bmean = 9999.99999
      else if( (smcosb .eq. 0.0) .and. (smsinb .lt. 0.0)) then
          bmean = 270.0
      else if( (smcosb .eq. 0.0) .and. (smsinb .gt. 0.0)) then
          bmean = 90.0
      else if( smcosb .gt. 0.0) then
          bmean = atan(smsinb/smcosb) / degtorad
      else
          bmean = 180.0 + (atan(smsinb/smcosb) / degtorad)
      endif
c
      if (bmean .gt. 180.0) bmean = bmean - 360.0
c
      do 15 k=1,npoints
         fa = val(k) - amean
           if (fa .gt. 180.0) then
```

```
            fa = fa - 360.0
               endif
          fa2 = fa**2
c
          fb = val2(k) - bmean
            if (fb .gt. 180.0) then
               fb = fb - 360.0
               endif
          fb2 = fb**2
c
          top = top + (fa*fb)
          sumfa2 = sumfa2 + fa2
          sumfb2 = sumfb2 + fb2
c
15      continue
c
        avgfa = sumfa2 / npoints
        avgfb = sumfb2 / npoints
        sqavgfa = sqrt(avgfa)
        sqavgfb = sqrt(avgfb)
        top = top/npoints
        bot = sqavgfa * sqavgfb
        coravg = top/bot
c
c
        write (6,*)'npoints',npoints
        write (6,*)'avg a', amean
        write (6,*)'avg b', bmean
        write (6,*)'avg correlation coefficient',coravg
        write (6,*)'rmsf a',sqavgfa
        write (6,*)'rmsf b',sqavgfb
        write (6,*)
c
        return
        end
```

APPENDIX II

The appendix describes two serious problems that I encountered with the molecular dynamics module of AMBER that were never solved. The description that follows merely attempts to document these problems. First, the problem of using the all-atom representation with small peptides is described, which was alluded to in Chapter 3. The second area involves documentation of a bug in the MD code for the Cray.

In testing the sigmoidal dielectric function, we wanted a small system so that long simulations could be performed. It was also important to have structural experimental data to evaluate our results. For these reasons, among others, we began to study the peptide described in Chapter 3. We were forced to compare the 2-D NMR data to our simulations with united atom models because our simulations using the all-atom representation were unsuccessful. Two 500 ps simulations were performed of the peptide with the all-atom model, one using the sigmoidal dielectric function and the other with the linear function. In both simulations the peptide collapsed and lost helix content. Figure II.1 shows the end-to-end distance as a function of time using the sigmoidal function. The structure collapsed after 300 ps and remained in that state for the following 200 ps. There was no discernable helical structure in the peptide after approximately 200 ps. Also, both simulations gave long distances corresponding to the observed NOEs. In fact, all of the distances were considerably greater than 5 Å.

Initially we thought that the fast heating (<1 ps to reach the simulation temperature) that is used with constant temperature simulations was causing the problem. To test this idea, we slowly heated the peptide to 278 K over 20 ps. There was no improvement with slower heating. It appears that the structures with all hydrogen atoms are inherently less stable than the united atom counterparts. We evaluated the energy of the final structure after 500 ps of MD using the all-atom model with both the all-atom and united atom force field parameters. The total energy was over 60 kcal/mole higher using the all-atom parameters. In fact, the only component of the potential energy function to favor the

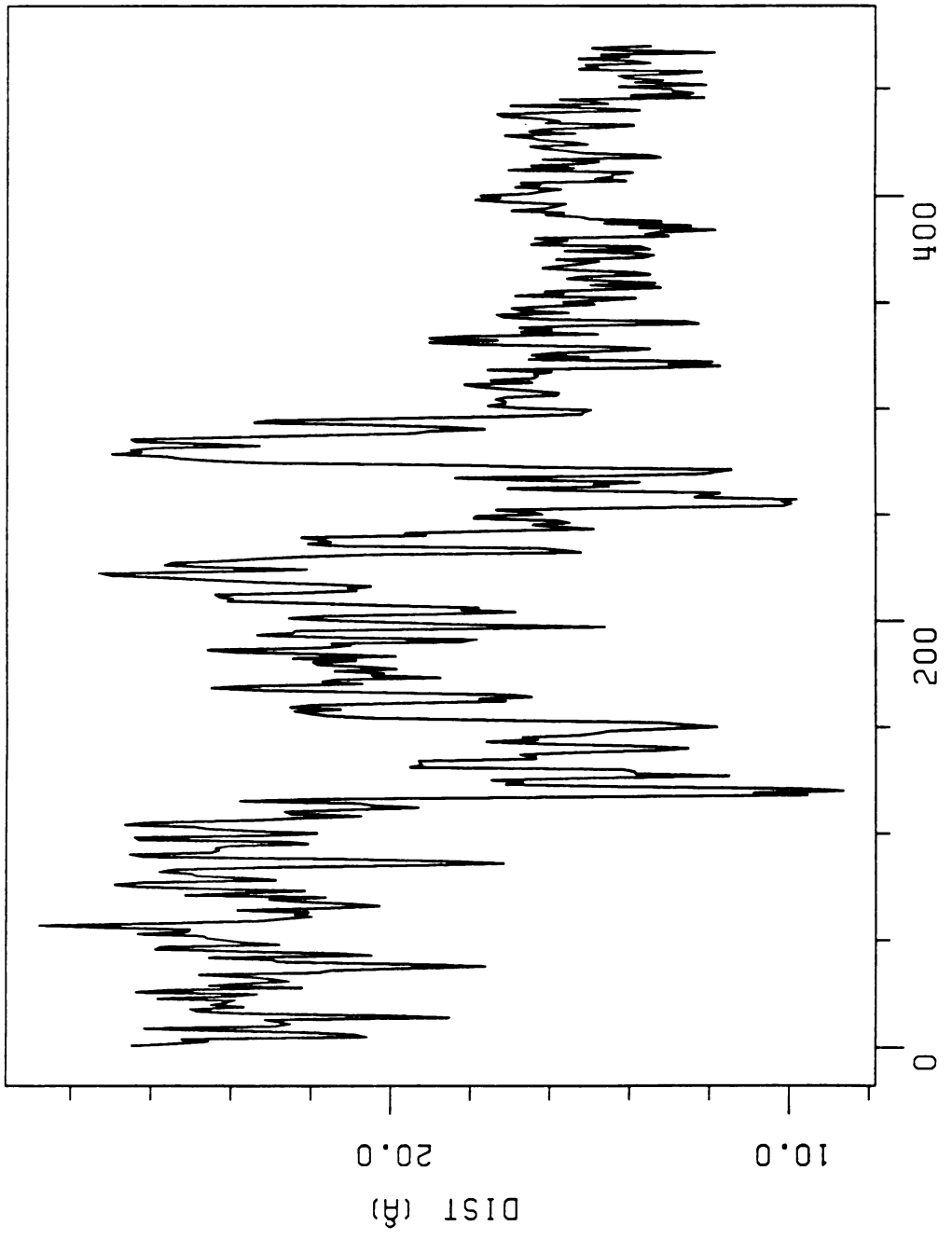


Figure II.1: End-to-end distance (Å) for the simulation of the peptide shown in Figure 3.1 using the all-atom representation.

all-atom parameters was the dihedral term. It may be that the high density of atoms and charges with the all-atom model causes strain that eventually leads to collapse. The 60 kcal/mole difference between the two representations for that particular structure was due almost entirely to higher angle energy (by 54 kcal/mole) with the all-atom parameters. This effect has not yet been reported with the all-atom force field and other studies may have reported reasonable results with these parameters because the simulations were relatively short (< 50 ps).

The second major problem that I encountered with molecular dynamics involves a bug in the Cray code. After running 4 ns of MD of polyalanine (Chapter 4) on the FPS 264 array processor, I wanted to repeat the simulation and collect structures more frequently for analysis. So, another 4 ns simulation was performed but this time on the Cray X-MP. The results from this simulation were very intriguing because the peptide showed almost the opposite behavior of the FPS simulation: the coil state was more stable than the helix (the overall average helix content was 20%). Initially the results from these two simulations seemed reasonable because the equilibrium constant between the helix and coild state is approximately one. Unfortunately, further testing showed that the structures were torn apart when run on the Cray. We took the final structure after 2 ns of MD on the FPS, and from that structure and its associated velocities we performed 100 ps of MD on the Cray.

Figure II.2 shows the end-to-end distance of the peptide for 2100-2200 ps from the original FPS run (Figure II.2A) and the same time period on the Cray beginning from the FPS structure (Figure II.2B). As can be seen from these plots, the peptide became much more compact when run on the Cray compared to the FPS (note the different scales). The two simulations should be essentially the same because the structures and velocities determine the trajectory. Figure II.3 shows the fraction of helix during this time frame for the two simulations. On the FPS the helix content remained fairly high during this time period; however, on the Cray the helical structure was disrupted almost immediately. I

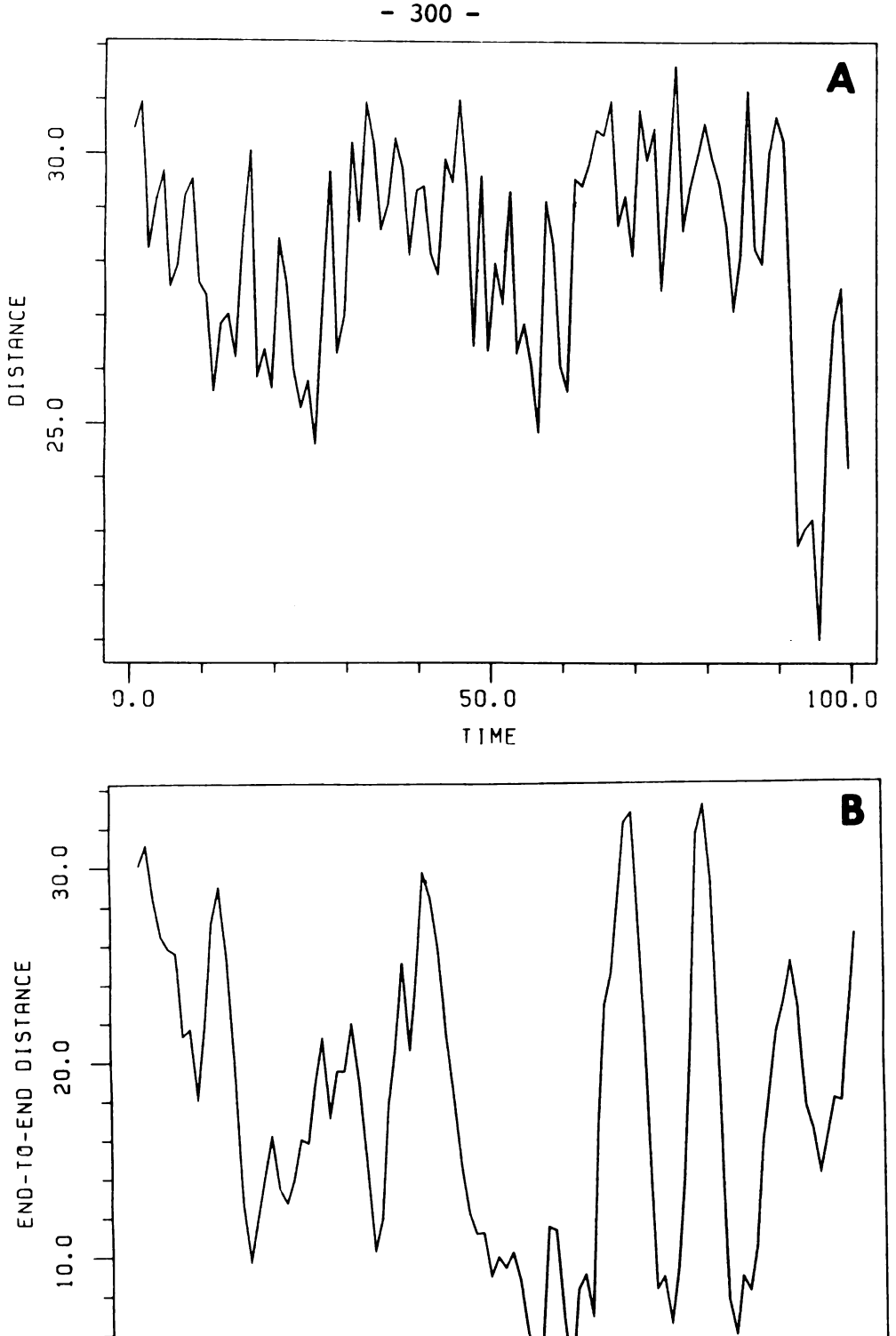


Figure II.2: End-to-end distance (Å) as a function of simulation time. Panel A is the original FPS simulation described in Chapter 4. Panel B is a simulation on the Cray beginning with the 2 ns structure from the FPS run.

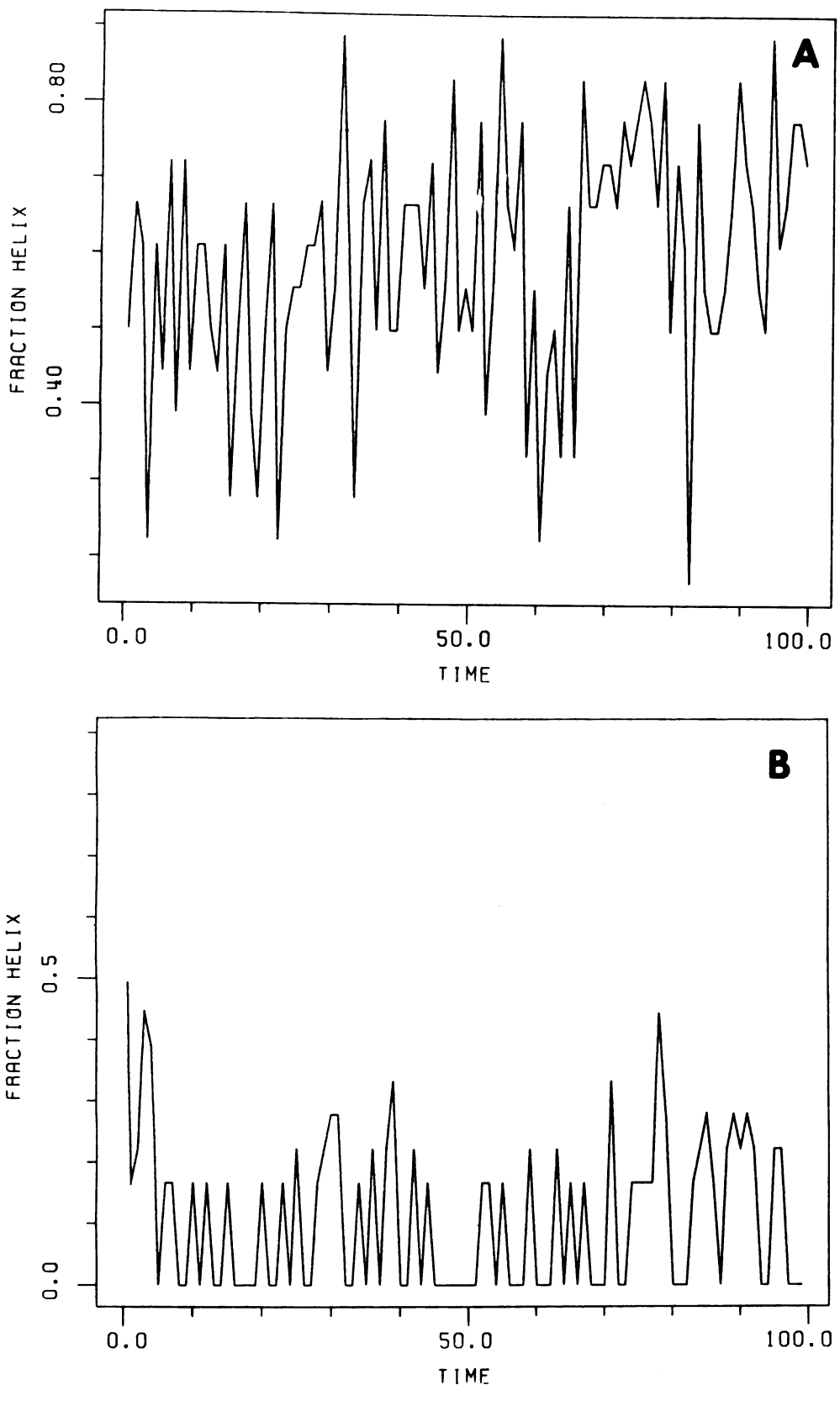


Figure II.3: Fraction helix as a function of time. (A) Original FPS run. (B) Cray run beginning with the FPS structure.

UC San Francisco LIBRARY
UNIVERSITY OF CALIFORNIA

UC San Francisco LIBRARY
UNIVERSITY OF CALIFORNIA
FOR REFERENCE
NOT TO BE TAKEN FROM THE ROOM
BRO DART
CAT. NO. 23 012
PRINTED IN U.S.A.