**Title**
Modeling protein stability



**Author**
Veenstra, David L.

**Publication Date**
1996

Peer reviewed|Thesis/dissertation

# MODELING PROTEIN STABILITY: MOLECULAR DYNAMICS AND FREE ENERGY CALCULATIONS OF T4 LYSOZYME

**by**

David L. Veenstra
B.S, University of California Santa Barbara, 1989
Pharm.D., University of California San Francisco, 1996

**DISSERTATION**

**Submitted in partial satisfaction of the requirements for the degree of**

**DOCTOR OF PHILOSOPHY**

**in**

Pharmaceutical Chemistry

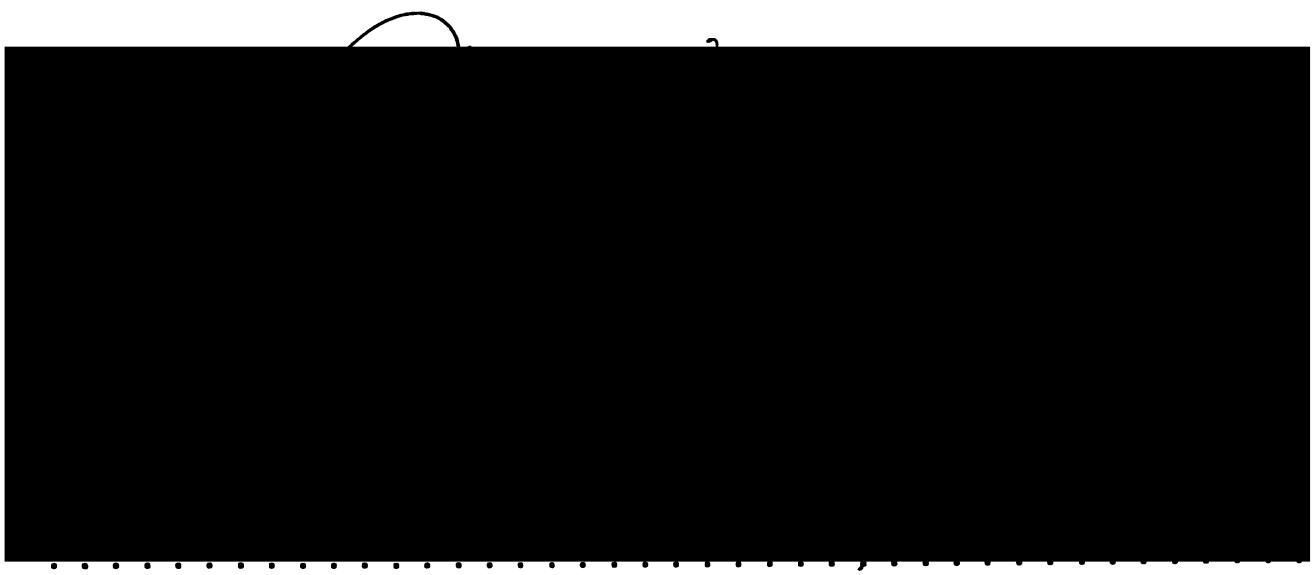
**in the**

**GRADUATE DIVISION**

**of the**

**UNIVERSITY OF CALIFORNIA**

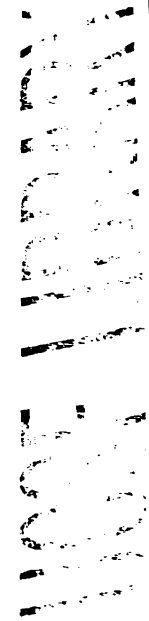
**San Francisco**



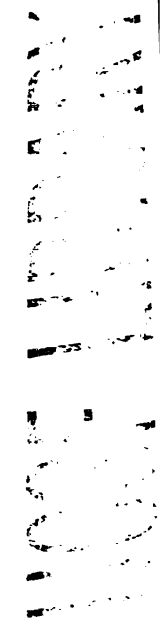
. . . . . .

**Date** **University Librarian**

**Degree Conferred:** . . . . . .



Many thanks to Peter Kollman for his endless enthusiasm and patience, my family for their support, and my friends for being just that. The La Playa gang - Spanky, the Muffster, Tiffer and Halls deserve a special thanks, as do the new and honorary members, Jonzer, Boutrous and B.C.



## Thesis Abstract

This dissertation encompases the theoretical study of protein stability with molecular mechanics methodologies. Chapter 1 provides an introductory review of this work and places it in a historical perspective. Chapter 2 describes the analysis and development of refined van der Waals parameters based on *ab initio* gas phase calculations. Presented in Chapter 3 is the design and analysis of unnatural amino acid mutants of T4 lysozyme, which was conducted in conjunction with experiment. In Chapter 4, extensive free energy calculations on a subset of the mutants designed in Chapter 3 are described. These calculations cover both methodological issues such as sampling and the use of constraints and lend insight to the role that the hydrophobic effect, packing in the protein interior, adjustment of the protein to mutation, and side chain entropy play in protein stability. The challenges awaiting theoretical study of proteins are discussed in Chapter 5 and future research directions presented.

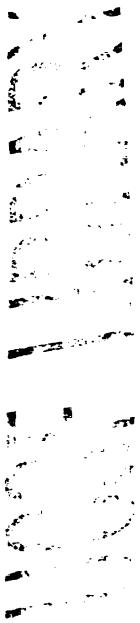
## Table of Contents

## List of Tables

## List of Figures

# Chapter 1

# Introduction



## Experimental Investigation of Protein Stability

An understanding of the intermolecular forces which determine protein folding and stability has been a principal goal of structural biology ever since Anfinsen showed proteins are capable of folding in solution based solely on the information contained in their amino acid sequence (Anfinsen *et al.*, 1961). The relatively modest stability of the folded state of a protein is essentially determined by large, opposing forces: the hydrophobic effect and conformational entropy (Dill, 1990). Obviously, a complete analysis of protein stability is more complicated than this. A variety of experimental techniques such as calorimetry (Privalov and Gill, 1989), UV spectroscopy (Elwell and Schellman, 1977), X-ray crystallography (Matthews, 1995), and more recently, multidimensional NMR spectroscopy (Englander, 1993) have been used to investigate the subtle balance of forces which leads to distinct folded states. The use of site-directed mutagenesis in conjunction with these techniques has been one of the most successful methods for examining the stability of proteins since the ability to substitute amino acids allow one to adjust the physical properties which determine the structure and stability of a protein. Small proteins such as barnase and T4 lysozyme have become paradigms for studying the effects of mutations on structure and the thermodynamics and kinetics of folding (Fersht, 1993; Matthews, 1995).

T4 lysozyme is the most thoroughly studied protein in terms of the structural and thermodynamic effects of amino acid mutation. There are approximately 152 crystal structures of T4 lysozyme mutants currently available in the Brookhaven Protein Databank (Matthews, 1995). Mutations have focused on the effects of adding disulfide bridges (Matsumura *et al.*, 1989), long-range electrostatic interactions (Daopin *et al.*, 1991), altering helix propensity (Blaber *et al.*, 1994), and nonpolar mutations (Eriksson *et al.*, 1992). These mutations have shown that the protein is quite amenable to changes in amino

acids. Large mutations such as insertions have resulted in foldable proteins that retain the tertiary structure of the native protein (Heinz *et al.*, 1994), while smaller mutants typically result in minor movement of the backbone and slight to moderate rearrangement of side chains (Baldwin *et al.*, 1993). Thus, T4 lysozyme is an ideal candidate for stability studies due to the wealth of information and conservative changes in mutant structures.

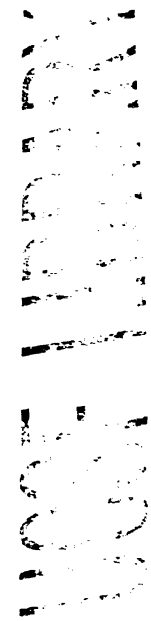
Mutations of T4 lysozyme, however, have been limited in the sense that changes made to the structure are restricted to the natural amino acids. A good example of this is the mutation of Leu 133 to Phe, which was done in an attempt to fill a hydrophobic cavity in the C-terminal domain. (Karpusas *et al.*, 1989) The Phe is not too large, but because the hybridization at the gamma carbon is $sp^2$ rather than $sp^3$, the side chain is placed in a position with unfavorable van derWaals contacts, resulting in a protein that is less stable than the native protein. It is difficult to change only one property of the side chain. Mutations involving Val or Ile cause changes in beta-branching as well as the size of the side chain. Schultz and coworkers have increased the specificity with which proteins can be engineered through the use of unnatural amino acids. The key to their success is the use of an amino-acylated tRNA which is designed to be inserted in response to the nonsense TAG codon (Noren *et al.*, 1989). These techniques have been used to study the stability of T4 lysozyme (Mendel *et al.*, 1992; Cornish *et al.*, 1994), catalysis of staphylococcal nuclease (Judice *et al.*, 1993), and conformational switching in *ras* p-21 (Chung *et al.*, 1993).

Despite the enormous amount of experimental work aimed at understanding the mechanisms by which proteins are able to form stable, unique tertiary structures, there remains unanswered a variety of questions. This is largely due to the difficulty in deconvoluting the contributions to protein stability from the various molecular forces involved in the folded and unfolded states based solely on static X-ray structures and

thermodynamic data. There is thus exceptional opportunity for theoretical methods to provide valuable insight in the study of protein stability.

## Theoretical Investigation of Protein Stability

Computational studies of proteins have generally used simple energetic models based on molecular mechanics. Scheraga and coworkers have attempted to derive the structure of folded proteins using a process of building amino acids onto a growing peptide chain in low energy conformations (Scheraga, 1981). These methods have met with difficulty, though, due to the enormous conformational space available to an amino acid polymer. Most studies using molecular mechanics force fields thus rely on obtaining the coordinates of the protein structure from X-ray crystallographic data. The first molecular dynamics simulations of a protein was published by McCammon and Karplus in 1977 (McCammon *et al.*, 1977). With the addition of water to the protein models, dynamics results were obtained which agreed well with experimental measures such as crystallographic B-factors (van Gunsteren and Berendsen, 1984). Global fluctuations of proteins (Heiner *et al.*, 1993), conformational changes upon substrate binding (Brown and Kollman, 1985), and the role of water in protein structure (Brooks and Karplus, 1985) are some of the topics that have been investigated with molecular dynamics simulations.



Lattice models have been developed which sacrifice a sophisticated description of the protein for the ability to more thoroughly explore conformational space. Dill has used these models to support the contention that the hydrophobic effect is the dominant force in protein folding (Dill, 1995); in contrast, others have emphasized the role of hydrogen bonding in the formation of secondary structure (Srinivasan and Rose, 1995). While lattice models lend themselves well to modeling the general thermodynamic properties of proteins and exploring conformational space, analysis at the atomic level is best carried out with molecular mechanics.

At the heart of a molecular mechanics modeling program is its force field. The parameters which describe the bonded and nonbonded interactions of the molecules must be derived from and tested against experimental results and high level theoretical calculations. The AMBER force field used by the Kollman group is based on deriving atomic centered point charges from *ab initio* quantum mechanics calculations, while the van der Waals (vdW) parameters have in general come from experimental data (Weiner *et al.*, 1986). Chapter 2 describes the *ab initio* calculations of the interaction of single water with various nonpolar hydrogens in order to develop refined vdW parameters of methyl hydrogens adjacent to electronegative groups. The motivation for this work is twofold: Ferguson and Kollman (1991) found that the methyl group of methylphosphonate oligonucleotides interacted with waters in a fairly polar fashion, and Gough *et al.* (1992) discovered that the r* vdW parameter of hydrogen for fluorinated hydrocarbons needed to be reduced to reproduce the densities of the liquid. The practice of reducing r* on methyl hydrogens adjacent to electronegative groups was utilized in the current force field (Cornell *et al.*, 1995) and in the calculations presented in Chapter 4.

The design and analysis of unnatural amino acids at position 133 in T4 lysozyme are presented in Chapter 3. This work is a result of mutual interest by the Schultz and Kollman groups in the C-terminal core of the protein, which contains a hydrophobic cavity. Molecular dynamics was used to determine the source of instability of the F133 mutant (Karpusas *et al.*, 1989) and to design side chains which would fill the hydrophobic cavity without creating bad vdW contacts, thus stabilizing the protein. Side chains which progressively removed nonpolar surface area were also simulated. Mendel *et al.* inserted the designed unnatural amino acids in T4 lysozyme and found that the protein indeed was stabilized by the larger, nonpolar side chains. Estimates of contributions from the hydrophobic effect, packing, and side chain entropy were made based on surface areas and

approximations of the relative conformational energies of the side chain rotamers. These estimates, however, were empirical in nature and did not clearly delineate the importance of the various factors, including the relative contributions from the folded and unfolded state.

Free energy calculations have the capability of providing important information about the free energy changes involved in protein folding (Sun *et al.*, 1996). This method allows for the calculation of the free energy difference between two similar states. It is based on using a molecular mechanics force field and molecular dynamics to generate ensembles of structures and evaluate their energies, which can then be used to find the free energy difference (Kollman, 1993). Free energy perturbation methods were originally applied to the solvation free energy differences of small molecules, the classic example being Jorgenson and Ravimohan's (1985) perturbation of methane to ethanol. It was quickly realized that this method could be applied in the study and design of enzyme inhibitors as well as in macromolecules (Bash *et al.*, 1987a,b). Dang *et al.* (1989) calculated the difference in stability between the native protein and the Thr157 $\rightarrow$ Val mutant in T4 lysozyme, using an extended tetrapeptide in solution as a model for the denatured state. Similar calculations have been conducted for electrostatic (Tidor, 1994) and hydrophobic (Prevost *et al.*, 1991; Sun *et al.*, 1996) mutations.

Despite the wide application of free energy calculations and the molecular insights gained from them, serious reservations about the technique have been raised in the literature recently. These concerns have focused on the need for increased sampling of conformational space in order to accurately calculate free energy differences (Mark *et al.*, 1994). It has been shown that hundreds of picoseconds of sampling are required to achieve consistent results for rather simple systems (Mitchell and McCammon, 1991). More recently, the practice of decomposing the free energy change into contributions from the force field, ie. electrostatic and vdW contributions, has been scrutinized (Smith and van

Gunsteren, 1994). And of most concern to this graduate student, the validity of using free energy calculations to determine protein has been specifically questioned (Yun-yu *et al.*, 1993). There are several major concerns. One is the reproducability of the calculations; perturbation calculations of macromolecules often give different results when conducted in the forward and reverse directions, most often due to structural hysteresis. Another problem is the representation of the denatured state. The protein stability calculations published to date have used extended peptides in solution as a model of the unfolded state, based on the assumption that the amino acid side chain is fully exposed to solvent. This assumption has been questioned, but the issue is more of matter of the interpretation of the physical nature of the unfolded state rather than an immediate flaw in the methodology. A more immediate and perhaps appropriate question is what influence does chain length have on the results. If the results are highly dependent on the peptide chain length, this would suggest that using an extended peptide may not be an accurate model even for a solvent exposed denatured state.

The free energy calculations of protein stability in Chapter 4 are presented with the above concerns in mind. The simulation time required for convergence of results was determined for the simulations conducted in gas phase, solution, and the folded protein. The structural and thermodynamic hysteresis for the mutation Nvl (norvaline) → Ala → Nvl was evaluated, and the small molecules dimethyl ether and propane were studied to test the force field parameters for the analogous side chain mutation. Various length peptides, from a dipeptide to a hexapeptide, were used to evaluate the denatured state model. In the folded protein, different constraint sets were used to determine the effects of backbone rearrangement on stability and the sensitivity of the calculations to constraint sets. And finally, side chain rotamers of the $\chi_1$ dihedral were sampled through the use of flatwell restraints. A number of issues raised by Yun-yu *et al.* have thus been addressed, and the conclusion from the work in Chapter 4 is that free energy calculations of protein stability,

while difficult, can provide valuable insights. The contributions from desolvation in the unfolded state, packing changes and backbone rearrangement in the folded protein, and side chain conformational freedom are determined and discussed.

A synopsis of the work presented in this thesis, from parameter development, to molecular modeling and design of proteins, to intensive free energy calculations, is given in Chapter 5. Areas of improvement for theoretical studies of protein stability are discussed with the hope that this work will inspire and guide more accurate and insightful studies of the forces which govern protein stability. And finally, proposals for studies with T4 lysozyme on protein unfolding, the stability of isolated protein fragments, and the accessible states of the protein backbone are presented.

## References


Anfinsen,C.B, Haber,E., Sela,M. and White,F.H.Jr. (1961) *Proc. Natl Acad. Sci. USA*, **47**, 1309-1314.

Baldwin,E.P., Hajiseyedjavadi,O., Baase,W.A. and Matthews,B.W. (1993) *Science*, **262**, 1715-1718.

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

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

Blaber,M., Zhang,X.J., Lindstrom,J.D., Pepiot,S.D. and Matthews,B.W. (1994) *J. Mol. Biol.*, **235**, 600-624.

Brown,F.K. and Kollman,P.A. (1985) *J. Mol. Biol.*, **198**, 533-546.

Brooks,C.L.3rd and Karplus,M. (1985) *J. Mol. Biol.*, **208**, 159-181.

Chung,H.H., Benson,D.R., Cornish,V.W. and Schultz,P.G. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 10145-10149.

Cornell,W.D., Cieplak,P., Bayly,C., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) *J. Am. Chem. Soc.*, **117**, 5179-5197.

Cornish,W., Kaplan,M., Veenstra,D., Kollman,P.A. and Schultz,P.G. (1994) *Biochemistry,* **33**, 12022-12031.

Dang,L., Merz,K.J. and Kollman,P. (1989) *J. Am. Chem. Soc.*, **111**, 8505-8508.

Daopin,S., Soderlind,E. Baase,W.A., Wazniak,J.A. and Matthews,B.W. (1991) *J. Mol. Biol.*, **221**, 873-887.

Dill,K.A. (1990) *Biochemistry*, **29**, 7133-7155.

Dill,K.A. (1995) *Protein Sci.*, **4**, 561-602.

Elwell,M.L. and Schellman,J.A. (1977) *Biochim. et Biophys. Acta*, **494**, 367-383.

Englander,S.W. (1993) *Science*, **262**, 848-849.

Eriksson,A.E., Baase,W.A., Zhang,X.-J., Heinz,D.W., Blaber,M., Baldwin, E.P., and Matthews,B.W. (1992)*Science* , **255**, 178-183.

Fergusson,D.M. and Kollman,P.A. (1991) *Antisense Res. and Dev.*, **1**, 243.

Fersht,A.R. (1993) *FEBS*, **325**, 5-16.

Gough,C.A., DeBolt,S.E. and Kollman,P.A. (1992) *J. Comput. Chem.*, **13**, 963-970.

Heiner,A.P., Berendsen,H.J.C. and van Gunsteren,W.F. (1993) *Protein Eng.*, **6**, 397-408.

Heinz,D.W., Baase,W.A., Zhang,X.J., Blaber,M., Dahlquist,F.W. and Matthews, M.W. (1994) *J. Mol. Biol.*, **236**, 869-886.

Jorgensen,W. and Ravimohan,C. (1985) *J. Chem. Phys.*, **83**, 3050.

Judice,J.K., Gamble,T.R., Murphy,E.C., de Vos,A.M. and Schultz,P.G. (1993) *Science*, **261**, 1578-1581.

Karpusas, M., Baase,W.A., Matsumura,M. and Matthews,B.W. (1989) *Proc. Natl Acad. Sci. USA,* **86**, 8237-8241.

Kollman,P.A. (1993) *Chem. Rev.*, **93**, 2395-2417.

Mark,A.E., van Helden,S.P., Smith,P.E., Janssen,L.H.M. and van Gunsteren,W.F. (1994) *J. Am. Chem. Soc.*, **116**, 6293-6302.

Matthews,B.W. (1995) *Adv. Prot. Chem.*, **46**, 249-278.

Matsumura,M., Signor,G. and Matthews,B.W. (1989) *Nature*, **342**, 291-293.

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

Mendel,D., Ellman,J.A., Chang,Z.Y., Veenstra,D.L., Kollman,P.A, and Schultz,P. (1992) *Science*, **256**, 1798-1802.

Mitchell,M.J. and McCammon,J.A. (1991) *J. Comput. Chem.*, **12**, 271- 275.

Noren,C.J., Anthony-Cahill,S.J., Griffith,M.C. and Schultz,P.G. (1989) *Science*, **244**, 182-186.

Prevost,M., Wodak,S., Tidor,B. and Karplus,M. (1991) *Proc. Natl Acad. Sci.USA*, **88**, 10880-10884.

Privalov,P.L. and Gill,S.J. (1989) *Pure and Appl. Chem.*, **61**, 1097-1104.

Scheraga, H. (1981) *Biopolymers* , **20**, 1877.

Smith,P. and van Gunsteren,W. (1994) *J. Phys. Chem.*, **98**, 13735-13740.

Srinivasan,R. and Rose,G.D. (1995) *Proteins*, **22**, 81-99.

Sun,Y.-C., Veenstra,D.L. and Kollman,P.A. (1996) *Protein Eng.*, **9**, 273-281.

Tidor,B. (1994) *Proteins*, **19**, 310-323.

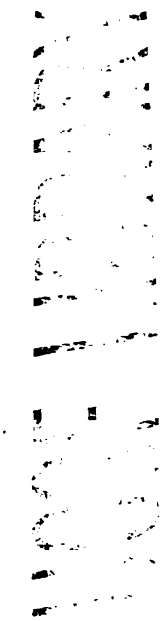
van Gunsteren,W.F. and Berendsen,H.J.C. (1984) *J. Mol. Biol.*, **176**, 559.

Weiner,S.J., Kollman,P.A., Nguyen,D.T. and Case, D.A. (1986) *J. Comput. Chem.*, **7**, 230-252.

Yun-yu,S., Mark,A.E., Wang,C.-x., Fuhua,H., Berendsen,H.J.C. and van Gunsteren,W.F. (1993) *Protein Engin.*, **6**, 289-295.

# Chapter 2

## Refinement of van der Waals Parameters for Nonpolar Hydrogens



# How Transferable Are Hydrogen Parameters In Molecular Mechanics Calculations ?

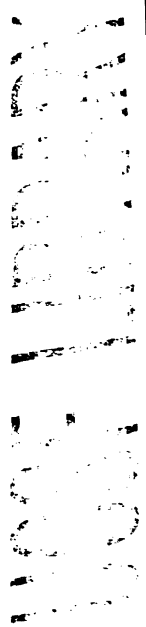
*David L. Veenstra, David M. Ferguson, and Peter A. Kollman*

*Department of Pharmaceutical Chemistry, University of California*
*San Francisco, CA 94143-0446*

*J. Comp. Chem.* **13,** 971-978 (1992).

## ABSTRACT

Interactions of water with various methyl hydrogens were studied using *ab initio* quantum mechanics and molecular mechanics. Our goal was to determine the effect of electronegative substituents on the nonbonded interactions of hydrogens. We found that methyl hydrogens are indeed very much affected by substituents and that standard van der Waals parameters used for these hydrogens in molecular mechanics do not reproduce these effects. Adjusted van der Waals parameters are presented which reproduce the *ab initio* values, and their use is discussed, as well as a general approach to van der Waals parameters on all types of hydrogens.

## INTRODUCTION

Empirical force field calculations have found widespread applications in chemistry and biochemistry over the last decade.[1] This is partly due to the computational expedience of the method that brings macromolecular and condensed phase calculations into the reach of the computational chemist. The accurate simulation of these complex systems, however, depends upon the accuracy of the force field (potential function and parameters) which is often developed from molecular mechanics (MM) studies of smaller molecules and fragments and transferred to larger systems of interest. The main assumption is that molecular interactions of specific types can be more or less generalized from structure to structure. Of course, this allows for the examination of chemical systems for which there is limited experimental data or structural information, but as one might guess, could have adverse consequences if environmental effects that are local to a particular system are not accounted for in the parameterization. The force fields MM2 and MM3 developed by Allinger and coworkers allow for various effects to be accounted for in the force field (e. g. electronegativity effects on bond lengths and anomeric effects).[2,3] These adjustments are directed toward optimizing the intramolecular properties of the molecule. While there is no doubt that this is an important contribution, our efforts have been focused more on intermolecular interactions and properties. A recent contribution in this area has been the development of a polarizable water model for inclusion in the Weiner *et al.* force field.[4]

Intermolecular interactions are determined by the electrostatic and van der Waals terms in molecular mechanics force fields, as well as special hydrogen bonding terms in some force fields. The electrostatic potential surface of a molecule, which is represented by atom-centered partial charges in the AMBER force field, is determined mainly by the valence electrons and is highly variable. Therefore, we cannot assume that these partial

charges are transferable between atoms in different molecules or fragments. However, *ab initio* calculations can be conducted on the molecules or fragments used in MM calculations at a high enough level to generate accurate potential surfaces. Fitting partial charges to these electrostatic potential surfaces has provided a systematic way of dealing with parameters which are not, in most cases, transferable.[5]

Van der Waals interactions, in contrast to electrostatics, are not as variable because they are largely influenced by the inner shell electrons. Molecular mechanics force fields often utilize the Lennard-Jones function to describe van der Waals interactions;

$$E_{vdw} = \frac{n\varepsilon^*_{i,j}}{(n-m)}\left\{\frac{m}{n}\left[\frac{R^*_{i,j}}{R}\right]^n - \left[\frac{R^*_{i,j}}{R}\right]^m\right\}, \qquad (1)$$

where m = 6 and n = 12 for the "6-12" potential. R defines the interatomic seperation of atoms i and j, $R^*_{i,j}$ is the sum of the van der Waals radii for the atom pair ($R^*_{i,j} = R^*_i + R^*_j$) and $\varepsilon^*_{i,j}$ is the geometric mean in the van der Waal well depths ($\varepsilon^*_{i,j} = (\varepsilon_i\varepsilon_j)^{1/2}$).[6] It has generally been assumed that van der Waals parameters are transferable for atoms with the same hybridization, ie. all tetrahedral carbons can be assigned the same van der Waals parameters.[7,8] These assumptions seem to be fairly accurate for heavy atoms. Problems arise, however, for hydrogens, which have no inner shell of electrons. Hydrogens attached to electronegative atoms, such as oxygen and nitrogen, have often been given zero vdw parameters, allowing electrostatics to control hydrogen bonding and the hetero atom to provide the repulsion.[8,9] This method, however, may require unphysically large vdw parameters on the heteroatoms to avoid hydrogen bonds which are too short.[7,10,11] The Weiner *et al.* force field uses a 10-12 Lennard-Jones function to fine tune hydrogen bond distances and energies[7]. But what about hydrogens attached to carbon? These hydrogens are given non-zero vdw parameters, which contribute significantly to

intermolecular interactions. The elements C, N, and O are increasingly electronegative, and it seems inconsistent to suddenly entirely remove the vdw potential from hydrogen on going from carbon to nitrogen. Thus, the general question arises, how should optimal vdw parameters for hydrogens be determined?

We recently became interested in the transferability of van der Waals parameters for methyl hydrogens during several independent studies. Hydrogen van der Waals parameters did not appear to be completely transferable while conducting liquid studies of alcohols and fluorocarbons.[12,13] We found that the densities were very sensitive to the choice of the methyl hydrogen van der Waals parameter. Some of the details of this work are described in the accompanying paper by Gough *et al.* Our second indication came from a study of methylphosphonate oligonucleotides.[14] Here, we found that the methyl group of the modified phosphate linkage was not hydrophobic and that the methyl hydrogen-oxygen water nonbonded distance was significantly less than that predicted for hydrocarbon methyl groups. It was clear that substituents attached to carbon affect the nonbonded interactions of the methyl hydrogens, and that these hydrogens appeared to require vdw parameters intermediate to the values used for a standard methyl hydrogen and those used for a hydrogen attached to nitrogen.

We seek a consistent approach for assigning hydrogen vdw parameters where the assignment is made based on the electronegativity of the attached hetero atom and its substituents. Studying the interaction of molecules containing methyl hydrogens with a single water molecule should enable us to develop this approach; this interaction may also be of great importance in relation to the hydrophobic effect and its consequences in biological systems.[15] Thus, we will examine several water...methyl-hydrogen interactions both quantum mechanically and classically, using the *ab initio* values as a guide to the effects of the substituents on nonbonded interactions. Two sets of vdw parameters for HC and

CT atom types will be used, one from the Weiner *et al.* force field and a more recent set derived from liquid hydrocarbon simulations.[16] The transferability of methyl hydrogen van der Waals parameters will be discussed as well as approaches to adjust these parameters to reproduce "correct" nonbonded distances.

## METHODS

*Ab initio* quantum mechanics (QM) calculations carried out at the 6-31G* basis set level have been shown to give reasonable intermolecular distances and energies, and are often used in the development of molecular mechanics force fields.[17,18,19] Single point QM calculations were thus conducted at the 6-31G* level using either Gaussian 80/UCSF or Gaussian 90.[5,20] The water molecules were placed with the water oxygen pointed toward the methyl hydrogen of interest, and the water hydrogens oriented to avoid any strong repulsions (Figure 1). The minimum energy distances between methyl hydrogens and water oxygens were found by sampling the energies at 0.05 Å intervals (G80) or by optimizing the distance (G90), and are reported relative to the isolated molecules. Standard geometries were used for the model built structures.[21,22,23] The water molecule had TIP3P geometry.[24]

The ANAL module of AMBER 3.0 Rev. A was used for the molecular mechanics calculations. The same geometries were used as for the QM calculations, and the energies were sampled at 0.05 Å intervals. The partial charges were fit to the electrostatic potential generated by 6-31G* *ab initio* calculations on the isolated molecules using Gaussian 80/UCSF, and are shown in Figure 2. Valence and vdw parameters were taken from the Weiner *et al.* force field; note that the only parameters which should effect the results are the nonbonds since the reported energies are relative to the isolated molecules.

The chlorine and fluorine vdw parameters used were R* = 2.25 Å, ε = 0.20 kcal/mol and R* = 1.55 Å, ε = 0.08 kcal/mol, respectively. Two sets of vdw parameters for the HC and CT atom types were used, one from the Weiner *et al.* force field (set A) and the other derived from liquid hydrocarbon studies (set B). The Weiner *et al.* force field uses R* = 1.540 Å and ε = 0.0100 kcal/mol for HC and R* = 1.800 Å and ε = 0.0600 kcal/mol for CT whereas the values obtained from the liquid hydrocarbon studies are R* = 1.487 Å and ε = 0.0157 kcal/mol for HC and R* = 1.908 Å and ε = 0.1094 kcal/mol for CT. The parameters derived from the liquid simulations gave accurate densities and enthalpies of vaporization for hydrocarbons. All calculations used a dielectric constant of 1.0.

## RESULTS AND DISCUSSION

### Quantum Mechanics

The interaction energies and distances of water with methyl hydrogens are given in Table 1. Electron withdrawing groups attached to methyl groups would be expected to pull electron density away from the methyl hydrogens, thus allowing for a closer approach of the water molecule. Conversely, electron donating groups should increase the distance. This trend is seen in the series $CH_4$, $CH_3CH_3$, $CH_3NH_2$, $CH_3OH$, $CHF_3$. The methyl substituent, in ethane, being electron donating, slightly increases the nonbonded interaction distance compared to methane. The methyls substituted by the increasingly electronegative elements nitrogen, oxygen and fluorine have progressively shorter nonbonded distances. The results for methylamine varied depending on which hydrogens were sampled; the two methyl hydrogens anti to the HN's gave different interactions than the methyl hydrogen gauche to both HN's. These differences are due to electrostatic interactions between the water molecule and the unsymmetrical $-NH_2$ substituent. Overall, however, methyl amine appears to produce HC--OW distances between those of ethane

and methanol. The same problem of unequivalent hydrogens arises in methanol, but the differences are not quite as dramatic, ie. 2.75 Å vs. 2.70 Å. Flourine, the most electronegative element, produces the shortest nonbonded distance of the second row substituents. Adding more fluorines further strengthens the nonbonded interaction; each fluorine changes the distance by about 0.2 Å. The shortest nonbonded distance observed for the molecules studied was for methyl ammonium. The positive charge is by far more electron withdrawing than the uncharged substituents; the interaction is stronger than most hydrogen bonds[18].

Interestingly, methyl groups substituted by the third row elements sulfur and chlorine have shorter nonbonded distances than for methanol and fluoromethane. This seems counterintuitive since sulfur and chlorine are less electronegative than oxygen and fluorine. However, the third row elements are more polarizable than the second row elements, and with a longer C-X bond distance the electron density of sulfur or chlorine is farther removed from the water oxygen, decreasing the electrostatic repulsion. The interaction of water with the methyl hydrogen on methanethiol which is anti to the sulfur hydrogen is almost twice as strong as the equivalent interaction for methanol. The other hydrogens of methanethiol also have stronger interactions than those of methanol, but the differences are not as great.

The nonbonded interaction distances for methyls adjacent to carbonyl groups, in acetaldehyde and N-methyl acetamide (NMA), were significantly shorter than for methane or ethane. The carbonyl group apparently has about the same effect on the interaction with water as fluorine or chlorine. The carbonyl carbon in NMA, being adjacent to nitrogen, is not as electron poor as the one in acetaldehyde, and thus is less electron withdrawing. The N-methyl hydrogens studied in NMA gave weaker interactions than for the acetyl group, but still stonger than for methane. It should be noted that the hydrogens eclipsed with the carbonyl oxygen in NMA and acetaldehyde were not studied. These hydrogens

would likely give longer nonbonded distances due to electrostatic repulsion between the water and carbonyl oxygens.

## Molecular Mechanics

The parameter set derived from liquid hydrocarbon simulations (set B) gave results which agreed more closely with the *ab initio* calculations than did results based on the parameter set of Weiner *et al.* (set A), although the differences were small. Parameter set B, with a deeper well (0.0157 vs 0.0100 kcal/mol) and shorter R* (1.487 vs 1.540 Å) for HC atom types produced consistently stronger interactions. Both sets, however, led to interaction energies that were too weak and intermolecular distances that were too long compared to the *ab initio* results. Emphasis will be place on the results from parameter set B since these parameters are believed to be more accurate, at least for hydrocarbons, than those of set A. The parameters of set B were derived from studies of pure hydrocarbons; they should be expected to give accurate nonbonded interactions for methyl hydrogens that are in a hydrocarbon-like environment.

Distances and especially energies were reproduced fairly well for NMA using parameter set B. This is encouraging since NMA serves as a simple model for peptide structure, and has been used as a test case for the development of MM parameters.[8,18,25] Thus, standard vdw parameters used for methyl hydrogens appear to be roughly appropriate for methyl groups in such an environment. Parameter set B reproduced the nonbonded distances to within 0.1 Å and the energies to within about ten percent. Both parameter sets reproduced the interactions of the acetyl methyl hydrogens of NMA better than the methyl hydrogens of acetaldehyde, most likely because the carbonyl in acetaldehyde is more electron withdrawing.

Not surprisingly, the interactions for all methyls substituted by electronegative groups

were reproduced least accurately. The distances did decrease the more electronegative the substituents were, but not enough to compensate for having a standard vdw parameter on HC. An excellent example of this is the series $CH_4$, $CH_3F$, $CH_2F_2$, and $CH_3F$. The discrepancy between *ab initio* and MM distances increases from 0.05 Å to 0.30 Å; the agreement gets worse the more electronegative the substituents. The interactions for methylamine, methanol, chloromethane, methanethiol, acetaldehyde, and methylammonium are also poorly reproduced, all of which would be expected to have shorter nonbonded distances and stronger interactions. It is clear that a standard vdw parameter on these hydrogens prevents stronger interactions from occuring.

As mentioned, no minima were found for certain methyl hydrogens on methylamine, methanol, and methanethiol. This is, in our opinion, due to the fact that the electrostatic potential near the methyl hydrogens of these molecules is dominated by the potentials from the -OH, $-NH_2$, and -SH substituents. For example, in methanol, the largest potentials felt by an approaching molecule will be the positive potential due to the hydroxyl hydrogen and the negative potential in the oxygen "lone pair region". Thus, water molecules approaching the methyl hydrogens may be in a region of negative potential dominated by the heteroatom lone pair potential and be electrostatically repelled. The *ab initio* calculations compensate for this somewhat with attractive polarization and charge transfer for $H_2$-O...H-C interactions, but in molecular mechanics, no such terms are included. Thus, the MM studies of water...methyl hydrogen interactions in $CH_3OH$, $CH_3NH_2$, and $CH_3SH$ should be considered suggestive only. The definitive analysis comes from the water...chloro- and fluorocarbon studies where the ambiguities from -OH, $-NH_2$, and -SH orientations do not exist.

The next step in our study was to find vdw parameters for the methyl hydrogens which would reasonably reproduce the *ab initio* results and hopefully produce some type of

minima for the troublesome hydrogens. The vdw parameters for the HC atom type which gave distances that agreed with the 6-31G* distances to within 0.05 Å are given in table II. In every case but one the value of R* had to be decreased. The only exception was for parameter set B; the hydrogen gauche to both HN's on methyl amine needed no adjusting. The magnitude of the decrease in R* was approximately the same as the difference between the *ab initio* and MM distances, ie. R* had to be decreased from 1.487 to 1.150 Å for trifluoromethane in order to decrease the distance from 2.60 Å to 2.30Å (see Figure 3). With R* adjusted to produce distances in agreement with the *ab initio* calculations, the discrepancy in the energies was reduced. NMA methyl hydrogens had interaction energies accurate to within 0.08 kcal/mol for parameter set A and within 0.05 kcal/mol for parameter set B. Most of the other energies were less negative than the QM energies by 0.1 to 0.3 kcal/mol. The agreement for the fluorine series decreases as fluorines are added, suggesting that slight changes in well depths may be required. Overall, parameter set B gave energies which were closer to the *ab initio* values. The only interactions at this point that were in poor agreement with the QM values were the troublesome ones mentioned earlier; the hydrogens anti to HS and HO in methanethiol and methanol, and the hydrogens anti to HN in methylamine. No reasonble minima could be found for these interactions, likely due to the reasons previously mentioned.

## CONCLUSION

Methyl group hydrogens are often given standard vdw parameters despite the fact that they are in different chemical environments. While these standard parameters appear to be fairly sufficient for methyl hydrogens in hydrocarbon side chains or in the protein backbone, we have shown that what one would intuitively assume is true; electronegative groups can have a dramatic effect on the electron density of hydrogens. Thus, changes must be made to the vdw parameters of hydrogens which are influenced by

electronegative groups to account for this difference.

Van der Waals parameters which reproduce distances from *ab initio* calculations are presented here. We have found that R* must be reduced by ~0.05 Å up to ~0.50 Å depending on the combined electronegativity of the substituents. But rather that being refined parameters for use in molecular mechanics, molecular dynamics (MD), and free energy perturbation (FEP) calculations, these parameters are meant more as a guide for exploration into the effects of substituents on vdw parameters. The *ab initio* calculations themselves, at the 6-31G* level, are accurated only to a point; they exaggerate electrostatics, but neglect dispersion. Determining how these offsetting factors affect the nonbonded distances is difficult.

In addition, these calculations have sampled only one approach of water to the hydrogens, and used fixed geometries. Thus, the vdw parameters were not adjusted to precisely reproduce the *ab intio* values. Liquid simulations, which sample many configurations and nonbonded interactions, can be used to adjust parameters to reproduce experimental densities and heats of vaporization. In the accompanying paper by Gough *et al.*, liquid simulations were run for tetrafluoromethane and trifluoromethane. The R* for the hydrogen in trifluoromethane required to reproduce experimental data is very close to the value derived in this study (1.21 vs. 1.15 Å). Further liquid simulations are in progress to refine parameters for methanol and NMA[13]. Free energy perturbation calculations can also be used to test parameters. In fact, FEP calculations using the HC and CT vdw parameters derived from the liquid hydrocarbon simulations have produced numbers in excellent agreement with experiment.[26]

We have shown in this study that hydrogen parameters should be correlated with the electronegativity of attached atoms. Using standard vdw parameters for methyl hydrogens which are adjacent to electronegative groups results in nonbonded interactions that

are too repulsive. The more electronegative the substituents within a given row of the periodic table, the smaller R* should be. It appears that the third row elements have a greater effect than the second row elements, so one cannot use electronegativity alone as a guide to quantitatively determining the optimum vdw parameters; ties must be made with experiment or high level quantum mechanics. The vdw parameters of hydrogens other than those attached to carbon can also be assigned on the same basis. Hydrogens attached directly to nitrogen would require a smaller R* than methyl hydrogens since nitrogen is more electronegative than carbon and in the same row of the periodic table. Studies of hydrogen bonding in amides have shown that an R* of 0.6 Å is required on HN atom types in order to reproduce *ab initio* and experimental distances when OPLS vdw parameters are used on the nitrogen along with 6-31G* electrostatic potential derived charges.[27] Substituent effects would also likely play an important role for the amides. Hydroxyl hydrogens, being attached to an even more electronegative atom, would not require vdw parameters, as with TIP3P water.[24] This general approach, based on the electronegativity of atoms $\alpha$ and $\beta$ to hydrogen, allows for a more consistent and systematic assignment of vdw parameters for all types of hydrogens, and has a sound physical basis rather than being based on the needs of a particular force field.

Acknowledgement: We gratefully acknowledge the support of the NIH through a Biotechnology Training Grant (GM-08388-02) to D.L.V. and grant GM-29072 to P.A.K., the San Diego Super Computer Center for computer time, and the U.C.S.F. Computer Graphics Labratory, supported by NIHRR-1081 to Robert Langridge.

## References


1. W.F. Van Gunsteren and H.J.C. Berendsen, *Angew. Chemie*, **29** 992 (1990).

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

3. a. J.-H. Lii and N.L. Allinger, *J. Comp. Chem.*, **12,** 186 (1991). b. N.L. Allinger, *J. Am. Chem. Soc.*, **111,** 8551 (1989).

4. J. Caldwell, L.X. Dang, and P.A. Kollman, *J. Am. Chem. Soc.*, **112,** 9144 (1990).

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

6. U. Burkert and N.L. Allinger, *Molecular Mechanics,* ACS Monograph No. 177, American Chemical Society, Washington, D.C. (1982).

7. a. S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case, *J. Comp. Chem.*, **7,** 230 (1986). b. S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, *J. Am. Chem. Soc.*, **106,** 765 (1984).

8. W.L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110,** 1657 (1988).

9. A.T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96,** 5319 (1974).

10. J. Blaney, P. Weiner, A. Dearing, P. Kollman, W. Jorgensen, S. Oatley, J. Burridge, and C Blake, *J. Am. Chem. Soc.*, **104,** 6424 (1982).

11. G. Wipff, A. Dearing, P. Weiner, J. Blaney, and P. Kollman, *ibid.*, **105,** 997 (1983).

12. C.A. Gough, S. DeBolt, and P.A. Kollman, *J. Comp. Chem,* submitted.

13. S. DeBolt and P.A. Kollman, work in progress.

14. D.M. Ferguson and P.A. Kollman, *Antisense Res. and Dev.*, **1,** 243 (1991).

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

16. D. Spellmeyer and P.A. Kollman, unpublished results.

17. W.J. Hehre, L. Radom, P.v.R. Schleyer, and J.A. Pople, *Ab Initio Molecular Orbital Theory,* Wiley, New York (1986).

18. P.A. Kollman, A. Howard, S.J. Weiner, and U.C. Singh, in *Three-Dimensional Structures and Drug Action,* Y. Iitaka and A. Itai, Eds., Univ. of Tokyo Press, Tokyo, 1987.

19. J. Pranata, S.G. Wierschke, and W.L. Jorgensen, *J. Am. Chem. Soc.,* **113,** 2810 (1991).

20. M.J. Frisch, M. Head-Gordon, G.W. Trucks, J.B. Foresman, H.B. Schlegel, K. Raghavachari, M.A. Robb, J.S. Binkley, C. Gonzalez, D.J. Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius, J. Baker, R.L. Martin, L.R. Kahn, J.J.P. Stewart, S. Topiol, and J.A. Pople, Gaussian, Inc., Pittsburgh PA (1990).

21. A. Streitwieser and C. Heathcock, *Introduction to Organic Chemistry,* 3rd ed., Macmillan Publishing Co., New York (1985).

22. A.J. Gordon and R.A. Ford, *The Chemist's Companion,* John Wiley and Sons, New York (1972).

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

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

25. W.L. Jorgensen and J Gao, *J. Am. Chem. Soc.,* **110,** 4212 (1988).

26. Y. Sun and P.A. Kollman, *J. Phys. Chem.*, submitted.

27. D.M. Ferguson and P.A. Kollman, unpulished results.

## Figure Captions

**Figure1.** Orientation of methane and TIP3P water

**Figure2.** Partial charges, in fractions of an electron, of molecules studied.

**Figure 3.** Energy vs. distance for $F_3C$-H···O-$H_2$ interaction for parameter set B (·····), parameter set B with adjusted R* for HC atom type (- - - ), and *ab initio* calculations (——).

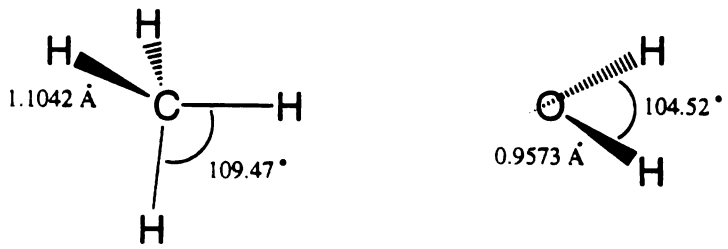



**Figure 1**

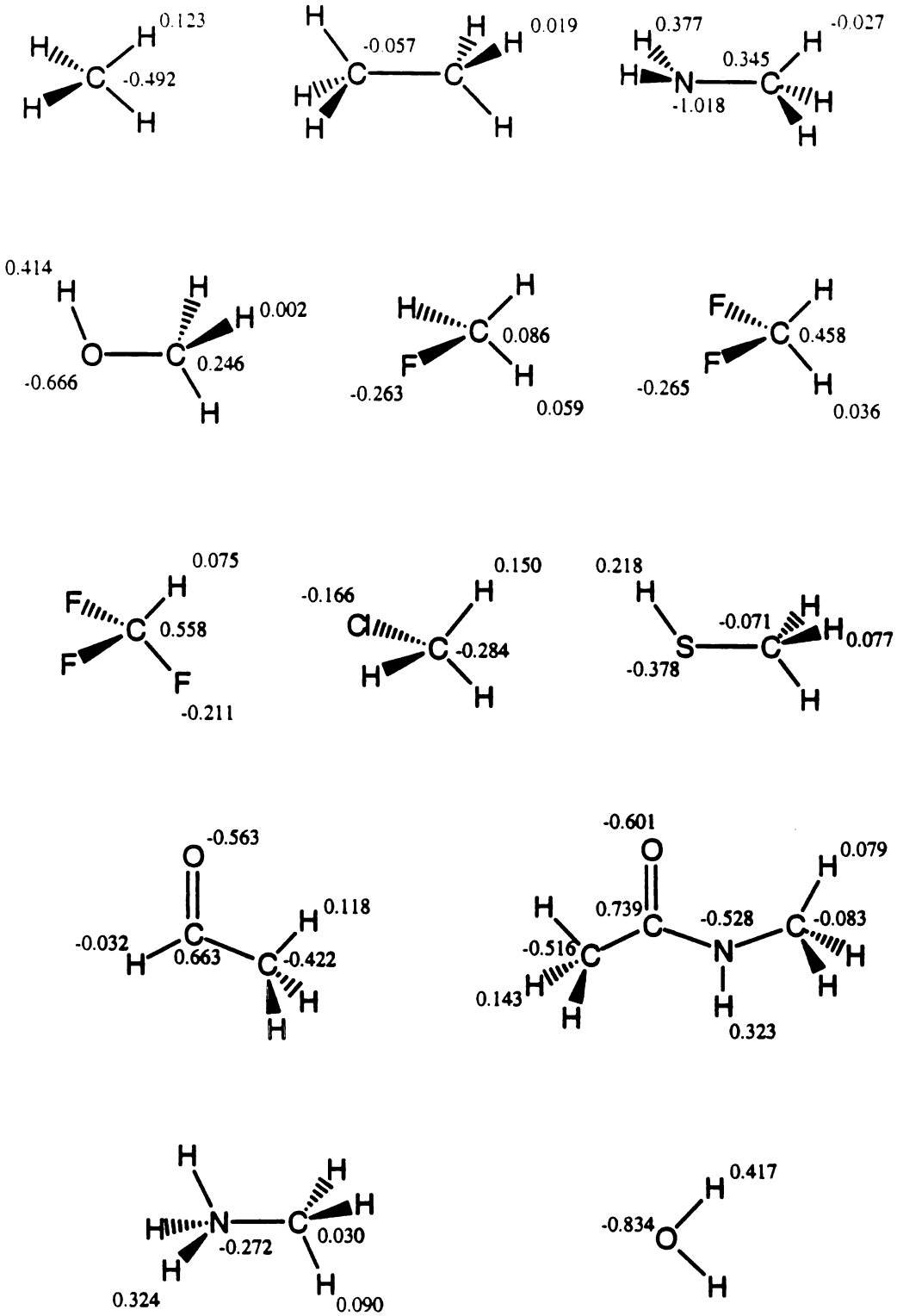

0.123
-0.492
-0.057
0.019
0.377
-0.027
0.345
-1.018
0.414
0.002
-0.666
0.246
0.086
-0.263
0.059
0.458
-0.265
0.036
0.075
0.558
-0.211
0.150
-0.166
-0.284
0.218
-0.071
0.077
-0.378
-0.563
0.118
-0.032
0.663
-0.422
-0.601
0.079
0.739
-0.528
-0.083
-0.516
0.143
0.323
0.030
-0.272
0.324
0.090
0.417
-0.834

**Figure 2**

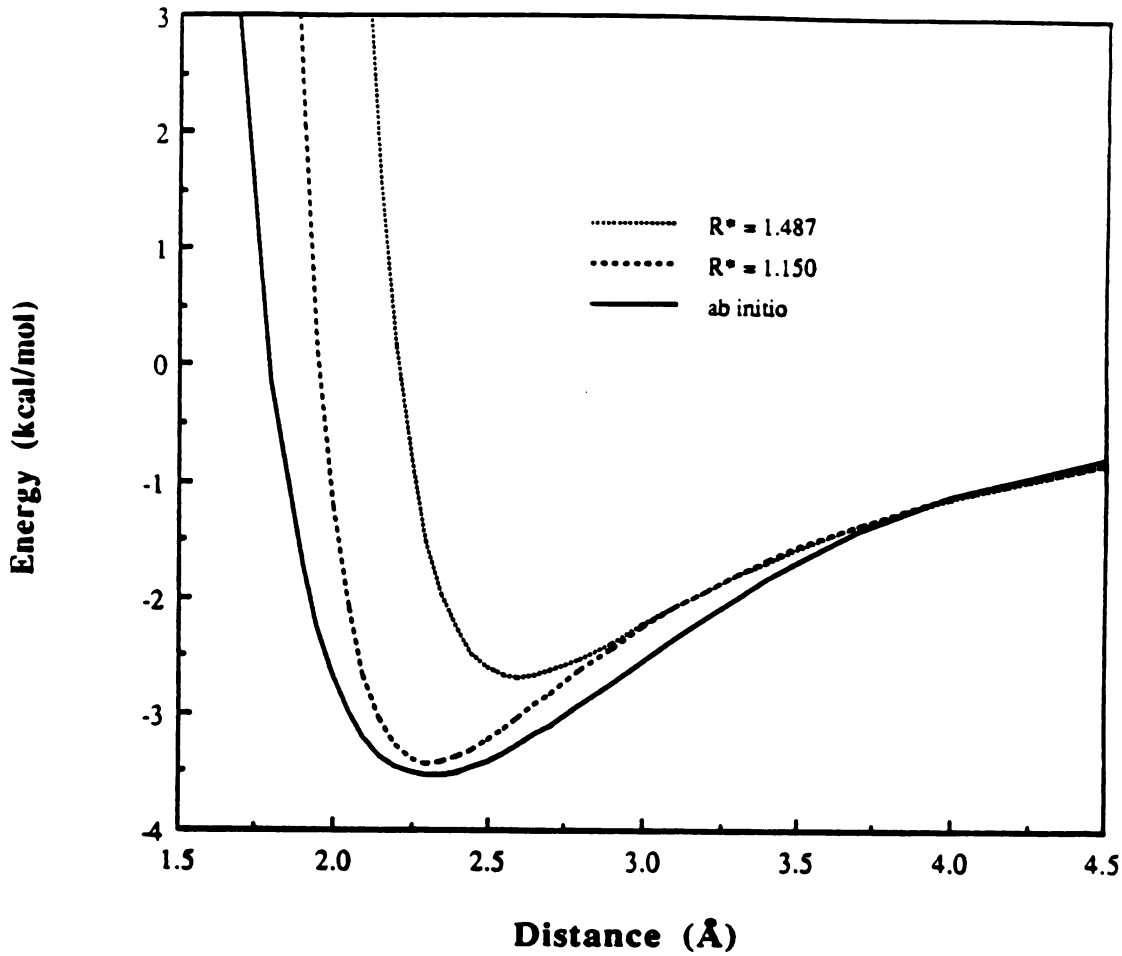

Energy (kcal/mol)
3
2
1
0
-1
-2
-3
-4
1.5
2.0
2.5
3.0
3.5
4.0
4.5
Distance (Å)
R* = 1.487
R* = 1.150
ab initio

**Figure 3**

**Table I.** Interaction Energies and Distances of Water with Methyl Hydrogens[a]

| Molecule | Energy (kcal/mol) | | | Distance HC--OW (Å) | | |
|---|---|---|---|---|---|---|
| | *ab initio*[b] | A[c] | B[d] | *ab initio* | A | B |
| $CH_4$ | -0.60 | -0.32 | -0.38 | 2.85 | 2.90 | 2.90 |
| $CH_3CH_3$ | -0.54 | -0.17 | -0.24 | 2.90 | 3.05 | 3.00 |
| $CH_3NH_2$[e] | -0.76 | -0.65 | -0.71 | 2.92 | 2.95 | 2.90 |
| $CH_3NH_2$[f] | -0.64 | -- | -- | 2.81 | -- | -- |
| $CH_3OH$[g] | -1.09 | -0.85 | -0.91 | 2.75 | 2.85 | 2.80 |
| $CH_3OH$[h] | -0.82 | -- | -- | 2.70 | -- | -- |
| $CH_3F$ | -1.43 | -1.06 | -1.13 | 2.65 | 2.75 | 2.75 |
| $CH_2F_2$ | -2.46 | -1.89 | -1.96 | 2.45 | 2.65 | 2.65 |
| $CHF_3$ | -3.55 | -2.64 | -2.70 | 2.31 | 2.60 | 2.60 |
| $CH_3Cl$ | -1.83 | -1.42 | -1.48 | 2.51 | 2.70 | 2.70 |
| $CH_3SH$[g] | -1.33 | -1.02 | -1.09 | 2.63 | 2.80 | 2.75 |
| $CH_3SH$[h] | -1.61 | -0.57 | -0.63 | 2.53 | 2.80 | 2.80 |
| $CH_3CHO$ | -1.99 | -1.72 | -1.78 | 2.56 | 2.70 | 2.70 |
| $CH_3CONHCH_3$ | -1.15 | -1.03 | -1.10 | 2.75 | 2.80 | 2.80 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $CH_3CONHCH_3$ | -1.76 | -1.60 | -1.67 | 2.65 | 2.75 | 2.70 |
| $CH_3NH_3^+$ | -10.30 | -8.29 | -8.35 | 2.15 | 2.50 | 2.50 |

[a] Oxygen of water was oriented toward hydrogen shown in bold. Distances are between methyl hydrogen (HC) and water oxygen (OW).

[b] Calculated using 6-31G* basis set.

[c] Calculated using Amber3a, non-bond parameters of R*= 1.540 and ε = 0.010 for HC and R* = 1.800 and ε = 0.0600 for CT. Distances were sampled at 0.05 Å intervals.

[d] Same as (b), but using HC and CT non-bond parameters derived from liquid hydrocarbon studies; R*= 1.487 and ε = 0.0157 for HC and R* = 1.908 and ε = 0.1094 for CT.

[e] Water oriented off of HC gauche to both HN's.

[f] Water oriented off of HC gauche to only one HN. No MM minimum was found.

[g] Water oriented off of HC gauche to H of hetero atom.

[h] Water oriented off of HC anti to H of hetero atom. No MM minimum was found for methanol.

**Table II.** Interaction Energies and Distances with Adjusted R* for HC Atom Type.[a]

| Molecule | Parameters | | Molecular Mechanics | | *ab intio* (6-31G*) | |
|---|---|---|---|---|---|---|
| | Set[b] | $R^*_{HC}$ (Å)[c] | dist. (Å)[d] | E(kcal/mol) | dist. (Å)[d] | E(kcal/mol) |
| $CH_4$ | A | 1.50 | 2.85 | -0.34 | 2.85 | -0.60 |
| | B | 1.45 | | -0.40 | | |
| $CH_3CH_3$ | A | 1.40 | 2.90 | -0.18 | 2.90 | -0.54 |
| | B | 1.40 | | -0.25 | | |
| $CH_3NH_2$[e] | A | 1.50 | 2.90 | -0.67 | 2.92 | -0.76 |
| | B | 1.49 | | -0.71 | | |
| $CH_3NH_2$[f] | A | -- | -- | -- | 2.81 | -0.64 |
| | B | -- | | -- | | |
| $CH_3OH$[g] | A | 1.45 | 2.75 | -0.90 | 2.75 | -1.09 |
| | B | 1.42 | | -0.95 | | |
| $CH_3OH$[h] | A | -- | -- | -- | 2.70 | -0.82 |
| | B | -- | | -- | | |
| $CH_3F$ | A | 1.40 | 2.65 | -1.18 | 2.65 | -1.43 |
| | B | 1.40 | | -1.20 | | |
| $CH_2F_2$ | A | 1.30 | 2.45 | -2.25 | 2.45 | -2.46 |
| | B | 1.25 | | -2.31 | | |
| $CHF_3$ | A | 1.20 | 2.30 | -3.41 | 2.31 | -3.55 |
| | B | 1.15 | | -3.44 | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| $CH_3Cl$ | A | 1.35 | 2.50 | -1.65 | 2.51 | -1.83 |
| | B | 1.25 | | -1.78 | | |
| $CH_3SH$[g] | A | 1.40 | 2.65 | -1.13 | 2.63 | -1.33 |
| | B | 1.35 | | -1.19 | | |
| $CH_3SH$[h] | A | -- | -- | -- | 2.53 | -1.61 |
| | B | -- | | -- | | |
| $CH_3CHO$ | A | 1.35 | 2.55 | -1.95 | 2.56 | -1.99 |
| | B | 1.30 | | -2.01 | | |
| $CH_3CONHCH_3$ | A | 1.50 | 2.75 | -1.07 | 2.75 | -1.15 |
| | B | 1.45 | | -1.13 | | |
| $CH_3CONHCH_3$ | A | 1.45 | 2.65 | -1.69 | 2.65 | -1.76 |
| | B | 1.45 | | -1.71 | | |
| $CH_3NH_3^+$ | A | 1.10 | 2.15 | -9.81 | 2.15 | -10.30 |
| | B | 1.00 | | -9.98 | | |

[a] R* for HC atom type was adjusted so that the *ab initio* and molecluar mechanics distances agreed to within 0.05 Å. The oxygen of water was oriented toward the hydrogen shown in bold.

[b] Parameter set A is from Weiner *et al.*[7a]. Set B is the same except it uses HC and CT atom type vdw parameters derived from liquid hydrocarbon studies.[16]

[c] R* for HC atom type required for agreement with *ab initio* distances.

[d] Distance between HC and OW.

[e] Water oriented off of HC gauche to both HN's.

[f] Water oriented off of HC gauche to only one HN. No reasonable MM minimum was found.

[g] Water oriented off of HC gauche to H on hetero atom.

[h] Water oriented off of HC anti to H on hetero atom. No reasonable MM minimum was found.

# Chapter 3

## Molecular Dynamics of T4 Lysozyme Mutants: Design and Analysis of Unnatural Amino Acid Side Chains

# Probing Protein Stability with Unnatural Amino Acids

David Mendel, Jonathan A. Ellman, Zhiyuh Chang, David L. Veenstra, Peter A. Kollman, Peter G. Schultz*

D. Mendel, J.A. Ellman, Z. Chang, P.G. Schultz, Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, and Center for Advanced Materials, Lawrence Berkeley Laboratory, Berkeley, CA 94720.

D.L. Veenstra and P.A. Kollman, Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94143.

*To whom correspondence should be addressed

*Science* **256**, 1798-1802 (1992)

# Abstract

Unnatural amino acid mutagenesis, in combination with molecular modeling and simulation techniques, was used to probe the effect of side chain structure on protein stability. Specific replacements at position 133 in T4 lysozyme included (i) leucine (wt), norvaline, ethyl-glycine, and alanine to measure the cost of stepwise removal of methyl groups from the hydrophobic core, (ii) norvaline and O-methyl serine to evaluate the effects of side chain solvation, and (iii) leucine, S,S-amino-4-methylhexanoic acid, and S-2-amino-3-cyclopentylpropanoic acid to measure the influence of packing density and side chain conformational entropy on protein stability. All of these factors (hydrophobicity, packing, conformational entropy, and cavity formation) significantly influence protein stability and must be considered when analyzing any structural change to proteins.

Mutational studies of the amino acids that form the hydrophobic core of proteins are beginning to define how these residues influence protein structure and stability (1-10). However, it is difficult to make mutations with the natural 20 amino acids that perturb one interaction without simultaneously affecting several others. For example, mutation of $Leu^{133} \rightarrow$ Phe or $Ala^{129} \rightarrow$ Val in T4 lysozyme (T4L) in an attempt to increase packing density, and as a consequence, thermal stability, resulted in a less stable protein because of a local increase in strain energy (11). The ability to site-specifically incorporate unnatural amino acids into proteins makes possible more precise changes in the steric or electronic properties of an amino acid and expands the scope of structural perturbations that can be made (12,13). We have used molecular modeling and simulation techniques to design unnatural amino acids to probe the effects of hydrophobicity, packing, cavity formation, and side chain conformational entropy on protein stability. The effects of these substitution on the thermal stability of T4L were evaluated from the melting properties of the purified mutant proteins determined by circular dichroism (CD) spectroscopy.

Amino acids were substituted for $Leu^{133}$ in T4L, a residue that is buried in the hydrophobic core of T4L and that defines one face of a large cavity (Fig. 1A). Mutations $Leu^{133} \rightarrow$ Ala and $Leu^{133} \rightarrow$ Phe have been shown not to significantly perturb overall protein structure, simplifying the analysis of mutations at this site (3,11). Substitutions for $Leu^{133}$ were designed to: (i) bury more hydrophobic surface area, (ii) incrementally enlarge the cavity, (iii) alter the hydrophobicity of isosteric amino acid side chains (14,15), and (iv) alter side chain configurational entropy. Molecular dynamics (MD) calculations were carried out to evaluate the structural effects of these mutations on the surrounding protein (Table 1). Simulations of wild-type protein ($Leu^{133}$) gave low root-mean-square (rms) values for cavity residues, as expected. Simulation of the $Leu^{133} \rightarrow$Phe mutant, as a test of the ability of MD to detect unfavorable interactions, yielded an unfavorable $\chi_1$ dihedral of -91°, close

to the Phe[133] crystal structure value of -87° (11), and a higher rms of Leu[121], which appears to have bad van der Waals contacts in the Phe[133] crystal structure (16).

Novel amino acid replacements for Leu were then designed and incorporated at position 133, and their influences on T4L stability were evaluated (scheme 1). S,S-2-amino-4-methylhexanoic acid **2** and S-2-amino-3-cyclopentylpropanoic acid **1** should extend the side chain of Leu by one $-CH_3$, and two $-CH_2-$ groups, respectively, and fill the cavity more completely than Leu (**3**) without creating unfavorable packing contacts (Fig. 1). For the cyclic side chain there should be less loss of conformational entropy on folding than with a straight chain alkyl group. The rms deviations of cavity residues surrounding Leu[133] $\rightarrow$ **1** and Leu[133] $\rightarrow$ **2** from MD simulations agree quite well with those of Leu[133]. Tert-leucine **7** buries approximately the same hydrophobic surface area as Leu but is expected to introduce significantly greater strain because of a clash between the $\gamma$ carbons and the carbonyl oxygen of Ala[129] in the preceding turn of the $\alpha$ helix. Rotation of the side chain of **7** about $\chi_1$ cannot alleviate this strain because it is symmetrically substituted, and so, to accommodate the steric bulk, $\phi$ increases in the simulation (-53° to -59° in the protein and -56° to -66° in the isolated helix). Norvaline **4**, ethylglycine **6**, and alanine **8** were included to assay the effects of systematic removal of $-CH_3$ ($-CH_2-$) groups from Leu[133]. Finally, O-methylserine **5**, which is isosteric with norvaline **4** but has a $CH_2 \rightarrow O$ substitution, was included to assess the effects of side chain solvation (polar and nonpolar) on protein stability. The simulations suggest that the norvaline **4**, methylserine **5**, and ethylglycine **6** side chains have similar conformations to Leu133, except for a ~7° change in $\psi$, which probably related to the smaller size of these residues and reflects slight repacking of the cavity (1,3). In addition, **5** has a smaller $\chi_1$ dihedral angle (-61°) than Leu (-78°), which may be due to decreased bulk at the $\gamma$ position and perhaps some electrostatic effects. Although the changes in dihedral angles and rms deviations are small, they do permit qualitative estimates of structural disruptions.

Incorporation of the above amino acids at position 133 of T4L was accomplished by in vitro suppression of a $Leu^{133}$→TAG stop mutation (encoded on plasmid pT4LL133am) (17) with a chemically aminoacylated suppressor tRNA derived from yeast $tRNA^{Phe}$ (12, 13, 18). We have used this method to site-selectively incorporate the analogs into β-lactamase (12) and a photoactivatable β-2-nitrobenzyl-Asp (19) as well as a wide variety of amino acids with novel backbone structures into T4L (20). When the in vitro-coupled transcription-translation system was programmed with pT4LL133am and supplemented with aminoacyl $tRNA_{CUA}$'s, full-length T4L was produced (lanes 4 to 11, Fig. 2). At 4 mM added magnesium acetate, amino acids **1** to **8** were incorporated with suppression efficiencies ranging from 27% (**8**) to 62% (**2**). In contrast, when $tRNA_{CUA}$ was omitted (lane 2, Fig. 2) or did not carry an amino acid (lane 3, Fig. 2), less than 1% full-length T4L was produced, compared to expression of wild-type T4L by pHSe54,97.TA (lane 1, Fig. 2) (20, 21). Thus the in vitro system does not contain endogenous suppressor tRNAs capable of reading through the amber stop codon, nor do the aminoacyl tRNA synthetases present in the *Escherichia col* S-30 extract aminoacylate the suppressor $tRNA_{CUA}$ with any of the 20 natural amino acids.

Mutant T4L's were purified to homogeneity (as judged by silver-stained denaturing polyacrylamide gel electrophoresis) from 5.0-ml in vitro protein synthesis reactions (22). The mutants have specific activities indistinguishable from that of wild-type T4L with the exception of $Leu^{133}$→ **7**, whose activity is reduced fivefold (23). This decrease likely reflects the unfavorable steric interactions that result from introduction of a tertiary β center within an α helix, leading to perturbation of the three-dimensional protein structure (24) (Table 1). The thermal stabilities of the mutants were evaluated with purified protein (extensively dialyzed against 20 mM potassium phosphate, 25 mM KCl, pH 2.5) by

determining the midpoint of the reversible thermal denaturation curve as monitored by CD with the use of a two-state denaturation model (25) (Fig. 3 and Table 2).

Semiquantitative estimates of stability differences based on hydrophobicity, packing effects, and side chain conformational entropy were also made for a comparison with the experimental values (Table 2). Transfer free energies, $\Delta G$'s of hydrocarbons between hydrocarbon liquids and water, were used to estimate the maximum $\Delta\Delta G$ of desolvating nonpolar amino acid side chains upon folding (26, 27). Two values were used: 24 cal $mol^{-1}$ $Å^{-2}$ of solvent-accessible surface area removed from water (28) and the more recent value of 47 cal $mol^{-1}$ $Å^{-2}$, which reflects an additional entropy term due to size differences between solvent and solute (26). The difference in packing interactions between the liquid-like denatured and solid-like folded states was estimated with the value of the heats of fusion for alkanes derived by Nicholls et al. of 20.3 cal $mol^{-1}$ $Å^{-2}$ (multiplying this value by the difference in surface area compared to Leu gives an upper estimate of the stabilization or destabilization due to changes in packing) (27). Rather than use the entropy of melting to estimate the contribution of freezing side chain motion, side-chain configurational entropy $\Delta S_{conf}$ was calculated from the partition functions of the side chain dihedrals, giving entropy terms that are a function of the accessible conformations of the side chains rather than a function of their surface area. In general, the packing and $\Delta S_{conf}$ terms nearly cancel one another (27). Cavity formation or filling must be accounted for when comparing mutations of hydrophobic residues to transfer experiments (1, 3). The cost of cavity formation has been estimated from sublimation $\Delta G$'s to be ~20.7 cal $mol^{-1}$ $Å^{-2}$ (27). Although this number gives a maximum value because it assumes that the entire change in surface area contributes either to filling a cavity or creating one, the value is very close to that determined experimentally (20 cal $mol^{-1}$ $Å^{-2}$) by Eriksson *et al.* (3), where repacking is taken into consideration. In vitro synthesized T4L containing Leu at position 133 (generated with leucyl $tRNA_{CUA}$) melted at 43.51° +/- 0.19° C, in good agreement with our

previous determinations under identical conditions (20). Stepwise removal of $-CH_3$ ($-CH_2-$) groups led to a stepwise decrease in protein stability (4, 5, 7), but the change was nonlinear; $\Delta\Delta G$ became increasingly large with decreasing side chain surface area ($\Delta\Delta G$ = 1.1 kcal $mol^{-1}$ for **3** versus **4** and $\Delta\Delta G$ = 2.2 kcal $mol^{-1}$ for **4** versus **6**). This trend is accurately reflected n the semiquantitative estimates of the $\Delta\Delta G$'s and results from the net effect of changes in the hydrophobic, entropic, packing, and cavity terms (29).

Unnatural amino acid mutagenesis allows us to directly correlate the transfer of amino acid side chains from aqueous solution to the protein hydrophobic core with transfer from water to octanol in the absence of other variables (30). The difference in stability between the mutant enzyme containing **4** and the mutant enzyme containing the isosteric residue **5** ($-CH_2-$ replace by -O-), 1.7 kcal $mol^{-1}$, compares favorable with the difference in the octanol-water partitioning ratios of the respective N-acetyl amide derivatives ($\Delta\Delta G$ = 1.8 kcal $mol^{-1}$) (31), suggesting in this case that octanol-water partitioning ratios provide an accurate measure of solvation effects without additional entropy terms (26).

The extended amino acids **2** and **1**, which were designed to fill the cavity with minimal strain stabilize T4L by 1.9° (0.60 kcal $mol^{-1}$) and 4.3°C (1.24 kcal $mol^{-1}$), respectively, demonstrating that amino acids that increase the bulk of buried hydrophobic residues without concomitant strain can significantly increase protein stability (4, 7, 9). The experimental difference between the stabilities of the mutant proteins containing these two amino acids at position 133 (0.54 kcal $mol^{-1}$) in part reflects the effects of increasing $\Delta S_{conf}$ while keeping surface area relatively constant ($\Delta$surface area = 0.9 $Å^2$). Again, this value is in reasonable agreement with the calculated $\Delta S_{conf}$ and emphasizes the importance of this term in interpreting mutagenesis results. The $Leu^{133} \rightarrow$ **1** and $Leu^{133} \rightarrow$ **2** mutants are likely not as stable as the calculated estimates because of nonideal packing interactions or incomplete filling of the cavity.

Hydrophobicity, packing effects, cavity formation, and side chain conformational entropy all play important roles in determining protein stability and must be considered in any mutational study. Additional mutations coupled with intensive free energy calculations should provide greater insight into the various contributions of these factors to protein folding and stability. This approach can be applied to other forces that determine protein folding, including hydrogen bonding, electrostatic interactions, and main chain effects.

## REFERENCES AND NOTES


1. K.A. Dill, *Biochemistry* **29**, 7133 (1990).
2. M.J. Behe, E.E. Lattman, G.D. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4195 (1991).
3. A.E. Eriksson *et al.*, *Science* **255**, 178 (1992).
4. M. Matsumura, W.J. Becktel, B.W. Matthews, *Nature* **334**, 406 (1988).
5. J.T. Kellis, Jr., K. Nyberg, D. Sali., A.R. Fersht, *ibid.* **333**, 784 (1988).
6. W.S. Sandberg and T.C. Terwilliger, *Science* **245**, 54 (1989).
7. K. Yutani, K. Ogasahara, T. Tsujita, Y. Sugino, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4441 (1987).
8. D. Shortle, W.E. Stites, A.K. Meeker, *Biochemistry* **29**, 8033 (1990).
9. W.A. Lim and R.T. Sauer, *J. Mol. Biol.* **219**, 359 (1991).
10. _____, *Nature* **339**, 31 (1989).
11. M. Karpusas, W.A. Baase, M. Matsumura, B.W. Matthews, *Proc. Natl. Acad. Sci. U.A.A.* **86**, 8237 (1989)
12. C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P.G. Schultz, *Science* **224**, 182 (1989).
13. J.A. Ellman, D. Mendel, C.J. Noren, S. Anthony-Cahill, P.G. Schultz, *Methods Enzymol.* **202**, 301 (1991).
14. T.L. Doering *et al.*, *Science* **252**, 1851 (1991).
15. R.O. Heuckroth, L. Glaser, J.I. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8795 (1988).
16. Phe adopted a low-enery dihedral angle ($\chi_1$ = -56°) in a minimized isolated helix ($Trp^{129}$ to $Ala^{134}$), suggesting that the unfavorable angle in the native protein is caused by unfavorable steric interactions with the carbonyl O of the $Asn^{132}$ side chain and is thus unlikely to move to accommodate $Phe^{133}$.

17. Mutagenesis was performed according to the Eckstein method with 5' TCT ACT TTT AGC CTA GTT AAC TGC 5' (mismatches underlined) as the mutagenic oligonucleotide; J.R. Sayer, W. Schmidt, F. Eckstein, *Nucleic Acids Res.* **16**, 791 (1988).

18. C.J. Noren *et al., ibid.* **18**, 83 (1990).

19. D. Mendel, J.A. Ellman, P.G. Schultz, *J. Am. Chem. Soc.* **113**, 2758 (1991).

20. J.A. Ellman, D. Mendel, P.G. Schultz, *Science* **255**, 197 (1992).

21. Plasmid pHSe54,97TA encodes a cysteine-free T4L behind a twin *tac* promoter.

22. Proteins werer purified by polyethyleneimine·HCl precipitation of the crude in vitro supression reaction followed by passing the supernatant over DEAE- and CM-cellulose cartridges in tandem. T4L's were eluted from the CM-cellulose cartridge with a salt gradent (pH 7.5). A second round of cation exchange chromatography at pH 4.9 aforded homogeneous protein: See also (20).

23. Suppression efficiencies, determined by scintillation counting of SDS-polyacrylamide gel slices, were used to estimate the amount of T4L produced. These values were then compared to the rates of lysis of *E. coli.* strain NAPIV cells by the mutant proteins to estimate the specific activity relative to wild-type enzyme. See also (19) and (20).

24. P.C. Lyu, J.C. Sherman A. Chen, N.R. Kallenbach, *Proc. Natl. Acad. Sci., U.S.A.* **88**, 5317 (1991).

25. W.J. Becktel and W.A. Baase, *Biopolymers* **26**, 619 (1987).

26. K.A. Sharp, A. Nicholls, R. Freidman, B. Honig, *Biochemistry* **30**, 9686 (1991); K. A. Sharp, A. Nicholls, R.F. Fine, B. Honig, *Science* **252**, 106 (1991).

27. A. Nicholls, K.A. Sharp, B. Honig, *Proteins* **11**, 281 (1991).

28. C. Chothia, *J. Mol. Biiol.* **105**, 1 (1976).

29. Cavity terms can also be estimated from cavity surface area (3). Changes in cavity surface areas using a 1.4 Å probe are: $Leu^{133} \rightarrow$ **1**, -7.7 $Å^2$; $Leu^{133} \rightarrow$ **2**, -14.6 $Å^2$;

Leu[133] → **4**, -16.4 Å$^2$; Leu[133] → **6**, -66.1 Å$^2$. Estimates using these values give cavity terms similar to those reported in Table 2.

30. The only isosteric replacement provided by the natural amino acids is Val to Thr; however, this would result in the introduction of an unsatisfied hydrogen bond.
31. Octanol-water partition values were determined according to the protocol of J. Fauchere and V. Pliska [*Eur. J. Med. Chem.* **18**, 369 (1993)] by vapor phase chromatography. The ΔΔG value of 1.8 kcal mol$^{-1}$ for **4** versus **5** represents the average of five determinations.
32. C. Schafmeister, LEaP, University of California, San Francisco (1991).
33. C. Huang *et al.*, MIDAS PLUS, Computer Graphics Lab, University of California, San Francisco (1991).
34. D.A. Evans, T.C. britton J.A. Ellman, R.Ll Dorow, *J. Am. Chem. Soc.* **112**, 4011 (1990).
35. R. Naef and D. Seebach, *Helv. Chim. Acta* **68**, 135 (1985).
36. G. Seibel, H.C. Singh, P.K. Weiner, J. Caldwell, P.A. Kollman, AMBER 3.0, Rev. A, University of California, San Francisco (1991).; S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A. Case, *J. Comput. Chem.* **7**, 230 (1986).
37. Temperature was measured as described (20).
38. W.J. Becktel and J.A. Schellman, *Biopolymers* **26**, 1859 (1987)
39. S. Dao-pin, D.E. Anderson, W.A. Baase, F.W. Dahlquist, B.W. Matthews, *Biochemistry* **30**, 11521 (1991).
40. T.J. Richmod, *J. Mol. Biol.* **178**, 63 (1984).
41. F.M. Richards, *Annu. Rev. Biophys. Bioeng.* **6**, 151 (1977).
42. We are grateful for support by the Director, Office of Energy Research, Office of Basic Energy Sciences, Division of Material Sciences, and also by the Division of Energy Biosciences of the U.S. Department of Energy (DE-AC03-76SF00098). D.M. was supported by American Cancer Society postdoctoral fellowship PF-4014A and J.A.E.

by NSF Waterman Awardee. D.L.V. was supported by NIH through a Biotechnology Training Grant (GM-08388-02), and P.A.K. thanks NIH (GM-29072), for research support. We also thank the NIH for support of the UCSF Computer Graphics Laboratory, R. Langridge, P.I. (NIH-RR-1081).

## Figure Captions

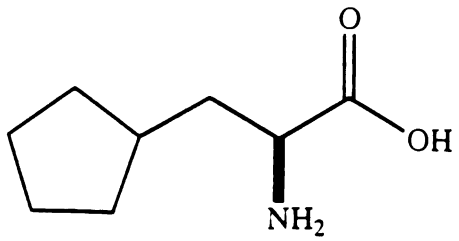
**Scheme 1.** Amino acid substitutions for $Leu^{133}$; Me, methyl.

**Figure 1.** Graphic representation of the side chain van der Waals surfaces at position 133 in T4L of: (**A**) Leu, (**B**) S,S-2amino-4-methylhexanoic acid **2**, and (**C**) S-2-amino-3-cyclopentyl-propanoic acid **1**. Amino acids were designed with the model building program LEaP starting from the coordinates of $Leu^{133}$ and by using standard geometries (32). The residues were then superimposed on the crystal structure. Visualization of the cavity was made possible by the generation of a molecular surface in the absence of a side chain at position 133 with MIDAS PLUS and represents the contact area between the protein and a water probe of 1.4 Å radius (33). Unnatural amino acid **1** was synthesized using the method of Evans (34) starting from 3-cyclopentylpropionyl chloride. Amino acid **2** was constructed according to the method Seebach (35) with S-2-methyl-1-butanol.

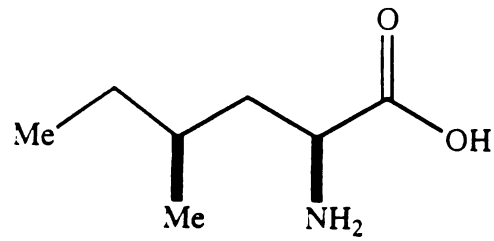
**Figure 2.** Autoradiogram of in vitro suppression reactions labeled with L-[$^{35}$S]Met and containing the following plasmids and tRNAs: lane 1, pHSe54,97.TA (wild type); lane 2, pT4LL133am without $tRNA_{CUA}$; lane 3, pT4LL133am and 8 μg of full-length, unacylated $tRNA_{CUA}$; lane 4, pT4LL133am and 8 μg of alanyl-$tRNA_{CUA}$; lane 5, pT4LL133am and 8 μg of ethylglycyl- $tRNA_{CUA}$; lane 6, pT4LL133am and 8 μg of norvalyl-$tRNA_{CUA}$; lane 7, pT4LL133am and 8 μg of O-methylseryl-$tRNA_{CUA}$; lane 8, pT4LL133am and 8 μg of leucyl-$tRNA_{CUA}$; lane 9, pT4LL133am and 8 μg of S,S-2-amino-4methylhexanoyl-$tRNA_{CUA}$; lane 10, pT4LL133am and 8 μg of S-2-amino-3-cyclopentylpropanoyl-$tRNA_{CUA}$; lane 11, pT4LL133am and 8 μg of *tert*-leucyl-$tRNA_{CUA}$. Lane M contains $^{14}$C-methylated molecular weight standards. Cleared supernatants (20μl) from terminated 30-μl in vitro reactions were incubated with 4 μl of 2.5 mg $ml^{-1}$ ribonuclease (Rnase) A for 15 min at

37°C and analyzed by 15% SDS-polyacrylamide gel electrophoresis. Each 30-µl reaction contained 7 µCi of L-[$^{35}$S]Met.

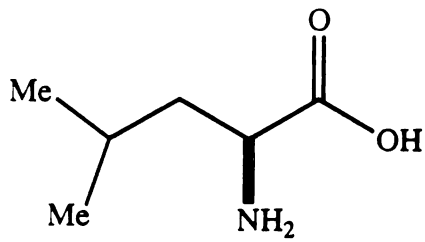
**Figure 3.** Melting curves for T4L variants containing the following amino acids at position 133: (○), O-methyl serine **5**; (▲), norvaline **4**; (□), leucine **3**; (×), S,S-2amino-4-methyl hexanoic acid **2**; and (●), S-2-amino-3-cyclopentylpropanoic acid **1**. Curves represent the average of three runs and were displayed with the relation $F_u = (CD_f - CD_x)/(CD_f - CD_u)$ where $F_u$ is the fraction of unfolded protein, $CD_f$ is the average CD signal (223 nm) at the earliest portion of the curve, $CD_u$ is the average CD signal at the latest part of the curve, and $CD_x$ is the average CD signal observed at given points between these two extremes.



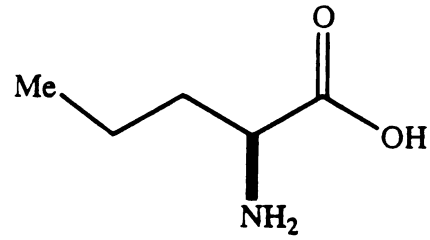

1



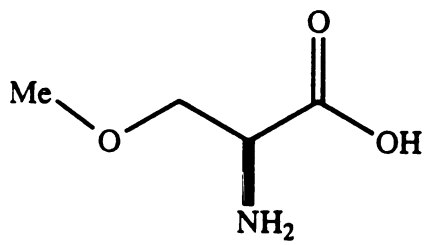

2



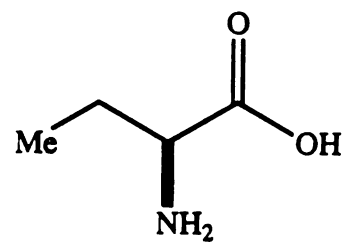

3



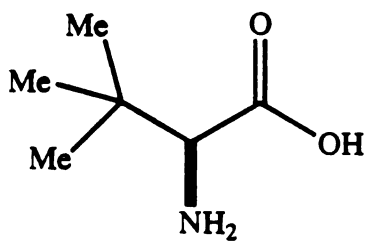

4



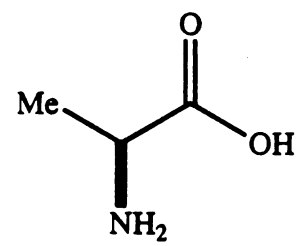

5




6




7




8

**Scheme 1**

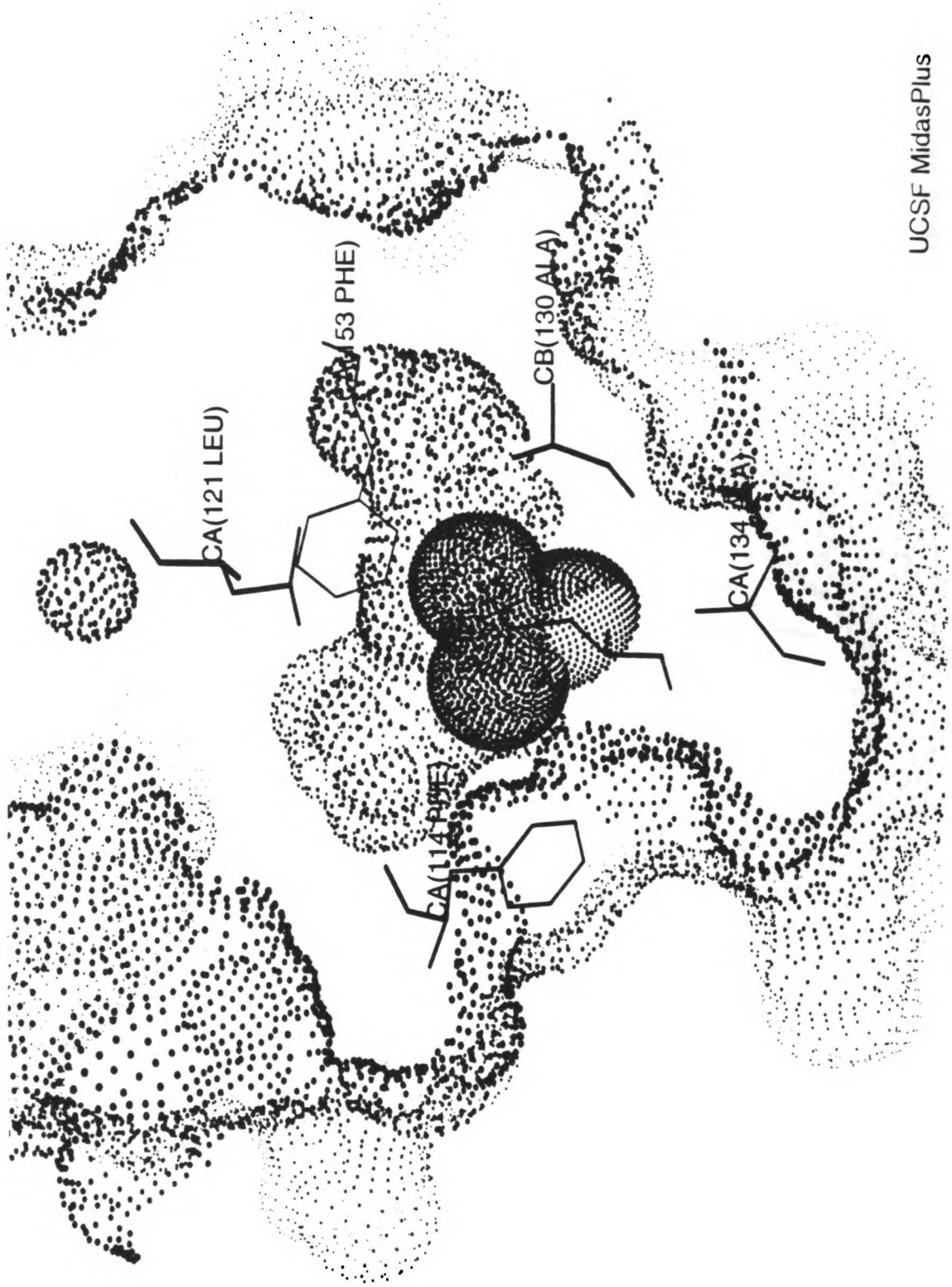



**Figure 1a**

**Figure 1b**

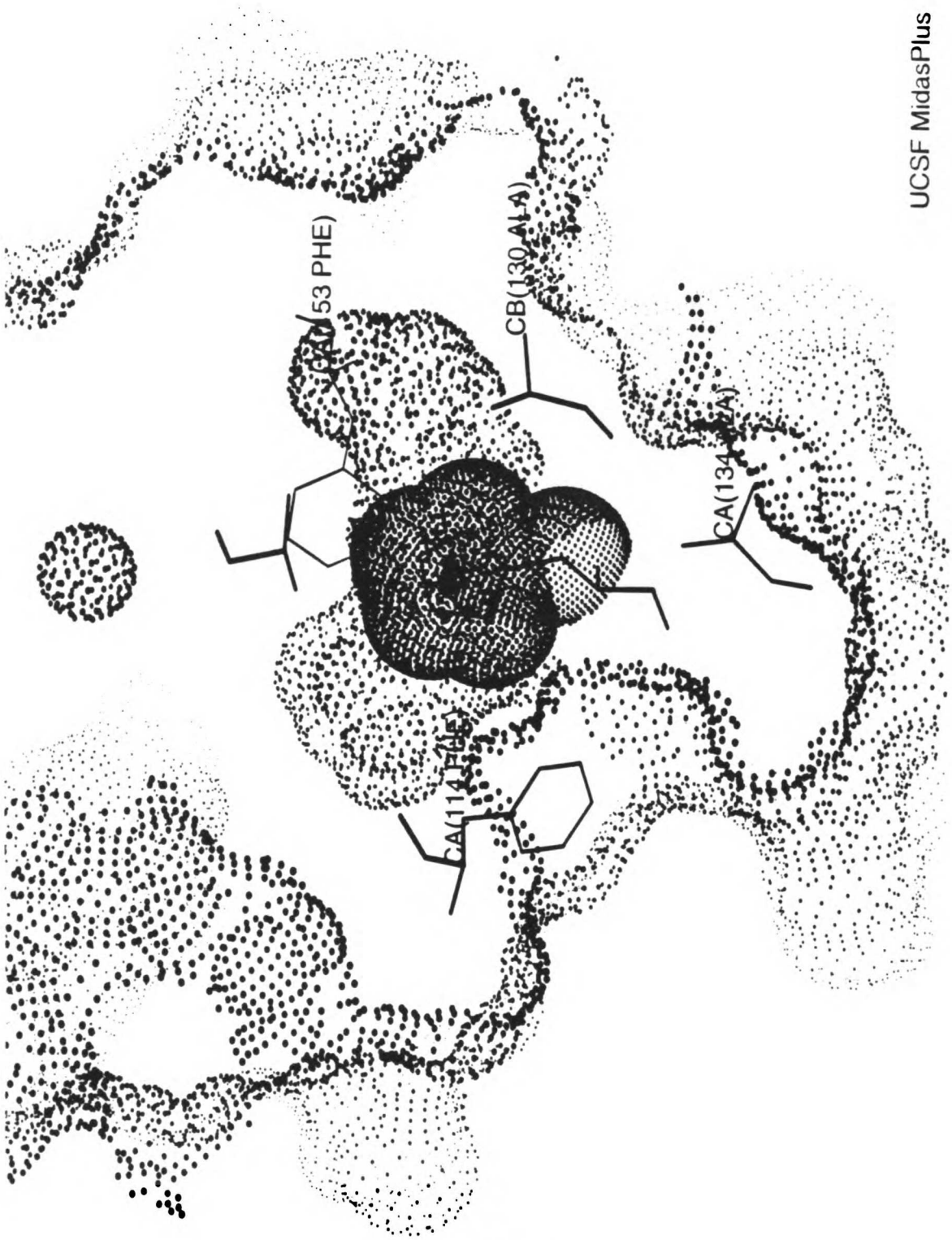



**Figure 1c**

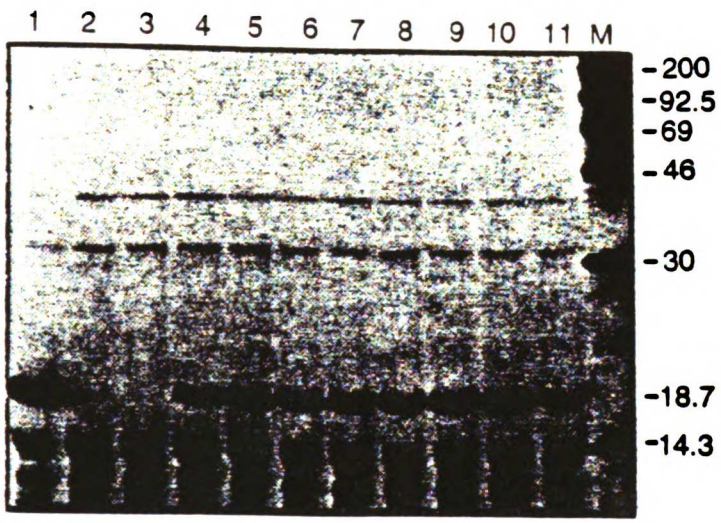





**Figure 2**

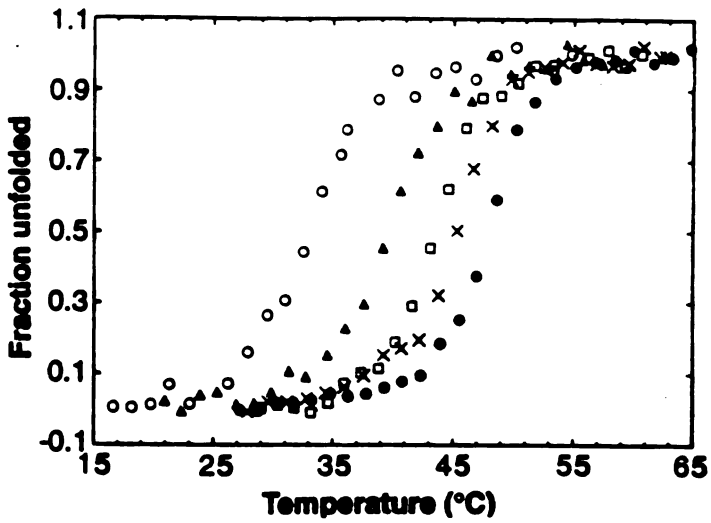

Fraction unfolded
1.1
0.9
0.7
0.5
0.3
0.1
-0.1
15
25
35
45
55
65
Temperature (°C)

**Figure 3**

**Table 1.** Root-mean-square deviations and dihedral angles averaged over 20 ps of molecular dynamics. A 17 Å cap of water molecules was centered on the γ carbon of 133. Only residues defining the cavity (98,99,102,106,111,114,117,118,120,121,129 to 134, 138, 139, 146, 149, 150, and 153) and waters within the cap were allowed to move (11) (an assumption supported by the x-ray crystal structures of the $Phe^{133}$ and $Ala^{133}$ mutants). Data were collected after minimization and 10-ps constrained molecular dynamics. Simulations were carried out with AMBER 3A with the Weiner *et al.* force field (36). All of the bond lengths were held constant. A time step of 2.0 fs and a constant dielectric were used. The rms deviations from the native crystal structure for heavy atoms only were calculated from coordinates collected every 0.5 ps and averaged.

| Residue 133 | Rms deviations of cavity residues near 133 | | | | | | Dihedral angles (degrees) | | |
|---|---|---|---|---|---|---|---|---|---|
| | 102 | 114 | 117 | 121 | 129 | 153 | φ | ψ | $\chi_1$ |
| 3 (wt) | 0.16 | 0.16 | 0.18 | 0.10 | 0.09 | 0.13 | -53 | -40 | -78 |
| Phe | 0.18 | 0.14 | 0.21 | 0.16 | 0.10 | 0.13 | -50 | -42 | -91 |
| 2 | 0.15 | 0.11 | 0.23 | 0.11 | 0.08 | 0.12 | -53 | -40 | -79 |
| 4 | 0.15 | 0.11 | 0.19 | 0.10 | 0.09 | 0.13 | -55 | -32 | -76 |
| 1 | 0.17 | 0.13 | 0.19 | 0.12 | 0.08 | 0.14 | -54 | -40 | -78 |
| 7 | 0.16 | 0.14 | 0.20 | 0.11 | 0.13 | 0.11 | -59 | -39 | -65 |
| 6 | 0.17 | 0.12 | 0.18 | 0.12 | 0.09 | 0.12 | -54 | -35 | -73 |
| 5 | 0.16 | 0.13 | 0.20 | 0.11 | 0.09 | 0.13 | -56 | -34 | -61 |

**Table 2.** Amino acids incorporated at position 133 of T4L and the experimental and estimated thermodynamic stabilities of the resulting enzymes. Proteins containing these replacements were purified to homogeneity from 5.0-ml scale in vitro suppression reactions (22). The CD melts were performed in triplicate (except for $Leu^{133} \rightarrow$ **6**, one determination) according to the procedure of Becktel and Baase (25) with a Jasco 600 polarimeter for CD readings (37, 38). Protein concentrations used in the CD measurements were between 5 and 10 $\mu g\ ml^{-1}$.

| Amino Acid | $T_m$* (°C) | $\Delta H_{expt}$† (kcal $mol^{-1}$) | $\Delta\Delta G_{expt}$‡§ (kcal $mol^{-1}$) | ΔSurface area ‖ (Å$^2$) | $\Delta\Delta G_{calc}$‡ (kcal $mol^{-1}$) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HE¶ | Cavity# | Pack** | $\Delta S_{conf}$†† | Total |
| **1** | 47.8 ± 0.3 | 98 ± 2 | 1.24 ± 0.09 | +22.0 | +0.5/+1.0 | +0.5 | +0.4 | 0.0 | +1.4/+1.9 |
| **2** | 45.4 ± 0.2 | 103 ± 16 | 0.60 ± 0.13 | +22.9 | +0.6/+1.1 | +0.5 | +0.5 | -0.5 | +1.1/+1.6 |
| **3** | 43.5 ± 0.2 | 96 ± 4 | 0.00 | 0.0 | 0.0/0.0 | 0.0 | 0.0 | 0.0 | 0.0/0.0 |
| **4** | 39.5 ± 0.3 | 79 ± 4 | -1.07 ± 0.10 | -21.0 | -0.5/-1.0 | -0.4 | -0.4 | +0.4 | -0.9/-1.4 |
| **5** | 32.9 ± 0.5 | 63 ± 13 | -2.63 ± 0.56 | -50.7 | -2.3 | -0.4 | -0.4 | +0.4 | -2.7§§ |
| **6** | ~31 | | ~-3.3‡‡ | -49.3 | -1.2/-2.3 | -1.0 | -1.0 | +0.5 | -2.7/-3.8 |
| **7** | | | | -4.6 | -0.1/-0.2 | -0.1 | -0.1 | +0.9 | +0.6/+0.5 |

*The melting temperatures ($T_m$) of the enzymes correspond to the average of three determinations. The error estimates correspond to the 95% confidence limit of these measurements. †The ΔH is the enthalpy of unfolding at the $T_m$ as determined by van't Hoff analysis of the CD melts. The error estimates correspond to the 95% confidence limit from three determinations. ‡Relative to the unfolded state so that a mutant with a positive ΔΔG is more stable than the wild-type Leu. §Isothermal ΔΔG values were calculated at 43.5°C with the use of a thermodynamic model in which a constant change in heat capacity $\Delta C_p$ was used, estimated (3) to be 2.5 kcal $mol^{-1}$ $K^{-1}$. ΔΔG values

were also calculated at the $T_m$ of the mutants by using $\Delta H$ (96 kcal $mol^{-1}$) and $\Delta C_p$ (1.80 kcal $mol^{-1}$) of the wild-type enzyme according to the method of Dao-pin *et al.* (39) and were within the error estimates obtained for the isothermal $\Delta\Delta G$ determinations. ||Nonpolar solvent-accessible surface area only. The difference in nonpolar solvent-accessible surface area compared to Leu was calculated with Richmond's algorithm (40), using a 1.4 Å radius water probe and Richards' van der Waals values (41) for the tetrapeptide N-acetyl-Asn-X-Ala-N-methyl amide in the extended conformation. ¶For the pairs of values, the hydrophobic effect (HE) is calcualted with a value of 24 cal $mol^{-1}$ $Å^{-2}$ of buried nonpolar surface area (28) for the first entry and with a value of 47 cal $mol^{-1}$ $Å^{-2}$ (26) for the second. #Estimated from $\Delta G$ of sublimation of hydrocarbons, 20.7 cal $mol^{-1}$ $Å^{-2}$ (27). This term accounts for changes in dispersion interactions as well as changes in the mobility of surrounding residues. **Estimated from $\Delta H$ of melting of hydrocarbons, 20.3 cal $mol^{-1}$ $Å^{-2}$ (27), which accounts for more favorable packing interactions in the solid-like native state versus the liquid-like denatured state. ††Estimated from $\Delta G = -RT\ln(q_2/q_1)$ where q is the partition function for side-chain rotamers and R is the gas constant; q was calcualted based on a *gauche* versus *trans* difference of 0.9 kcal $mol^{-1}$ and was assumed to be 1 for the native state, except for **4**, which can assume two conformations. ‡‡$\Delta\Delta G$ was calculated at the mutant $T_m$ (31°C) with $\Delta G$ (96 kcal $mol^{-1}$) and $\Delta C_p$ (1.80 kcal $mol^{-1}$) of the wild-type enzyme according to (39). §§Cavity, dispersion, and conformational entropy estimates are assumed to be the same as for norvaline (**4**).

# Chapter 4

## Free Energy Calculations of Protein Stability

# Modeling Protein Stability: A Theoretical Analysis of the Stability of T4 Lysozyme Mutants

*David L. Veenstra and Peter A. Kollman**

*University of California San Francisco*
*Department of Pharmaceutical Chemistry*
*San Francisco, CA 94143-0446*

Keywords: T4 lysozyme/free energy calculations/protein stability

Running Tittle: Calculating protein stability: T4 lysozyme

*To whom correspondence should be addressed

## Abstract

Free energy calculations were conducted to determine the relative stability of the unnatural amino acid mutants norvaline and methyl-serine and of alanine at site 133 in T4 lysozyme. These calculations were performed to both assess the validity of the methodology and to gain a better understanding of the forces which contribute to protein stability. Different length peptides were used to model the denatured state. Restraints were employed to force sampling of the side chain $\chi_1$ dihedral of the perturbed side chain, and the effect of backbone rearrangement upon mutation was studied through the use of different constraint sets. In addition, the convergence behavior and hysteresis of the simulations in the folded and unfolded states were determined. We find that free energy calculations can provide useful insights to protein stability when conducted carefully on a well chosen system. Our results suggest that loss of packing interactions in the native state is a major source of destabilization for mutants which decrease the amount of buried nonpolar surface area and that subtle responses of the backbone affect the magnitude of the loss of stability. We show that the conformational freedom of the $\chi_1$ dihedral has a noticeable effect on protein stability and that the solvation of amino acid side chains is strongly influenced by interactions with the peptide backbone.

## Introduction

The use of site-directed mutagenesis in conjunction with X-ray crystallography and thermodynamic measurements has been one of the most successful methods for examining the stability of proteins (Feire, 1993; Matthews ,1993). A classic example is the enzyme T4 lysozyme, one of the most studied proteins in the field of structural biology. A stunning array of structural and thermodynamic data has been generated; there are over 150 crystal structures of mutants available in the protein data bank (Matthews, 1995). T4 lysozyme is a single polypeptide chain of 164 residues which folds into two distinct domains and contains no disulfide bridges. The mutations studied have probed, for example, the hydrophobic effect and packing in the protein interior (Eriksson, 1992,1993; Baldwin, 1993), helix-dipole interactions (Nicholson, 1991), hydrogen bonding (Blaber, 1993), and insertions within helices (Heinz, 1994).

One of the most interesting sites of mutation is Leu 133, which is located in the C-terminal core in a four-helix bundle. Its side chain faces two hydrophobic cavities 39 $Å^3$ and 23 $Å^3$ in size. Thus, any changes in the physical properties of the side chain will have significant impact on the stability of the protein and both larger and smaller residues can be explored. Karpusas *et al.*(1989), in an effort to increase the stability of T4 lysozyme, substituted phenylalanine for leucine at position 133, but the protein was less stable due to poor van der Waals (vdW) contacts. In contrast, the L133A mutant is 3.5 kcal/mol less stable due to the loss of buried nonpolar surface area (Eriksson *et al.*, 1992). Mendel *et al.* (1992) utilized unnatural amino acids at position 133 to successfully stabilize the protein by filling the hydrophobic cavities. Side chains were also designed which progressively decreased the size of the side chain and thus the stability of the protein. In addition, an isosteric substitution was employed to study desolvation effects. Unnatural amino acids offer the advantage of more specifically

modifying the amino acid side chain, thus simplifying the task of deconvoluting the contributions to protein stability. However, correlations between estimated contributions to protein stability and experimental results are still empirical in nature. Further analysis of the intermolecular interactions involved in forming the native state of a protein can be accomplished with more detailed theoretical models.

Molecular dynamics (MD) and free energy calculations have proven to be valuable in analyzing the results of site-directed mutagenesis studies of protein stability as well as guiding experimental design (Heiner *et al.*, 1993; Cornish *et al.*, 1994). Indeed, MD was used to help in the design of the unnatural amino acids inserted into T4 lysozyme (Mendel *et al.*, 1992). Free energy calculations of protein stability have been conducted for a variety of substitutions. For example, Dang *et al.* (1989) investigated the role of hydrogen bonding with the mutation Thr → Val in T4 lysozyme. Tidor has studied electrostatic perturbations in both T4 lysozyme (Tidor and Karplus, 1991) and the λ Cro protein (Tidor, 1994), and the hydrophobic mutation Ile → Ala in barnase has been examined by Prevost *et al.* (1991) and more recently by Sun *et al.* (1996a).

Despite this variety of work, difficulties such as adequate conformational sampling, structural hysteresis, and modeling the non-specific structure of the denatured state have led some to question the worthiness of such efforts (Yun-yu *et al.*, 1993). In addition, the practice of decomposing free energy changes into individual contributions, ie. vdW versus electrostatics (Yun-yu *et al.*, 1993; Boresch *et al.*, 1994; Mark and van Gunsteren, 1994; Smith and van Gunsteren, 1994; Boresch and Karplus 1995), and certain free energy calculation methodologies have been debated (Sun *et al.*, 1996a). We thus felt that a thorough investigation of the techniques and difficulties presented by protein stability calculations would be valuable for methodology development and the evaluation of what insights can be gathered from free energy calculations.

We present here free energy calculations of the unnatural amino acid mutations norvaline (Nvl) → Ala and methyl-serine (Mse) → Nvl at site 133 in T4 lysozyme (Fig.1) (Mendel *et al.*, 1992). T4 lysozyme is an ideal candidate for theoretical study due to its relatively small size, abundance of structural and thermodynamic data, and general absence of significant structural rearrangement upon mutation. Both the L133F and L133A crystal structures show slight backbone movement and no changes in side chain rotamer states (Karpusas *et al.*, 1989; Eriksson *et al.*, 1992). The mutation Mse → Nvl involves essentially no changes in topology, thus simplifying differences in conformational preferences, and serves as a probe of desolvation properties of the side chain with minimal changes in packing in the folded protein. The mutation Nvl → Ala has significant effects in the unfolded state because of the hydrophobic effect as well as in the folded state due to changes in packing. It does not result in a change in β-branching, which would introduce significant contributions from side chain - helix interactions. In addition, mutations through Ala provide a pathway for changing the conformation of the Mse and Nvl side chains.

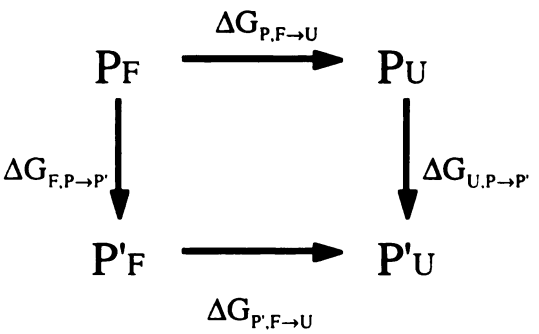
We have developed and tested force field parameters using the small molecules dimethyl ether (Dme) and propane (Prp) to model the Mse → Nvl perturbation. All perturbations were tested for convergence behavior, and several perturbations were conducted in the reverse direction to determine structural and thermodynamic hysteresis. Different length peptides were used as models of the denatured state to seek errors associated with the unfolded state calculations. Constraints were used on the protein backbone to assess the effect of structural relaxation on protein stability and the sensitivity of the results to the constraint set employed. Finally, computational techniques, as applied by Hermans *et al.* (Yun and Hermans, 1991; Hermans *et al.*, 1991, 1992; Wang *et al.*, 1995), were used to enhance sampling of the mutated side chain rotamers. This work

has produced consistent results which have given us greater understanding of the roles that packing, backbone rearrangement, side chain entropy, and solvation of sidechains play in influencing protein stability. While we do not suggest these calculations are straightforward or applicable in a wide array of settings, we show that free energy calculations can provide valuable insights which are difficult to gain experimentally.

## Methods

### *Calculating Protein Stability*

The difference in the stability of two protein sequences is not calculated directly since simulating the actual unfolding of a protein is not feasible due to the tremendous sampling that would be required. Instead, the difference is calculated indirectly using the thermodynamic cycle:




In this figure, P and P' refer to two different protein sequences and the subscripts F and U refer to the folded and unfolded states, respectively. Since free energy is a state function, these free energy differences are related through the equation

$$\Delta\Delta G = \Delta G_{P,F\rightarrow U} - \Delta G_{P',F\rightarrow U} = \Delta G_{F,P\rightarrow P'} - \Delta G_{U,P\rightarrow P'} \quad (1)$$

$\Delta G_{P,F\rightarrow U}$ and $\Delta G_{P',F\rightarrow U}$ are measured experimentally while $\Delta G_{F,P\rightarrow P'}$ and $\Delta G_{U,P\rightarrow P'}$ are calculated. $\Delta G_{F,P\rightarrow P'}$ was calculated using the using the X-ray crystallographic coordinates to model the folded state while $\Delta G_{U,P\rightarrow P'}$ was determined using model-built peptides in the extended conformation to model the denatured state.

### Thermodynamic Integration

Thermodynamic integration was used to calculate the difference in free energies between two amino acid residues. A coupling parameter $\lambda$ is used in the Hamiltonian $H(\lambda)$ to connect the two thermodynamic states, $H(\lambda=0)$ and $H(\lambda=1)$, which represent the initial and final states, respectively. The free energy change between these two states is given by

$$\Delta G = \int_0^1 d\lambda \left\langle dH(\lambda)/d\lambda \right\rangle_\lambda \approx \sum_i \left\langle dV(\lambda)/d\lambda \right\rangle_{\lambda_i} \Delta\lambda \qquad (2)$$

where $(\cdots)\lambda$ represents the ensemble average of $H(\lambda)$ (van Gunsteren and Weiner, 1989; Kollman, 1993). The transformation from one state to the other, e.g. Mse to Nvl, is achieved by changing the force field parameters which describe the amino acid residue, with the in-between states being described by a hybrid force field. All intra-residue contributions to the free energy were included. The single topology method was used (Pearlman, 1994; Sun *et al.*, 1996a). Bond lengths to dummy atoms were not shrunk, but a bond-PMF correction was applied due to the changing bond lengths inherent in the single topology method (Pearlman and Kollman, 1991; Wang and Hermans, 1994).

**Conformational sampling**

In order to enhance sampling of the $\chi_1$ side chain dihedral of Mse and Nvl, flatwell restraints were applied to keep $\chi_1$ in one of its three preferred rotamers, *gauche+* (*g+*), trans (*t*), and *gauche-* (*g-*) centered approximately at -60° (*g+*), 180° (*t*), and +60° (*g-*). Conformational space was then explored by mutating to Ala and regrowing the side chain in a different rotameric state, e.g.

$$\mathrm{Mse}_{\chi_1=g+} \rightarrow \mathrm{Nvl}_{\chi_1=g+} \rightarrow \mathrm{Ala} \rightarrow \mathrm{Nvl}_{\chi_1=t} \rightarrow \mathrm{Mse}_{\chi_1=t} \qquad (3)$$

In this manner the relative conformational free energies can be calculated. Sampling via use of a forcing potential was not employed since large barriers to rotation in the interior of a protein render such an approach less efficient. The flatwell restraints used

were taken from the SANDER module of AMBER 4.0 (Pearlman *et al.*, 1991) and implemented in the GIBBS routine, and have the form

$$\begin{array}{ll} r < r_1: & e = 2k_2(r_1 - r_2)(r - r_1) + k_2 (r_1 - r_2)^2 \\ r_1 \geq r < r_2: & e = k_2 (r - r_2)^2 \\ r_2 \geq r \leq r_3: & e = 0.0 \\ r_3 < r \leq r_4: & e = k_3 (r - r_3)^2 \\ r > r_4: & e = 2k_3 (r_4 - r_3)(r - r_4) + k_3 (r_4 - r_3)^2 \end{array} \qquad (4)$$

where r is the torsion angle, $r_1$, $r_2$, $r_3$, and $r_4$ define the placement of the restraints, and $k_2$ and $k_3$ are force constants. The restraints are applied so that $r_2$ and $r_3$ coincide with the barriers which define the rotamer. The restraint potential affects the dynamics of the system only at the occasionally sampled saddle-points and not the high probability, low energy regions of dihedral space and thus does not effect equilibrium properties, ie. the configurational free energy (Hermans *et al.*, 1992; Hodel *et al.*, 1993). For example, for the *t* rotamer centered at $\chi_1 = 180°$, the flatwell restraint parameters $r_1$, $r_2$, $r_3$, and $r_4$ were set to 110°, 125°, 235°, and 250°, and $k_2$ and $k_3$ were set to 50 kcal/mol/rad$^2$ (Fig. 2).

Imposing a conformational restraint changes the conformational free energy of the system. Thus, the free energy differences must be corrected for the altered distribution. This correction consists of adding the free energy cost of imposing the restraint at the beginning of the simulation and subtracting the cost of the restraint from the final state (Yun and Hermans, 1991; Hermans *et al.*, 1991, 1992; Wang *et al.*, 1995). The free energy cost of selecting a state j can be calculated if the relative conformational free energies, $\Delta G_i$, of the other conformational states are known. The partition function for the restrained state j is $q_{rst} = 1$ and for the unrestrained state is

$$q_{unr} = 1 + \sum_i \exp(-\Delta G_i / kT)] \qquad (5)$$

The free energy for imposing the restraint is then given by

$$\Delta G_r = -kT \ln[q_{rst} / q_{unr}] \quad (6)$$

If the selected state j is the lowest energy rotamer, the restraint free energy will tend to be small. Since the relative conformational free energies of $\chi_1$ have been determined from the perturbation cycle through Ala (eq. (3)), $\Delta G_r$ can easily be calculated. The overall free energy change for Mse → Nvl, for example, using $\chi_1 = g+$ as the reference state, is given by

$$\Delta G_{Mse \rightarrow Nvl,\ overall} = \Delta G_{r,Mse,\chi_1=g+} + \Delta G_{Mse \rightarrow Nvl,\ \chi_1=g+} - \Delta G_{r,Nvl,\ \chi_1=g+} \quad (7)$$

It is thus necessary to calculate the relative free energies of all rotamers for each dihedral of interest. The $\chi_2$ side chain dihedral was not sampled since doing so would require performing 9 calculations for each mutation and initial results indicated that $\chi_2 = t$ is the preferred rotamer.

The free energy change due to gain of conformational freedom about $\chi_1$ upon "moving" from the relatively restricted interior of the protein to the unfolded state (peptide) can be calculated since the partition functions for the unrestrained dihedrals are known for the unfolded state, $q_{unr,unf}$, and the folded state, $q_{unr,fold}$ from eq. (5). The free energy change is determined from

$$\Delta G_{conf} = -kT \ln[q_{unr,unf}/q_{unr,fold}] \quad (8)$$

and is most likely entropic in origin, although the relative contribution cannot be determined in this formalism (Hermans *et al.*, 1992).

## Parameters

A modified version of the Wiener *et al.* (1986) force field was used for all calculations. The potential function describing the interatomic interactions is given by

$$V = \sum K_r(r-r^{eq})^2 + \sum K_\theta(\theta - \theta^{eq})^2 + \sum V_n/2[1 + \cos(n\phi - \gamma)] +$$

$$\sum_{bonds} (A_{ij}/R_{ij}^{12} - B_{ij}/R_{ij}^{6}) + \sum_{angles} q_i q_j/\varepsilon R_{ij} + \sum_{dihedrals} (C_{ij}/R_{ij}^{12} - D_{ij}/R_{ij}^{10})$$

The first three terms represent bonds, angles, and dihedrals respectively while the last three terms describe non-bonded interactions: van der Waals, electrostatics, and hydrogen bonding. Modifications to the force field included using, for the perturbed residues, point charges fit to a 6-31G* basis set potential with the RESP methodology (Bayly *et al.*, 1993) and vdW parameters derived from liquid simulations in conjunction with *ab initio* calculations (Gough *et al.*, 1992; Veenstra *et al.*, 1992). This charge derivation protocol has been shown to produce more accurate free energies of solvation and conformational potentials (Cornell *et al.* 1993), and the hydrocarbon parameters have been favorably tested in free energy calculations (Gough *et al.*, 1992; Sun *et al.*, 1992). These methodologies for parameter development were those used in the development of the latest AMBER force field, but the final version of the Cornell *et al.* (1995) force field was not utilized since it had not been completed when calculations in this study were initiated.

*Small Molecules*

The small molecules dimethyl ether (Dme) and propane (Prp) were studied as a model for the Mse → Nvl. Atom centered point charges for dimethyl ether and propane were determined as described above (Fig. 3). The vdW parameters for the ether oxygen were the same as those used for the sugar ring oxygen in DNA in the Wiener *et al.* (1986) force field (Table I). The vdW parameters for propane are those derived from liquid simulations of hydrocarbons (Sun *et al.*, 1992). The hydrogens in dimethyl ether have the same vdW parameters as those in propane except r* has been reduced by 0.1 Å due to the electron withdrawing effect of the nearby oxygen (Veenstra *et al.*, 1992).

*Amino Acids*

The charges for Mse, Nvl, and Ala were derived analogously to those for dimethyl ether and propane. The potentials were calculated and charges fit with acetyl (Ace) and

N-methyl acetyl (Nme) blocking groups on the amino acid. The backbone charges of Mse, Nvl, and Ala were constrained to be equal (Fig. 3). The vdW parameters for the side chains were the same as those used for dimethyl ether and propane (Table I). Although the Weiner *et al.* (1986) force field contains parameters for alanine, we felt it was important that the parameters for all three perturbed side chains were derived in an analogous manner. The parameters for the non-perturbed residues were taken directly from the Weiner *et al.* (1986) force field.

**Simulations**

Simulations were carried out using the GIBBS and MINMD modules of either the AMBER 4.0 (Pearlman *et al.*, 1991) or AMBER 4.1 (Pearlman *et al.*, 1995) suite of molecular mechanics programs. A time step of 2.0 fs was used and all bonds were constrained via the SHAKE algorithm (Ryckaert *et al.*, 1977). The TIP3P water model was used (Jorgenson *et al.*, 1983). Simulations were conducted at 300 K and periodic box simulations were carried out under NTP conditions.

*Small molecules*

The simulations of dimethyl ether and propane in solution were conducted in a periodic box of water which extended at least 9 Å from the solute, containing approximately 240 solvent molecules. All structures were minimized then equilibrated for 50 ps before beginning the perturbations. Simulations were conducted for varying lengths of time to test for convergence. Analogous calculations were conducted *in vacuo*.

*Unfolded State*

The unfolded state of T4 lysozyme was modeled variously with a capped di-, tetra-, and hexapeptide in an extended conformation in a periodic box of water. The neighboring residues in the native sequence, Val-Asn-X-Ala-Lys, were used in the peptides. The peptides were restrained in a non-specific extended state using the flatwell restraints described above (Eq. (4), Fig. 2). $\phi$ was restrained to be in the range 0° to -180° using $r_1 = 170°$, $r_2 = 185°$, $r_3 = 355°$, and $r_4 = 370°$ and $k_2$ and $k_3$ equal to 50 kcal/mol/rad$^2$

while $\psi$ was restrained to be positive, 0° to +180°, using $r_1$ = -10°, $r_2$ = 5°, $r_3$ = 175°, and $r_4$ = 190° and $k_2$ and $k_3$ equal to 50 kcal/mol/rad$^2$. Restraints were not applied to the non-perturbed side chains. The periodic box of water extended 9 Å from the solutes. Minimization was followed by 50-100 ps of equilibration before free energy calculations were commenced. The $\chi_1$ side chain dihedral was sampled for Mse and Nvl as described above (Eq. (3)). The perturbations were carried out over a total time of 200 ps for the Mse → Nvl mutation and 400 ps for the Nvl → Ala mutation based on results of the dimethyl ether → propane simulations and tests of convergence for the dipeptide. The perturbation in the Ala→Nvl direction was conducted for the dipeptide to test hysteresis. Simulations without forced side chain sampling, ie. without side chain restraints, were conducted for Nvl→Ala and Mse→Nvl in the dipeptide. Calculations *in vacuo* were performed in a similar fashion.

*Folded State*

Coordinates for the folded state of T4 lysozyme were obtained from the Brookhaven Protein Data Bank, entry 3LZM (1.7 Å, pH 5.3) (Weaver and Matthews, 1987). The N-terminal domain consists of one α-helix and a four stranded antiparallel β-sheet, and is connected via a 21 residue α-helix to the C-terminal domain, which is essentially all helical and contains a four-helix bundle (Fig. 4). The 152 crystallographic waters were included in all calculations. Counterions were added to neutralize unpaired charged side chains. Chloride ions were placed near 4 arginine and 10 lysine side chains and sodium ions were matched with 3 aspartic acid and 1 glutamic acid side chains. Twenty-six charged side chains were treated as salt bridges, including Asp 70···His 31. The overall charge on the system was zero. A 17 Å sphere of water was placed centered on the β atom of residue 133. This cap consisted of about 190 waters and included within its radius 55 crystal waters and the amino acids which make up the hydrophobic cavity which residue 133 faces.

All atoms were allowed to move during the perturbation, but two different sets of light constraints (10 kcal/mol/Å$^2$) were placed on the backbone mainchain atoms (N,$C_\alpha$,C). One constraint set left the residues adjacent to residue 133 unconstrained ($\alpha$-constraints, Table II), while the other also left the residues lining the hydrophobic cavity unconstrained (cavity-constraints, Table II). These constraint sets were used to examine the effect of movement of the backbone atoms surrounding the hydrophobic cavity into which the side chain of residue 133 protrudes (Fig. 5). Both constraint sets were used for the Nvl→Ala mutation, while only the $\alpha$-constraints were used for the Mse→Nvl calculation since this mutation involves little change in side chain volume. The "belly" option in AMBER, which allows specified residues to be held immobile, was not used during the perturbations because of widely inconsistent results when used in conjunction with the bond-PMF correction.

The system was first minimized with 50 kcal/mol/Å$^2$ positional constraints on all the heavy atoms of all the amino acid residues, except residue 133, then with 10 kcal/mol/Å$^2$ constraints on the same atoms. This was followed by minimization without constraints, but using the belly option in AMBER to hold the non-cavity residues immobile. This approach was utilized to allow the residues surrounding site 133 to fully adapt to either the Nvl or Mse side chain from the native Leu crystal structure. The system was equilibrated at 300 K for 50 ps with the belly option. The belly option was then turned off, and positional constraints, using the equilibrated belly structure as a reference, were applied using either the $\alpha$-constraint or the cavity-constraint set discussed above (Table II), which was applied throughout the data collection portion of the simulation.

The Nvl → Ala calculations were run for 400 ps since tests of convergence and hysteresis for simulations of the folded state indicated that the free energy differences

were relatively converged in this time frame. Mse - > Nvl mutations were conducted over 200 ps for analogous reasons. The Nvl → Ala perturbation was run in the reverse direction with $\chi_1 = g+$ to determine the hysteresis present in the folded state simulations. Simulations without side chain restraints were also conducted for both mutations.

## Results and Discussion

### Precision and Accuracy of Calculations

*Convergence*

The convergence behavior of the mutations Dme → Prp, Mse → Nvl, and Nvl → Ala are shown in Figure 6. All the amino acid calculations were conducted with the $\chi_1 = g+$ rotamer since this is typically a preferred rotamer in α-helices in protein structures (McGregor, 1987). The Dme → Prp mutation provides the simplest test case since there are no heavy-atom dihedrals which would require sampling, the topology change is conservative, and electrostatic perturbations have been shown to converge more quickly that those involving vdW changes. The Dme → Prp calculation *in vacuo* is actually converged within several ps (Fig. 6a). This is not surprising considering that the ether oxygen of Dme has no non-bonded interactions with the rest of the molecule. The free energy change arises mainly from electrostatic interactions between the methyl hydrogens and the methylene hydrogens of Prp as well as bond, angle, and dihedral changes, which should be small considering the similarity of the molecules. In solution, the Dme → Prp perturbation requires 200 ps for convergence (Fig. 6b). The increase in required simulation time is due to the need for solvent reorganization, and is consistent with previous studies of convergence for small molecule calculations (Mitchell and McCammon, 1991; Pearlman and Kollman, 1991; Pearlman, 1994b)

The analogous mutation of the amino acid side chain, Mse → Nvl, was tested in the dipeptide Ace-X-Nme. In contrast to the Dme → Prp *in vacuo*, Mse → Nvl *in vacuo* gradually approaches convergence by 100 - 200ps, although the overall change in ΔG is small (Fig. 6c). The dipeptide has more extensive nonbonded interactions as well as dihedral angles which can be sampled. The dynamics of the peptides is discussed in more detail in the results of the unfolded state models, but in brief, $\chi_1$ is restrained to the *g*+ well, and the backbone remains in the $C_7^{eq}$ conformation, so this convergence time does not necessarily reflect extensive torsional sampling. The perturbation in solution is similar to Dme → Prp, but the results do not converge as cleanly (Fig. 6d). The range in ΔG from 200 - 500 ps is 0.33 kcal/mol, while for the small molecule calculation, it is 0.15 kcal/mol. Sampling of $\phi$,$\psi$ is the likely cause of the greater variability for these calculations; both the $C_7^{eq}$ and $P_{II}$ backbone conformations are explored. The Mse → Nvl mutation in the protein converges more quickly than in solution (Fig. 6e). Although proteins can present complex environments in which to sufficiently sample conformational space, in some instances relevant torsional space may be reduced due to the rigid structure of a protein (Mark *et al.*, 1994; Smith *et al.*, 1995). In the case of site 133 in T4 lysozyme, the residue is in an α-helix, and sampling about $\phi$ and $\psi$ is limited compared to solution; the standard deviation for $\phi$ and $\psi$ in solution was in the range 15° to 30°, while for the protein simulations, the values were typically near 10°. In addition, there are no significant structural rearrangements upon mutation at this site in the X-ray crystal structures or our simulations. There are no water molecules within the protein cavity, and, as will be discussed in more depth later, there are no exceptionally good hydrogen bonding partners for the ether oxygen of Mse. Since $\chi_1$ and $\chi_2$ are restrained to the *g*+ and *t* rotamers, respectively, this convergence time essentially reflects sampling of torsions within defined energy wells.

The mutation of Nvl to Ala *in vacuo* surprisingly required about 200 ps for convergence, although the absolute change in $\Delta G$ is only about 0.5 kcal/mol (Fig. 6f). The need for more sampling is likely due to poor coupling between changes in the vdW parameters of the nonpolar atoms and the dynamics of the rest of the structure. The perturbation in solution shows the most dramatic convergence behavior, with a change in free energy difference of about 1 kcal/mol between the 50 ps and the converged result at 400 - 500ps (Fig. 5g). This doubling of the sampling time required for consistent results is a consequence of the significant rearrangement of water that occurs upon changes in the size of a solute (Mitchell and McCammon, 1991; Pearlman and Kollman, 1991; Pearlman, 1994b). As seen with the Mse $\rightarrow$ Nvl mutation, the Nvl $\rightarrow$ Ala mutation in the protein shows shorter convergence time and a smaller range of $\Delta G$'s than in solution (Fig. 5h). Although the Nvl $\rightarrow$ Ala perturbation in the protein requires at least 200 ps for convergence compared to 100 ps for Mse $\rightarrow$ Nvl in the protein, the similarity of the convergence curves suggests that the ensemble of structures which must be sampled by the surrounding protein is similar for both these perturbations . Although convergence was not tested for the $\chi_1 = g-$ rotamer, this conformation may require more sampling due to unfavorable vdW contacts with residues in the previous turn of the $\alpha$-helix.

The Dme $\rightarrow$ Prp and Mse $\rightarrow$ Nvl simulations as short as 50 ps give results which are similar to the results from simulations which are four times as long. Thus, electrostatic perturbations may in some cases be conducted over simulation times as short as 50 ps with consistent results. The perturbation of Nvl to Ala, on the other hand, clearly indicates that longer simulation times are required, even *in vacuo*. The results found here for the protein simulations may not be applicable to mutations which involve greater structural rearrangement of the protein, larger perturbations of the side chain, or mutations at the protein surface. In other words, the experimental evidence which

shows only small changes in the protein structure in response to mutation at site 133 has allowed us to design a model for which we can sample to converged results. As noted by Mark *et al.* (1994), however, the benefit of using a fairly rigid model may be at the cost of ignoring low-frequency motions and the accompanying entropy contributions.

*Hysteresis*

A more definitive test of the reliability of free energy calculations is to run the calculation in both perturbation directions to determine the hysteresis (Mark *et al.*, 1994). Hysteresis can result from the lag of the Hamiltonian from the representative structures, insufficient sampling, or from structural hysteresis (Pearlman and Kollman, 1989). The latter probably presents the greatest difficulty in protein simulations (Dang *et al., 1989;* Yun-yu *et al.*, 1993). Most recent force fields and simulation protocols result in a protein structure which tends to drift from the X-ray crystal structure as the simulation progresses. When testing for hysteresis, it is often difficult to simulate the return to the starting state structure, especially if it involves specific interactions such as hydrogen bonding. Constraints are often used to retain the native structure, but this sometimes sacrifices proper sampling of the mutant state. A structural analysis of the native state is discussed more thoroughly in the folded state calculation results, but we present here the thermodynamic results of the perturbation cycle Nvl → Ala → Nvl in solution and in the protein to support the credibility of our results.

We chose to test hysteresis with the Nvl → Ala → Nvl perturbation for the dipeptide and in the protein because this mutation should present the greatest challenge due to the change in side chain volume. The results for Nvl → Ala → Nvl in solution are shown in Figure 7a. The perturbation was conducted for 300 ps in each direction since this simulation time was converged, although the final results are reported for 400 ps simulations. The overall hysteresis is 0.2 kcal/mol. The free energy is slightly more

positive in the Ala → Nvl direction in the middle of the perturbation, and probably indicates the repulsive nature of forming a larger nonpolar vdW surface in water. Hamiltonian lag may account for the less positive free energy at the end of the simulation (Ala → Nvl, $\lambda$ = 1). The mutation in the protein for Ala → Nvl was conducted for 400 ps. Of the two constraint sets used in the Nvl → Ala calculations (Table II), we used the cavity-constraints set, which allows for greater protein movement, and is thus a more stringent test case for hysteresis than the $\alpha$-constraint set. Figure 7b shows the protein results; the hysteresis is again about 0.2 kcal/mol. The free energy difference is less positive in the range $\lambda$ = 0.25 to 0.75 than in the Nvl → Ala mutation direction. This is in contrast to the solution simulation, and probably indicates the favorableness of growing a vdW volume into a preformed cavity as opposed to having to form a cavity in solution. The structural changes are fairly reversible and are presented later.

Our results for tests of convergence and for hysteresis indicate that the models employed in our simulations for the folded and unfolded states give reproducible results which have sufficiently sampled local conformational space. The range of converged free energies of about 0.2 kcal/mol for the Dme → Prp, Nvl → Ala, and Mse → Nvl mutations both in solution and in the protein and the hysterisis of 0.2 kcal/mol for Nvl → Ala → Nvl in solution and the protein suggest that, in general, the precision of our results is also about 0.2 kcal/mol.

*Dme → Prp solvation free energy*

The Dme → Prp mutation was conducted primarily to test the parameters used for the ether moiety. The solvation free energy for Prp versus Dme is 3.89 kcal/mol as measured experimentally (Hine and Mookerjee, 1975). The calculated result, obtained from the difference in $\Delta$G in solution and *in vacuo*, is 3.90 kcal/mol, in excellent agreement, thus supporting the use of 6-31G* basis set derived charges and vdW

parameters derived from liquid simulations. We also employed modified vdW parameters for the hydrogens in Dme based on gas phase *ab initio* calculations conducted in conjunction with liquid simulations (Veenstra *et al.*, 1992; Gough *et al.*, 1992). These hydrogens have r* reduced by 0.1 Å, which allows for a closer approach and stronger interaction of water with these atoms. Simulations with the standard vdW parameters on these hydrogens produced solvation free energies about 0.3 kcal/mol less favorable for Dme, providing support for the practice of reducing r* on nonpolar hydrogens adjacent to electronegative atoms. The only difference between our paramterers and those used in the Cornell *et al.* (1995) force field are the vdW parameters for the ether oxygen (r* = 1.6837, $\varepsilon$ = 0.1700 ), which when used give a relative solvation free energy for Dme and Prp of -3.62 kcal/mol, also in good agreement with experiment. The slightly smaller value may be due to the absence of a 10-12 vdW term.

## Unfolded State Calculations

*Nvl→ Ala*

The free energy changes for Nvl → Ala in the three peptide models, Ace-X-Nme (dipeptide), Ace-Asn-X-Ala-Nme (tetrapeptide), and Ace-Val-Asn-X-Ala-Lys-Nme (hexapeptide) in solution and *in vacuo* are shown in Table III. The results are quite similar for the solution calculations regardless of the peptide model used. The largest discrepancy is for the hexapeptide in the *t* rotamer and the other two peptides, -3.11 kcal/mol versus -2.25 and -2.36 kcal/mol. This difference is likely due to the fact that the hexapeptide adopts a different backbone conformation; $\phi$ and $\psi$ for the mutated residue are in the $C_7^{eq}$ conformation ($\phi$,$\psi$ near -80°,+70°) (Table IV), which results in a 1,7 intra-backbone hydrogen bond. The $C_7^{eq}$ conformation is unfavorable for the $\chi_1 = t$

rotamer because of repulsion from the backbone carbonyl. The di- and tetrapeptides, in contrast, sample both the $C_7^{eq}$ conformer and the $C_5$ conformer ($\phi,\psi$ near -160°,+170°).

The relative conformational free energies for the average of the three peptide models are shown in Figure 8a and Table V. The three rotamers are close in stability; indeed, the *g*+ and *t* rotamers are essentially equal in stability. This is not surprising considering the Nvl side chain is not branched and is nonpolar; interactions with the backbone would be expected to be minimal. The slightly higher free energy of the *g*- rotamer is probably due to the steric repulsion of the backbone carbonyl which occurs both in the $C_7^{eq}$ and $C_5$ backbone conformation. Hermans *et al.* (1992) determined the relative free energy for the *t* and *g*- rotamers of α-amino-n-butyric acid (ethyl side chain) in the extended conformation to be 0.36 kcal/mol and 0.86 kcal/mol, in relatively good agreement with our results of 0.09 kcal/mol and 0.54 kcal/mol. The average partition function for $\chi_1$ is 2.34, and the average free energy cost to restrain $\chi_1$ to the *g*+ rotamer is 0.51 kcal/mol (Table V). The overall ΔG, that is, the free energy change correcting for the presence of all the rotamers, is thus slightly less negative than the values calculated for any one rotamer since the free energy for restraining $\chi_1$ must be added to $\Delta G_{overall}$ (eq. (7)).

The free energy changes *in vacuo* show a greater dependence on the peptide model employed. The di- and tetrapeptide are similar, but for the hexapeptide, the *t* and *g*- rotamer results are more than 1 kcal/mol more positive. The positively charged lysine side chain bends back on the peptide to coordinate with the backbone carbonyls, leading to changes in the backbone hydrogen bonding pattern. Instead of $\phi,\psi$ angles in the $C_7^{eq}$ conformation, $\phi$ and $\psi$ are in the $P_{II}$ conformation ($\phi,\psi$ near -80°, +150°) (Table IV). This leads to stabilization of the *t* and *g*- rotamers of Nvl since the backbone carbonyl oxygen of Nvl is pulled toward the Lys side chain and no longer "bumping into" Cγ of Nvl (Fig. 9).

The difference in solvation free energy between Nvl and Ala is calculated to be -1.82 kcal/mol. The analogous difference between propane and methane has been calculated to be 0.4 kcal/mol (Sun *et al.*, 1992), while the experimental value is ~ 0.0 kcal/mol (Ben-Naim and Marcus, 1984) As discussed by Sun *et al.* (1992), however, the correlation between the solvation free energy of isolated small molecules and peptide side chains is not straightforward due to screening of the backbone from water by the side chain. Bai *et al.* (1993,1994) have found experimental evidence of this in exchange rates for amide protons. Although these studies indicate that side chains with β-branching have the largest effect on the solvation of the backbone, Leu was also found to have a significant blocking effect. The equal helix forming propensity of Nvl and Ala found by Lyu *et al.* (1991) is probably a result of the side chain blocking effect of Nvl, which would favor helix formation, being offset by the loss of conformational freedom of the side chain in the α-helix. Sun *et al.* (1992) found the relative solvation free energy of Val and Ala to be 1.1 kcal/mol, compared to 1.82 kcal/mol for Nvl and Ala in our study. It is puzzling that the solvation free energy for Nvl is greater than for Val if β-branching causes greater blocking of peptide groups. However, in the study by Sun *et al.*, the dipeptide was constrained to the $C_5$ backbone conformation, which places the carbonyl and amide hydrogen of the perturbed residue in weak, bent intra-molecular hydrogen bond. In contrast, in our work the peptides sampled the $C_7^{eq}$ conformation the majority of the time, which places the carbonyl and amide hydrogen in solvent exposed positions and closer to the side chain. A larger solvation free energy for Nvl is thus not surprising, and emphasizes the importance of sampling backbone conformations in peptide free energy calculations (Yun and Hermans, 1991; Hermans *et al.*, 1991, 1992; Wang *et al.*, 1995). These results, in conjunction with those of Sun *et al.*. (1992) and Bai *et al.* (1993,1994) suggest that the use of small molecules to model side chain solvation properties (Wolfenden, 1981) may not be appropriate.

*Mse → Nvl*

The results for the mutation Mse → Nvl are shown in Table III. The calculations in solution give fairly consistent results for the different length peptides in all three rotamers despite the change in polarity of the side chain. The dipeptide, Ace-X-Nme, contains no other side chains which might interact with the perturbed side. The Asp and Lys side chains of the longer peptides could interact either with the backbone or perturbed side chain, but remain in the extended conformation and are well solvated in all of the solution simulations.

The change in free energy is least favorable in the *g*+ rotamer and most favorable in the *t* rotamer. The $\phi,\psi$ angles of the mutated residue for the three peptides are near the $C_7^{eq}$ conformation during the perturbation (Table IV). This places the side chain ether oxygen within hydrogen bonding distance of the backbone in the *g*+ rotamer (Fig.10), near the backbone carbonyl in the *t* rotamer, and bridging the backbone HN and O at *g*-. The observed free energy changes thus appear to be driven by interactions of the side chain oxygen with the backbone amide groups. The average relative free energies of the $\chi_1$ rotamers are shown in Fig. 8b. Since the *g*+ rotamer is essentially the only accessible rotamer, the partition function for the $\chi_1$ dihedral is near unity, and the cost to restrain $\chi_1$ to the *g*+ rotamer is small, 0.09 kcal/mol (Table V).

The *in vacuo* calculations show much greater variability (Table III). The difference between the di- and tetrapeptides is caused by the Asn side chain in the tetrapeptide, which hydrogen bonds with its own backbone carbonyl, forcing the backbone of Mse into the $C_7^{eq}$ conformation and causing repulsion between Oγ of Mse its carbonyl oxygens (Fig. 10). In the hexapeptide, the Lys side chain coordinates with the backbone carbonyl of Mse/Nvl in the *g*+ and *g*- rotamers. The relative free energy of

the *g*- rotamer for the Mse side chain is thus significantly lowered because of reduced repulsion with the carbonyl. This is similar to what is seen for the Nvl → Ala mutation, but the effect is larger because of the electronegativity of the ether oxygen. Interestingly, in the *t* rotamer, the Lys side chain coordinates with the Mse side chain, leading to stabilization of this rotamer by 2.94 kcal/mol compared to the *g*+ rotamer (Fig.12). $\Delta G_{rst}$ for the hexapeptide is thus fairly large, 3.03 kcal/mol (Table V). When $\Delta G_{overall}$ is calculated, however, the three peptides give somewhat more consistent results, ranging from 0.64 kcal/mol for the hexapeptide to 2.18 kcal/mol for the tetrapeptide, with an average of 1.31 kcal/mol (Table III).

The solvation free energy, $\Delta\Delta G_{unf\text{-}vac}$, for Mse versus Nvl, using the averages over all the peptides, is 1.31 kcal/mol (Table III). Interestingly, this value is significantly smaller than the solvation free energy found for Dme versus Prp (3.90 kcal/mol). Obviously, there is greater error introduced in the peptide calculations due to the variability of the sampling, mainly in the *in vacuo* calculations. However, as with the Nvl → Ala calculations, the presence of the peptide backbone affects the solvation properties of the side chain and vice-versa (Sun *et al.*, 1992; Bai *et al.*, 1994; Bai and Englander, 1995). The radial distribution function for water hydrogens around the ether oxygen of Mse in the dipeptide (*g*+, $C_7^{eq}$ conformation) and of Dme is shown in Figure 13. While the RDF is a qualitative measurement and does not necessarily reflect solvation free energies directly (Chipot *et al.*, 1994), it can be seen that the oxygen of Dme has a greater occupancy of water hydrogens than Oγ Mse, which interacts with the backbone amide hydrogen in the *g*+, $C_7^{eq}$ conformation (Figure 10). This interaction is lost in both the solution and gas phase perturbations to Nvl, resulting in a smaller relative solvation free energy for Mse and Nvl compared to Dme and Prp.

*Unfolded state models*

Yun-yu *et al.* (1993) suggest that using peptides in an extended conformation is an inadequate model for the denatured state. There are several issues implicite in this statement. One is the appropriateness of assuming the denatured state is completely unfolded and exposed to solvent. The structure of the unfolded state of proteins is, in general, uncertain. However, there is no particular evidence in the case of T4 lysozyme to suggest otherwise. We thus feel that using a solvent exposed model in no way inherently negates the validity of the results. What can be concluded is whether or not the results using a solvent exposed model agree with experiment. A more immediate issue is the precision of using an extended peptide as a model for a solvent exposed unfolded state. Our calculations show that these models do give consistent results for simulations in solution. Increasing the length of the peptide changes overall results little, even with the side chains Asn and Lys, which are quite capable of intra-molecular interactions as seen in the gas phase simulations. The free energy changes appear to be determined mainly by interactions with the backbone of the mutated residue itself. Thus, while Yun-yu *et al.* (1993) state that important intra-molecular interactions are lacking in these models, we believe that the most important interactions, those with the peptide backbone and with water, are included in models as short as a dipeptide. Yun-yu *et al.* (1993) claim that the residues near the mutation site in the folded state need to be included in the unfolded state calculation. While these residues could be expected to make contributions in a partially unfolded structure, there is no particular reason that these specific residues should be included over any others in a completely unfolded state. Indeed, in calculations of protein stability for barnase, using a partially unfolded state in solution gives essentially the same results as an extended tetrapeptide in solution (Sun *et al.*, 1996b).

In summary, our results indicate that a dipeptide in the extended conformation is a reasonable model for a solvent exposed unfolded state. It is necessary, however, to allow enough flexibility in the peptide structure to allow for adjustment of the backbone in response to positioning of the side chain. The variability of the gas phase results illustrate that conformational sampling *in vacuo* is difficult to achieve. This difficulty does not limit the calculation of protein stability per se, but complicates the calculation of solvation free energies and thus analysis of the protein stability results. More extensive sampling protocols using driving potentials and conformational restraints would be easily applicable in the gas phase and improve the accuracy and precision of the calculations. Hermans *et al.* (1992) have shown that these protocols are useful for mutations involving residues that affect backbone flexibility such as valine.

## Folded State Calculations

*Nvl* $\rightarrow$ *Ala*

The results for the mutation from Nvl to Ala using both backbone constraint sets (Table II) are shown in Table VI. The free energy change, calculated over all $\chi_1$ rotamers, is 1.92 kcal/mol using the $\alpha$-constraint set, which allows little adjustment of the surrounding backbone in response to the mutation. As can be seen in Table VII, the distances between C$\beta$ of residue 133 and surrounding atoms change little during the mutation when the $\alpha$-constraints are employed. The interatomic distances for both the Nvl and Ala structures are more similar to those in the native X-ray crystal structure than those in the A133 mutant. The root-mean-square deviations (RMSD's) of the heavy atoms from the Leu X-ray structure for all residues, the cavity residues, and the backbone atoms of the cavity residues increase by ~0.1-0.3 Å (Table VIII). The $\phi$, $\psi$, and $\chi_1$ torsions for Nvl are similar to the those found in the Leu crystal structure, -58°, -41°, and -80° compared to -65°, -30°, -80° (Table IX). The $\chi_2$ side chain dihedral is able to adjust from 161° in the crystal structure to a more ideal 177° in the simulation

since there is only one δ-carbon on the side chain. Upon mutation to Ala, ψ decreases from -41° to -35° while ϕ remains essentially unchanged. The Ala crystal structure shows somewhat similar, but larger changes from the Leu crystal structure, with a ϕ,ψ of -70°, -19°.

The simulation of Nvl → Ala with the cavity-constraints gives an overall ΔG of 1.50 kcal/mol. This free energy change is 0.42 kcal/mol less than the simulation with tighter constraints, and suggests the protein has adjusted to the mutation more successfully (Table VI). Upon mutation to Ala, the four distances measured which are shorter in the Ala crystal structure versus the Leu crystal structure also decrease in the simulation (Figure 14). The structural changes occur mostly in the helices which are nearest the surface of the protein, Helix G (115 - 123), Helix H (126 - 134), and Helix I (137 - 141) (Figure 4). Helix J (143 - 155) shows some minor movements, while Helix E (93 - 106) varies little. The largest difference between the crystal structures is the 133Cβ - 114Cα distance, which decreases by 1.09 Å. This distance decreases by 0.65 Å in the simulation, compared to 0.24 Å for the α-constraint set simulation. The 133Cβ - 117Cα distance decreases by 0.98 Å compared to 0.55 Å in the crystal structures. Residues 136 and 138, at the junction between Helices H and J of the four-helix bundle, show similar movement. The X-ray structures show an increase in distance between the carbonyl oxygens in the preceding turn of the α-helix and 133Cβ. The simulations display a decrease in the 133Cβ - 129O distance, but a slight increase in the 133Cβ - 130O distance. The ϕ,ψ angles before mutation, -62° , -35°, match the Leu crystal structure values of -65°, -30°. The Ala simulation structure, however, has ϕ,ψ angles of -58°, -42° compared to -70°, -19° for the Ala crystal structure. The structural changes seen in the crystal structures show Cβ of Ala shifted in toward the cavity via changes in ϕ and ψ, and movement of the backbone and certain side chains in specific regions. In the simulations, the structural changes are associated with a more

generalized repositioning of Helix H as well as tighter packing of residues mostly in turn regions. The magnitudes of these movements are consistent with those seen in the various mutations in the C-terminal domain of T4 lysozyme, and are suggestive of a backbone that has sufficient flexibility in certain regions to accommodate amino acid substitutions (Baldwin *et al.*, 1993; Baldwin and Matthews, 1994; Matthews, 1995; Zhang *et al.*, 1995).

The perturbation in the Ala → Nvl direction with the cavity-constraints was also conducted, and gave a free energy change similar to the Nvl → Ala direction (1.27 kcal/mol vs. 1.47 kcal/mol, Fig.7b). The distances discussed above were also determined for the Nvl endpoint. All of the distances reverse their trend from the Nvl → Ala simulation and return, for the most part, to near their values before the perturbation cycle was begun. The distances which show the most hysteresis are for residues 114 and 138, which are not central components of the four-helix bundle and are in fairly flexible regions. When residue 133 is mutated back to Nvl, the RMSD's for all residues and the cavity residues decrease, although they do not return to their original values. This decrease is encouraging, though, considering that the total length of the perturbation cycle is over 1 ns. The RMSD's for the backbone atoms in the cavity show a small increase, indicating that the backbone structure has accommodated the increased side chain volume slightly differently than in the starting structure. The $\phi$ backbone dihedral is fairly unchanged from the Ala simulation structure as well as the equilibrated Nvl structure, while $\psi$ has increased from -35° to -52°.

It can be seen when comparing both the structural and thermodynamic results of the simulations with the two constraint sets that using the less rigid model allows for adjustment of the protein to the mutation and stabilizes the protein by 0.42 kcal/mol (Table VI). This free energy difference can be viewed as the free energy gained from

repacking of the protein backbone. Eriksson *et al.* (1992) estimated that the cost of creating a cavity in the interior of a protein is approximately 24 cal/mol/Å$^3$. The decrease in cavity volume for the L133A mutant due to repacking of the protein was 31 Å$^3$, which corresponds to 0.74 kcal/mol gained from adjustment of the protein. The mutation Nvl → Ala causes less of a volume change (~60 Å$^3$) than the Leu → Ala substitution (~90 Å$^3$), resulting in less repacking. Our estimate of 0.42 kcal/mol is therefore quite consistent with the estimate from experiment, and represents the magnitude of stabilization that can be gained by allowing increased backbone movement. While the simulations do not reproduce precisely the structural changes seen in the Ala crystal structure, the types of changes and their magnitudes are similar. Since the α-constraint set does not constrain residues 131-135 and does allow some motion of the backbone for other residues, the value of 0.42 kcal/mol may be conservative.

Clearly, the results from free energy calculations are sensitive to constraints placed on the structure. These restraints are needed to prevent the structure from drifting too far from the crystal structure. This is in contrast to the desire for increased conformational sampling and reproduction of structural changes seen upon mutation. The constraints, cutoffs, and frozen regions of protein structure must thus be carefully chosen, hopefully in conjunction with experimental data. In contrast to the view that a longer simulation time for the folded state of a protein will lead to more converged results (Yun-yu, 1993), the regions of conformational space sampled can become irrelevant, and hysteresis, if checked, will likely be larger than found here. The use of periodic water boxes for the simulation of proteins and improved force fields should lead to a decreased need for positional constraints (Fox and Kollman, 1995).

The relative free energies of the $\chi_1$ rotamers of Nvl for the cavity-constraint simulations are shown in Figure 15a and Table V. Not surprisingly, the *g*- rotamer is highly disfavored. Residue 133 is near the C-terminal end of an α-helix, and thus its side chain in the *g*- rotamer will encounter steric repulsion from the previous turn in the helix (Figure 16). The average distance between 133 Cγ and 129 O is 3.40 Å in the equilibrated Nvl structure. The steric clash causes $\phi$ to increase to -70° and $\psi$ to decrease to -19°, allowing the α-helix to "open-up". These changes in backbone dihedral are similar to those seen in simulations of t-butyl side chain groups which cause disruption of α-helices (Cornish *et al.*, 1994). The *t* rotamer is also less favored than the *g*+ rotamer; the most likely cause is both the backbone carbonyl of residue 133 and the side chain of Trp 138, which protrudes into the cavity near the *t* rotamer. In the cavity-constraint simulation, $\chi_1$ adopts an average angle of -134° and the Cγ is about 3.80 Å away from the Trp 138 side chain. If $\chi_1$ is set to 180° in the X-ray crystal structure, this distance is only 2.76 Å, indicating the adjustment required of the protein to accommodate this rotamer. The calculated relative free energies of the $\chi_1$ rotamers are similar to those found for the Leu188 mutant of viral protein 1 of HRV14 in the liganded conformation (Wade and McCammon, 1992), and are consistent with the prevalence of the *g*+, t, and *g*- rotamers of $\chi_1$ for Leu found in α-helices in X-ray crystal structures (McGregor *et al.*, 1987).

**Mse→Nvl**

The free energy change in the folded state for the perturbation from Mse to Nvl is 0.03 kcal/mol (Table VI). The cavity surrounding residue 133 is nonpolar; besides the backbone, the only nearby polar group is the hydroxyl of Ser 117. This side chain, however, has hydrogen bonding partners, and is not positioned well to interact with the ether oxygen of Mse. The backbone is also a rather poor hydrogen bond donator since its amide hydrogens are involved in intra-helical hydrogen bonding. There is thus no significant increase in free energy upon mutation to Nvl, which would be expected if

there were an unpaired hydrogen bonding partner in the Nvl structure (Blaber, 1993). Since the side chains are isosteric, contributions from packing should be minimal; this was the motivation for using the α-constraint set for this mutation. The interatomic distances between 133 Oγ and nearby atoms are similar for the *g*+ rotamer of Mse and the Leu X-ray structure (Table X). The ϕ,ψ angles for Mse are similar to the Leu X-ray values, and are relatively unchanged for the Nvl endpoint (Table IX). The RMSD's of the heavy atoms for the Mse equilibrated structure are similar to the starting structure for the Nvl → Ala mutation (Table VIII). Upon mutation to Nvl, the values increase, but to a lesser extent than in the Nvl → Ala perturbation.

Because of the electrostatic interactions of the ether moiety in the protein interior, the *g*- and *t* rotamers are even more disfavored than for NVL (Fig.15a,b, Table V). 133Oγ can hydrogen bond with 133HN in either the *g*+ or *g*- rotamer, which offsets the proximity of 129O (~3.1 Å); however, in the *g*- conformation, Oγ also in near O130 (Table X). $\chi_1$ adopts an average angle of 39° to minimize this repulsion (Table IX). In the *t* rotamer, the backbone carbonyl oxygen of residue 133 and Trp 138 are near Oγ, and there are no hydrogen bond donors in the vicinity. However, in the *g*+ rotamer, Oγ can interact with both 133 HN and the hydroxyl hydrogen of Ser117, although the angles are not ideal (Fig.17, Table X). The relative free energies of the *t* and *g*- rotamers may have been closer to that of the *g*+ rotamer if the cavity-constraints had been used for this perturbation since these two rotamers involve unfavorable vdW interactions. The free energy change in the preferred *g*+ conformation, however, would likely remain fairly unchanged since it is in a low energy state and there is no strong driving force for conformational changes and repacking.

**Calculated Stability Changes**

The final calculated stability changes are given in Table XI. The values are reported relative to unfolding; a negative value indicates the protein is more amenable to

unfolding, ie. is less stable. There is no direct experimental data for the difference in stability between the Nvl and Ala mutants. However, Mendel *et al.* (1992) measured the 2-carbon ethyl side chain mutant at site 133 as being ~2.2 kcal/mol less stable than Nvl. This value can serve as a lower bound for the loss in stability. Eriksson *et al.* (1992) determined the stability change for the L133A mutant to be -3.6 kcal/mol. Since this is a larger perturbation then the Nvl → Ala mutation, -3.6 kcal/mol is likely the maximum destabilization which could be caused by the Nvl → Ala mutation. The calculated value of -3.48 kcal/mol is thus within range of the estimated experimental value. The experimental difference in stability between the Mse and Nvl mutants is 1.56 kcal/mol as determined via comparison to the native Leu. The calculated $\Delta\Delta G$ of 1.84 kcal/mol agrees quite well with this value. The agreement between the experimental and calculated free energies, while encouraging, does not necessarily insure that the results have been obtained by accurately modeling the physical processes involved in protein stability. However, calculating results which agree with experiment for two mutations which present different physical perturbations to the protein suggests that the key interactions have been included in our model. In the following analysis, we discuss the contributing factors to the changes in stability which can be derived from the simulations and compare them to experimental data.

*Nvl → Ala*

Mendel *et al.* estimated contributions to the stability of T4 lysozyme mutants based on changes in surface area of the nonpolar side chains they studied. The mutation Nvl → Ala reduces the nonpolar surface area by about 56 $Å^2$. It has been estimated, based on octanol to water partition coefficients, that the value of 24 cal/mol/$Å^2$ provides a reasonable estimate of the contribution to protein stability from the hydrophobic effect. The corresponding free energy using this empirical estimate for the Nvl → Ala mutation is then -1.34 kcal/mol. Following the work of Nicholls *et al.* (1991), estimates were also made for cost of cavity formation in the protein based on the enthalpy of

sublimation for hydrocarbons (20.7 cal/mol/Å$^2$) and for the tighter packing in the solid-like native state than in solution (Lazaridis *et al.*, 1995) from the enthalpy of melting for hydrocarbons (20.3 cal/mol/Å$^2$); these two factors combined contribute -2.30 kcal/mol to the destabilization of the native state. And finally, an estimate of entropic changes due to restriction of side chain motion in the folded protein favors Ala over Nvl by +0.32 kcal/mol if the relative energy of the *gauche* versus *trans* rotamers is estimated as 0.90 kcal/mol and it is assumed $\chi_2$ can adopt two conformations in the protein. The overall estimate is -3.28 kcal/mol, quite close to the simulation results of -3.48 kcal/mol and within the experimental range of -2.2 to -3.6 kcal/mol.

Free energy perturbation calculations enable the direct determination of the contribution from the unfolded state, which is the difference in the free energy change in solution and the gas phase Unfortunately, the *in vacuo* free energy changes are variable for the different length peptides. Using the values from the dipeptide calculations may give more straightforward results because there are no interactions with other side chains, but doing so would change $\Delta\Delta G_{overall}$ by only 0.38 kcal/mol. We thus use the values averaged over all the peptides in the *in vacuo* runs for both Nvl → Ala and Mse → Nvl in Table XI and in the following discussion. The calculated solvation free energy difference for Ala and Nvl is $\Delta\Delta G_{unf\text{-}vac}$ = -1.82 kcal/mol. It is difficult to make analogies to transfer free energies of amino acids since there is no similar mutation amongst the natural amino acids. The transfer free energies from octanol to water for N-acetyl amino acid amides were measured by Fauchere and Pliska (1983); the difference between Ala and Leu is -1.90 kcal/mol. The Ala to Leu mutation is a greater change in side chain volume, but $\Delta\Delta G_{unf\text{-}vac}$ includes losses due to packing changes and would be expected to be larger than the octanol to water transfer value. In any case, our results suggest that a significant portion of the destabilization of for the Nvl → Ala mutation arises from stabilization of the unfolded state.

The free energy change in the protein can be determined in an analogous manner. $\Delta\Delta G_{fold\text{-}vac}$ is 1.66 kcal/mol, and reflects the loss of packing interactions in the protein interior. This compares to 2.30 kcal/mol as calculated empirically using surface areas (Nicholls *et al.*, 1991; Mendel *et al.*, 1992). Eriksson *et al.* (1992) estimated that of the -3.6 kcal/mol difference in stability of the L133A mutant, approximately -1.6 kcal/mol were from the loss of packing interactions in the native state, although this number would have been greater if repacking had not occurred. As discussed previously, the cavity-constraint set used in the simulation allows for some repacking, which stabilizes the protein by about 0.4 kcal/mol (Table VI). The calculated results thus appear to be in agreement with both theoretical and experimental estimates of stability changes which occur upon loss of packing in the protein interior, and suggest that this contribution is on the same order of magnitude as the contribution from the hydrophobic effect. This is not to say that vdW interactions contribute to protein stability equally as much as the hydrophobic effect, but that the formation of cavities in proteins is highly unfavorable (Dill, 1990; Lazaridis *et al.*, 1995).

Finally, the free energy gained when a dihedral is moved from the constricted interior of the protein to the conformationally degenerate unfolded state can be calculated for $\chi_1$ since the partition functions, $q_{unr}$, are known (Table XII). There are 2.34 equally accessible rotameric states in the unfolded state and essentially only one in the folded state, leading to a free energy gain of -0.48 kcal/mol upon unfolding. This value is slightly larger than the estimate of -0.22 kcal/mol using the method of Mendel *et al.* (1992) because the $\chi_1$ rotamers in the unfolded state only differ in free energy by 0.09 and 0.54 kcal/mol as calculated from the simulations as opposed to the estimate of 0.90 kcal/mol. The result from the simulations agree well with the empirical estimates by Pickett and Sternberg (1993) and by Nicholls *et al.* (1991) of the cost of restraining side

chain dihedrals in folded structures as 0.45 kcal/mol per dihedral. Creamer and Rose (1994), using Monte Carlo simulations to explore accessible conformations of side chains, calculated the loss of conformational free energy for Nvl upon forming an α-helix to be 0.23 kcal/mol for the entire side chain, ie. both $\chi_1$ and $\chi_2$. Hermans *et al.* (1992) calculated that about 0.25 kcal/mol was lost for α-amino n-butyric acid, which is similar to Nvl but lacking the δ-carbon, upon going from a β-sheet to an α-helix conformation. Our value of 0.48 kcal/mol for $\chi_1$ of Nvl is likely higher since there is more restriction of the side chain in the protein.

*Mse→Nvl*

The calculated solvation free energy, $\Delta\Delta G_{unf\text{-}vac}$, of Nvl vs. Mse is 1.31 kcal/mol. As discussed earlier, this value is less than the experimental difference in solvation free energy between Dme and Prp, 3.89 kcal/mol, and is likely due to hydrogen bonding of the ether oxygen with the peptide backbone. Mendel *et al.* (1992) measured the transfer free energy between octanol and water for Nvl vs. Mse to be 1.8 kcal/mol. Since the packing changes are relatively minor for this mutation, relative agreement between the gas phase to water and the octanol to water transfer free energies is reasonable.

The change in free energy in the folded state versus gas phase, $\Delta\Delta G_{fold\text{-}vac} = -0.53$, indicates that Nvl is slightly more stable than Mse in the protein interior. Since these side chain are isosteric, the greater stability of Nvl is not likely caused by improved packing. Instead, the difference in stability is likely due the lack of good hydrogen bonding partners for Mse. Our data thus support the assumption of Mendel *et al.* (1992) that the difference in stability between Mse and Nvl is due largely to changes in stability of the unfolded state, and conversely, the data of Mendel *et al.* support our use of a short peptide in the extended, solvent exposed conformation as a model of the denatured state.

In contrast to the Nvl side chain, there is little cost for restricting $\chi_1$ of Mse in the protein interior since Mse in the unfolded state has, on average, only ~1.2 equally accessible $\chi_1$ rotamers. The free energy gained upon unfolding is -0.09 kcal/mol (Table XII). This was a surprising but logical result. The Mse side chain is clearly unique. The $\chi_1$ dihedral of Ser, which also has an Oγ, does not show a similar, highly restricted distribution because it is able to form hydrogen bonds between its hydroxyl hydrogen and carbonyl oxygens of the backbone (McGregor *et al.*, 1987). Mse has the conformational restrictions typically seen with Val and Ile, but in this case due to electrostatic rather than steric confinement.

*Free Energy Decomposition*

The appropriateness of decomposing free energy changes into contributions from components of the Hamiltonian, ie. vdW and electrostatics, or from various groups of atoms, has been debated recently in the literature (Yun-yu *et al.*, 1993; Boresch *et al.*, 1994; Mark and van Gunsteren, 1994; Smith and van Gunsteren, 1994; Boresch and Karplus 1995). Since we have studied both an electrostatic perturbation with almost no vdW changes and a vdW perturbation with minimal electrostatic changes, we felt it was appropriate to present free energy components for our calculations, which are shown for the dipeptide and the folded state in Table XIII. In order to simplify analysis, only the components for the low-energy *g+* rotamer are presented.

The contribution to $\Delta\Delta G_{unf\text{-}fold}$ for Nvl → Ala arises mainly from the vdW component in the folded state (3.09 kcal/mol), which is consistent with the loss of packing interactions. The vdW component in the unfolded state contributes -1.07 kcal/mol, as would be expected for a change in volume of a solute in water. The electrostatic contribution is fairly small, while the intra-residue terms and the bond-PMF correction are negligible. This in contrast to results of Prevost *et al.* (1991) and Yun-yu *et al.*

(1993), who found that intra-residue terms, ie. bond, angle, and dihedral terms, make a significant contribution to changes in stability (Sun *et al.*, 1996a).

The difference in stability of Mse and Nvl is due mainly to electrostatic contributions. However, these contributions arise from the folded state rather than in solution (-1.48 vs. 0.03 kcal/mol). A favorable change in the electrostatic component in the protein is reasonable considering the nonpolar nature of the environment surrounding site 133. This is largely offset, though, by an increase in the vdW component (1.15 kcal/mol). In solution, the change in free energy is due almost entirely to the vdW component (1.59 kcal/mol). This is in contrast to what would be expected for an electrostatic perturbation in solution, and highlights the difficulty of analyzing free energy components (Yun-yu *et al.*, 1993). In the case of the Mse → Nvl mutation in solution, the free energy change may be dominated by the vdW term because upon mutation to Nvl, waters "bump" the side chain in an effort to hydrogen bond with the amide hydrogen, whereas with Mse, this amide hydrogen can form a hydrogen bond with Oγ of the side chain. This is also consistent with our hypothesis earlier that the relatively small calculated solvation free energy for Mse vs. Nvl is due to the presence of the intra-molecular hydrogen bond in solution and the gas phase.

Our analysis of free energy components exemplifies both the usefulness of these components and the risk of over-interpreting them. When supported by other thermodynamic and structural data, free energy components can provide useful insights to intermolecular interactions. However, due to the inherent interdependence of vdW and electrostatic interactions which will especially be present in complicated systems such as folded proteins and peptides in solution, it cannot be assured that these components provide a definitive assessment of the source of free energy changes.

*Simulations Without Side Chain Restraints*

The results from 400 ps simulations without side chain restraints are shown in Table XIV. Calculations were conducted only for the folded state and the dipeptide in solution. All simulations were started with $\chi_1$ and $\chi_2$ in the preferred *g*+,*t* conformation. The result for the Nvl → Ala mutation, -3.36 kcal/mol, is essentially the same as when sampling of all the $\chi_1$ rotamers was accounted for. In the dipeptide perturbation, the side chain stays in the *g*+,t conformation for the first 100 ps, then $\chi_2$ begins to extensively sample all three rotamers as the nonbond parameters for Nvl decrease (Fig. 18a). $\chi_1$ remains in the *g*+ rotamer throughout the calculation. In the protein, a similar trend is seen, except that both side chain dihedrals begin sampling other rotamers after about 100 ps (Fig. 18b). It is not surprising, then, that the free energy change is similar since only the preferred rotamers are sampled while the system is near the Nvl state.

The results for the unrestrained Mse → Nvl perturbation are also very close to the values obtained with forced sampling of $\chi_1$, 1.75 vs. 1.84 kcal/mol. In the dipeptide run, $\chi_2$ almost immediately changes conformation to the *g*+ rotamer, returning to the *t* rotamer for over 200 ps and sampling both rotamers for the remainder of the simulation (Fig. 18c). This sampling, however, likely has little effect on the free energy since the δ-carbon is a relatively unperturbed atom in the Mse → Nvl mutation. $\chi_1$ does not leave the *g*+ rotamer. In the protein, $\chi_1$ and $\chi_2$ remain in the *g*+,t conformation for over 300 ps (Fig. 18d). Near the end of the simulation, however, $\chi_2$ flips to the *g*+ rotamer, followed by $\chi_1$ switching briefly to *t* 10 ps later. These conformational changes are consistent with the Nvl side chain in the protein, where the *t* rotamer is more accessible for Nvl than for Mse.

The results without side chain restraints are essentially the same as those with thorough sampling of the $\chi_1$ rotamers. Wade and McCammon (1992) also found that inclusion

of all rotamers changed the final results by less than 0.5 kcal/mol. Extensive sampling of side chain rotamers thus may not be necessary if the simulation is conducted in the favored rotamer and there are no significant structural changes associated with the mutation. The success of side chain packing algorithms in eliminating high energy rotamers suggests that a combination of calculation methodologies may enable both appropriate sampling and high quality energetic calculations (Tuffery *et al.*, 1993; Roitberg and Elber, 1994)).

## Conclusion

We have presented free energy calculations of a van der Waals dominated and an electrostatic dominated perturbation to the amino acid sequence of T4 lysozyme. The results were obtained with converged simulations, display little hysterisis, reproduce subtle structural changes in the folded protein, and are in good agreement with similar calculations in other systems and with the experimental free energy differences.

Extensive insight and correlation with empirical estimates were obtained from the simulations. The calculated free energy differences emphasize the need to account for changes in packing when analyzing the stability of amino acid mutants. We have shown the importance of subtle backbone movement in response to mutation, and the role of side chain conformational freedom was presented in quantitative terms. Our results are consistent with , but no definitive of, an unfolded state which has its side chains well exposed to solvent, and suggest that the solvation of amino acid side chains is complex due to interactions with the peptide backbone.

Determining protein stability with free energy calculations is a demanding task. There are many methodological and theoretical difficulties, the most challenging of which is adequate and relevant conformational sampling. It is clear that when simulations are conducted that do not properly sample conformational space the interpretation of results

is problematic and can lead to misleading conclusions (Yun-yu *et al.*, 1993). Since it is difficult to know if structural changes seen in simulations are representative of low energy regions of conformational space without experimental structural data, we agree with Yun-yu *et al.* (1993) that free energy calculations, with current methodologies, cannot often be used to predict protein stability. We disagree, however, that useful insights cannot be gained from these calculations when both thermodynamic and structural correlations have been made with experiment.

## References


Baldwin,E.P., Hajiseyedjavadi,O., Baase,W.A., and Matthews,B.W. (1993) *Science*, **262**, 1715-1718.

Bai,Y., Milne,J.S., Mayne,L. and Englander,S.W. (1993) *Proteins*, **17**, 75-86.

Bai,Y. and Englander,S.W. (1994) *Proteins*, **18**, 262-266.

Baldwin,E.P., Hajiseyedjavadi,O., Baase,W.A. and Matthews,B.W. (1993) *Science*, **262**, 1715-1718.

Baldwin,E.P. and Matthews,B.W. (1994) *Curr. Opin. Biotech.*, **5**, 396-402.

Bayly,C.I., Cieplak,P, Cornell,W.D. and Kollman,P.A. (1993) *J. Phys. Chem.*, **97**, 10269-10280.

Ben-Naim,A. and Marcus,Y.J. (1984) *J. Chem. Phys.*, **81**, 2016.

Blaber,M., Lindstrom,J.D., Gassner,N. Xu,J., Heinz,D.W. and Matthews,B.W. (1993) *Biochemistry*, **32**, 11363-11373.

Boresch,S., Archontis,G. and Karplus,M. (1994) *Proteins*, **20**, 25-33.

Boresch,S. and Karplus,M. (1995) *J. Mol. Biol.*, **254**, 801-807.

Chipot,C., Millot,C., Maigret,B. and Kollman,P.A. (1994) *J. Chem. Phys.*, **101**, 7953-7962.

Cornell,W.D., Cieplak,P., Bayly,C.I. and Kollman,P.A. (1993) *J. Am. Chem. Soc.*, **115**, 9620-9631.

Cornell,W.D., Cieplak,P., Bayly,C., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) *J. Am. Chem. Soc.*, **117**, 5179-5197.

Cornish,W., Kaplan,M., Veenstra,D., Kollman,P.A. and Schultz,P.G. (1994) *Biochemistry*, **33**, 12022-12031.

Creamer,T.P. and Rose,G.D. (1994) *Proteins*, **19**, 85-97.

Dang,L., Merz,K.J. and Kollman,P. (1989) *J. Am. Chem. Soc.*, **111**, 8505-8508.

Dill,K.A. (1990) *Biochemistry*, **29**, 7133-7155.

Eriksson,A.E., Baase,W.A., Zhang,X.-J., Heinz,D.W., Blaber,M., Baldwin, E.P., and Matthews,B.W. (1992)*Science* , **255**, 178-183.

Eriksson,A.E., Baase,W.A., Matthews,B.W. (1993)*J.Mol. Biol.*, **229**, 747-769.

Fauchere,J. and Pliska,V. (1983) *Eur. J. Med. Chem.*, **18**, 369-375.

Fox,T. and Kollman,P.A. (1996) *Proteins*, **25**, 315-334.

Freire,E. (1993) *Arch. Biochem. Biophys.*, **303**, 181-184.

Gough,C.A., DeBolt,S.E. and Kollman,P.A. (1992) *J. Comput. Chem.*, **13**, 963-970.

Heiner,A.P., Berendsen,H.J.C. and van Gunsteren,W.F. (1993) *Protein Eng.*, **6**, 397- 408.

Heinz,D.W., Baase,W.A., Zhang,X.J., Blaber,M., Dahlquist,F.W. and Matthews, M.W. (1994) *J. Mol. Biol.*, **236**, 869-886.

Hermans,J., Yun,R.H. and Anderson,A.G. (1991) *J. Comput. Chem.*, **13**, 429-442.

Hermans,J., Anderson,A.G. and Yun,R.H. (1992) *Biochemistry*, **31**, 5646-5653.

Hine,J. and Mookerjee,P.K.. (1975) *J. Org. Chem.*, **40**, 292-298.

Hodel,A., Simonson,T., Fox,R.O. and Brunger,A.T. (1993) *J. Phys. Chem.*, **97**, 3409-3417.

Jorgensen,W., Chandrasekhar,J, Madura,J, Impey,R. and Klein,M. (1983) *J. Chem. Phys.*, **79**, 926-935.

Karpusas, M., Baase,W.A., Matsumura,M. and Matthews,B.W. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 8237-8241.

Kollman,P.A. (1993) *Chem. Rev.*, **93**, 2395-2417.

Lazaridis,T., Archontis,G. and Karplus,M. (1995) *Adv. Prot. Chem.*, **47**, 231-306.

Lyu,P.C., Sherman,J.C., Chen,A. and Kallenbach,N.R. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 5317-5320.

Mark,A. and van Gunsteren,W.F. (1994) *J. Mol. Biol.*, **240**, 167-176.

Mark,A.E., van Helden,S.P., Smith,P.E., Janssen,L.H.M. and van Gunsteren,W.F. (1994) *J. Am. Chem. Soc.*, **116**, 6293-6302.

Matthews,B.W. (1993) *Ann. Rev. Biochem.*, **62**, 139-160.

Matthews,B.W. (1995) *Adv. Prot. Chem.*, **46**, 249-278.

Matthews,B.W. (1996) *Faseb J.*, **10**, 35-41.

McGregor,M.J., Islam,S.A. and Sternmerg,M.J.E. (1987) *J. Mol. Biol.*, **198**, 295-310.

Mendel,D., Ellman,J.A., Chang,Z.Y., Veenstra,D.L., Kollman,P.A, and Schultz,P. (1992) *Science*, **256**, 1798-1802.

Mitchell,M.J. and McCammon,J.A. (1991) *J. Comput. Chem.*, **12**, 271- 275.

Nicholls,A., Sharp,D.A. and Honig,B. (1991) *Proteins*, **11**, 281-296.

Nicholson,H., Anderson,D.E., Daopin,S. and Matthews,B.W. (1991) *Biochemistry*, **30**, 9816-9828.

Pearlman,D.A. and Kollman,P.A. (1989) *J. Chem. Phys.*, **91**, 7831-7839.

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

Pearlman,D.A. and Kollman,P.A. (1991) *J. Chem. Phys.*, **94**, 4532-4545.

Pearlman,D.A. (1994a) *J. Phys. Chem.*, **98**, 1487-1493.

Pearlman,D.A. (1994b) *J. Comput. Chem.*, **15**, 105-123.

Pearlman,D.A., Case,D.A., Caldwell,J.C., Ross,W.,III,T.C., DeBolt,S., Ferguson, D., Seibel,G. and Kollman,P.A. (1995) *AMBER 4.1*, University of California, San Francisco.

Pickett,S.D. and Sternberg,M.J.E. (1993) *J. Mol. Biol.*, **231**, 825-839.

Roitberg,A and Elber,R. (1994) In Merz,Jr.,K. and Le Grand,S., (eds),*The Protein Folding Problem and Tertiary Structure Prediction*.. Birkhauser, Boston, pp. 1-41.

Prevost,M., Wodak,S., Tidor,B. and Karplus,M. (1991) *Proc. Natl Acad. Sci.USA*, **88**, 10880-10884.

Ryckaert,J., Ciccotti,G. and Berendsen,H.J.C. (1977) *J. Comput. Phys.*, **23**, 327-341.

Smith,P. and van Gunsteren,W. (1994) *J. Phys. Chem.*, **98**, 13735-13740.

Smith,L.J., Mark,A.E., Dobson,C.M. and van Gunsteren,W.F. (1995) *Biochemistry*, **34**, 10918-10931.

Sun,Y., Spellmeyer,D., Pearlman,D.A. and Kollman,P.A. (1992) *J. Am. Chem. Soc.*, **114**, 6798-6801.

Sun,Y.-C., Veenstra,D.L. and Kollman,P.A. (1996a) *Protein Eng.*, **9**, 273-281.

Sun,Y.-C., Daggett,V. and Kollman,P.A. (1996b) in preparation.

Tidor,B. and Karplus,M. (1991) *Biochemistry*, **30**, 3217-3228.

Tidor,B. (1994) *Proteins*, **19**, 310-323.

Tuffery,P., Etchebest,C., Hazout,S. and Lavery,R. (1993) *J. Comput. Chem.*, **14**, 790-798.

van Gunsteren,W.F. and Weiner,P. (1989) *Computer Simulation of Biomolecular Systems*. ESCOM Science Publishers, BV, Leiden, The Netherlands and references therein.

Veenstra,D.L., Ferguson,D.M. and Kollman,P.A. (1992) *J. Comput. Chem.*, **13**, 971-978.

Wade,R. and McCammon,J.A. (1992) *J. Mol. Biol.*, **225**, 679-696.

Wang,L. and Hermans,J. (1994) *J. Phys. Chem.*, **100**, 9129-9139.

Wang,L., O'Connell,T., Tropsha,A. and Hermans,J. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 10924-10928.

Weaver,L.H. and Matthews,B.W. (1987) *J. Mol. Biol.*, **193**, 189-199.

Weiner,S.J., Kollman,P.A., Nguyen,D.T. and Case, D.A. (1986) *J. Comput. Chem.*, **7**, 230-252.

Wolfenden,R., Anderson,L., Cullis,P. and Southgate,C. (1981) *Biochemistry*, **20**, 849-855.

Yun,R.H. and Hermans,J. (1991) *Protein Eng.*, **4**, 761-766.

Yun-yu,S., Mark,A.E., Wang,C.-x., Fuhua,H., Berendsen,H.J.C. and van Gunsteren,W.F. (1993) *Protein Engin.*, **6**, 289-295.

Zhang,X-j., Wozniak,J.A. and Matthews,B.M. (1995)*J. Mol. Biol.*, **250**, 527-552.

## Figure Captions

**Fig. 1.** Amino acid perturbations studied by free energy calculations.

**Fig. 2.** Flatwell restraints (···) with $r_1 = 110°$, $r_2 = 125°$, $r_3 = 235°$, $r_4 = 250°$, and $k_2$, $k_3 = 50$ kcal/mol/rad$^2$ (see Eq. (4)) and representative cosine torsional term (---) with 3 kcal/mol barrier.

**Fig. 3.** Point charges and atom types of molecules studied.

**Fig. 4.** T4 lysozyme. Residue 133 is located in the C-terminal in the four-helix bundle consisting of Helices E (93 - 106), G (115 - 123), H (126 - 134)and J (143 - 155).

**Fig. 5.** Hydrophobic cavity into which the side chain of residue 133 protrudes. The cavity shown here was generated without a side chain at 133. In the native protein (L133), the cavity is divided into two smaller cavities.

**Fig. 6.** Convergence behavior of (a) Dme → Prp dipeptide *in vacuo*, (b) Dme → Prp in solution, (c) Mse → Nvl dipeptide *in vacuo*, (d) Mse → Nvl dipeptide in solution, (e) Mse → Nvl in the folded state, (f) Nvl → Ala dipeptide *in vacuo*, (g) Nvl → Ala dipeptide in solution, (h) Nvl → Ala in the folded state.

**Fig. 7.** Hysteresis of Nvl → Ala → Nvl in (a) the dipeptide in solution and (b) in the folded state.

**Fig. 8.** Conformational free energies of $\chi_1$ rotamers relative to $g+$ rotamer for average of (a) Nvl peptides in solution and (b) Mse peptides in solution.

**Fig. 9.** Nvl hexapeptide *in vacuo*. The Lys side chain interacts with the backbone carbonyls, leading to the $P_{II}$ conformation ($\phi,\psi$ near -80°, +150°) of Nvl.

**Fig. 10.** Interaction of O$\gamma$ of Mse in solution with the backbone amide hydrogen in the $\chi_1 = g+$ rotamer. A similar interaction is seen *in vacuo*.

**Fig. 11.** In the Mse tetrapeptide *in vacuo*, Asn hydrogen bonds with the backbone, leading to the $C_7^{eq}$ conformation and an unfavorable interaction between Oγ and the carbonyl oxygen of Mse.

**Fig. 12.** Lys bends back to interact with Oγ of Mse in the hexapeptide *in vacuo* when $\chi_1$ is in the *t* rotamer, stabilizing this rotamer over the *g*+ rotamer by 2.94 kcal/mol.

**Fig. 13.** Radial distribution function of water hydrogens around the ether oxygens of Mse (—) and Dme (---).

**Fig. 14.** Repacking difference between the Nvl → Ala endpoints for the simulations conducted with α-constraints (dark lines) and with cavity-constraints (light lines). Distances are given in Table VII.

**Fig. 15.** Conformational free energies of $\chi_1$ rotamers relative to *g*+ rotamer for (a) Nvl and (b) Mse in the folded state. The values shown for Nvl were obtained from the cavity-constraints simulations and those for Mse from the α-constraints simulations.

**Fig. 16.** Clash of Cγ of Nvl with carbonyl oxygens in the previous turn of the α-helix for the $\chi_1 = g-$ rotamer in the folded state.

**Fig. 17.** Mse in the $\chi_1 = g+$ rotamer in the folded state. Oγ is near the backbone amide hydrogen and the hydroxyl hydrogen of Ser117, although hydrogen bonding angles are not ideal.

**Fig. 18.** Sampling of φ, ψ, $\chi_1$, and $\chi_2$ in simulations without side chain restraints for (a) Nvl → Ala dipeptide in solution, (b) Nvl → Ala in the folded state, (c) Mse → Nvl dipeptide in solution, and (d) Mse → Nvl in the folded state. All simulations were started in the $\chi_1 = g+$ and $\chi_2 = t$ rotamers.

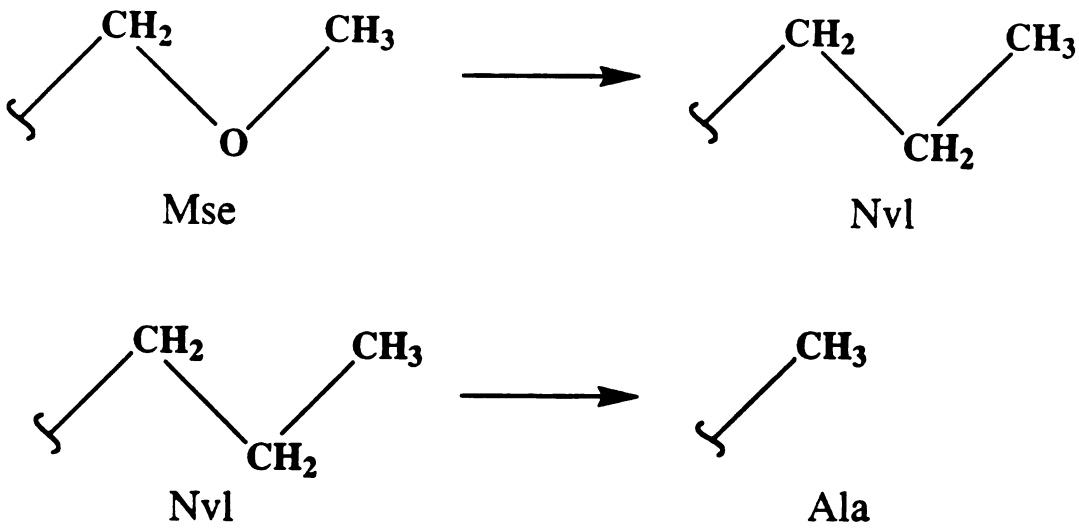



**Figure 1**

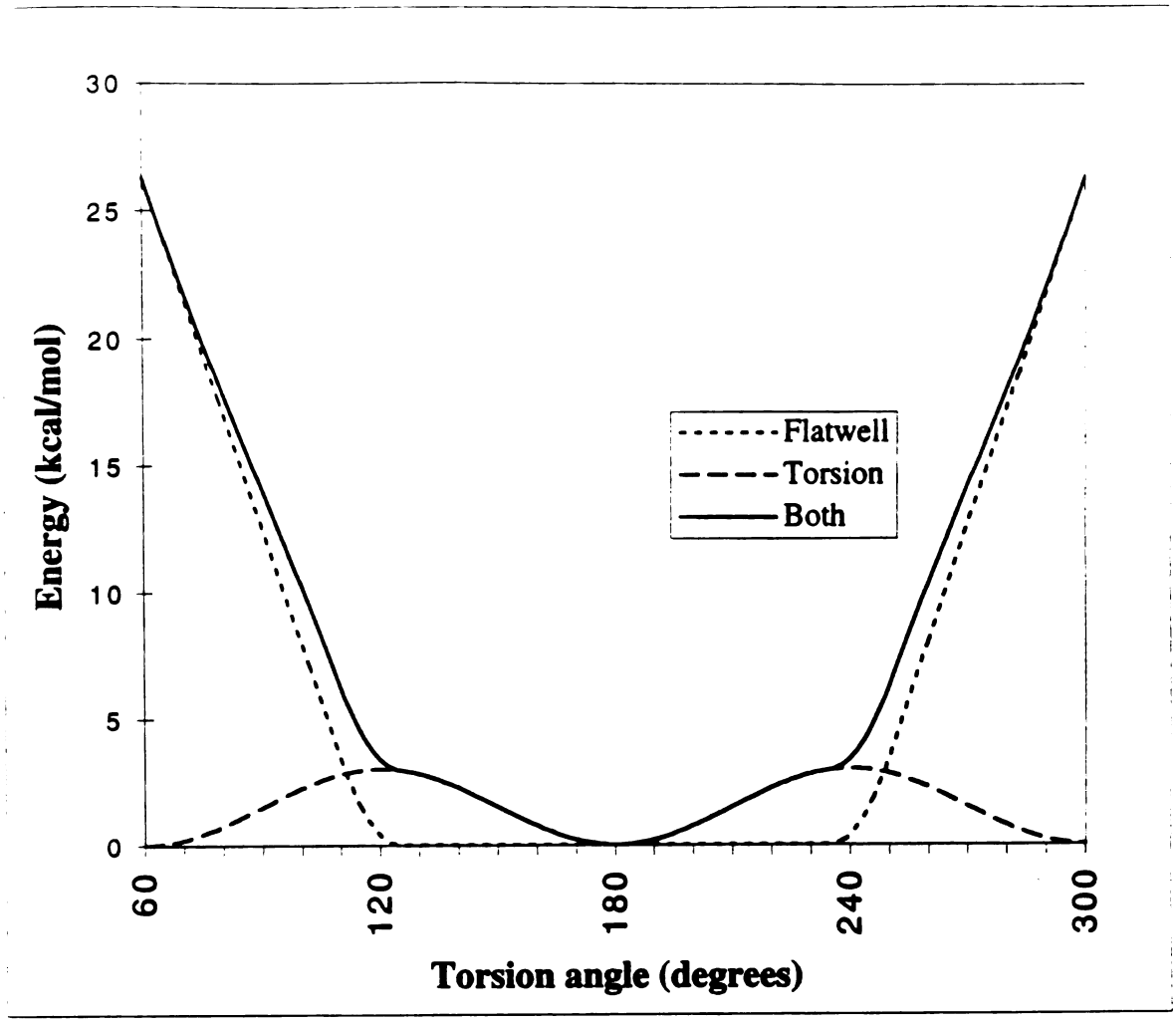

30
25
20
15
10
5
0
Energy (kcal/mol)
Flatwell
Torsion
Both
60
120
180
240
300
Torsion angle (degrees)

**Figure 2**

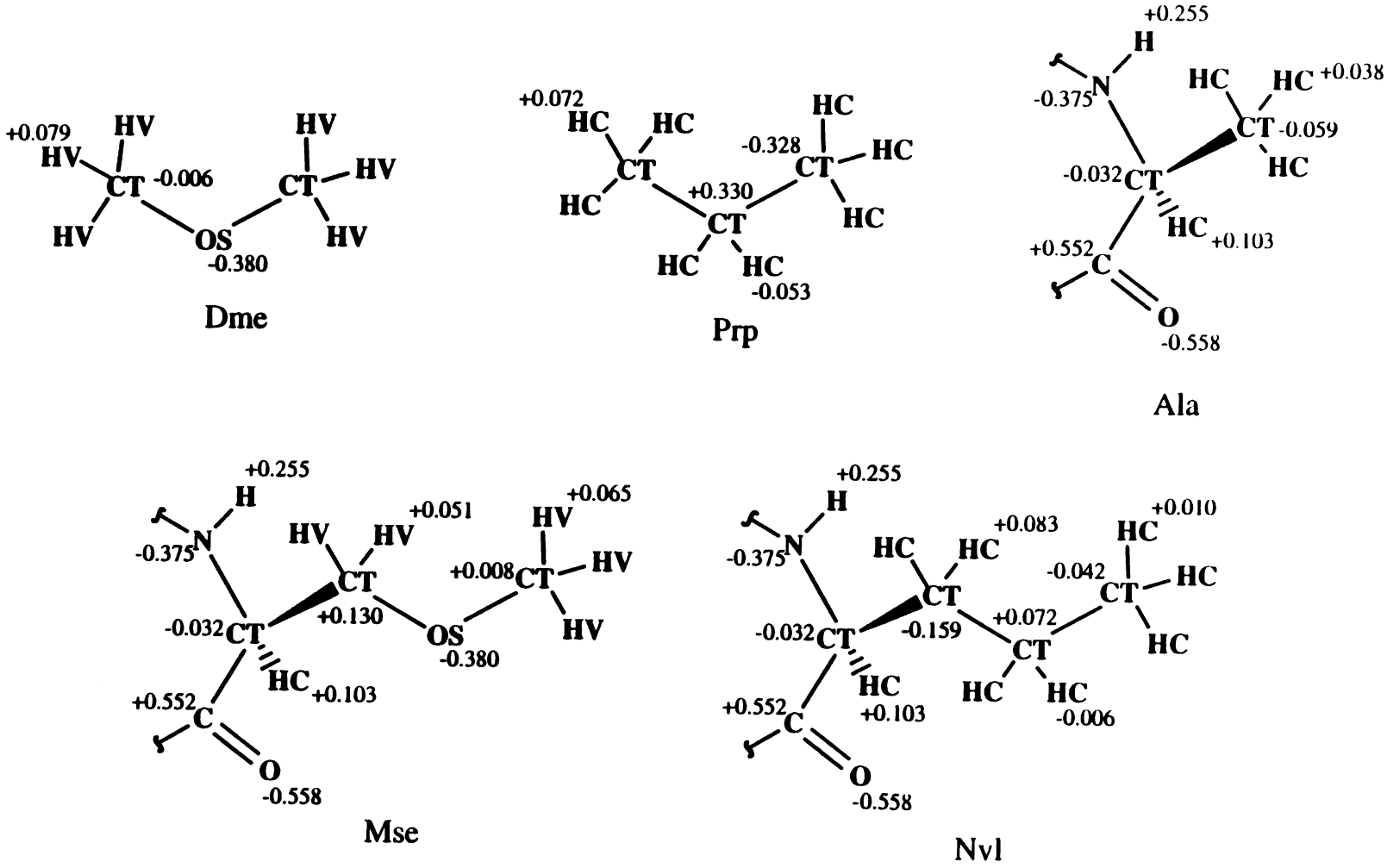

+0.079
HV
HV
CT
-0.006
HV
CT
HV
HV
OS
HV
-0.380
Dme
+0.072
HC
HC
CT
HC
-0.328
CT
HC
+0.330
CT
HC
HC
HC
-0.053
Prp
+0.255
H
N
-0.375
HC
HC
+0.038
CT
-0.059
HC
-0.032
CT
HC
+0.103
+0.552
C
O
-0.558
Ala
+0.255
H
N
-0.375
HV
HV
+0.051
CT
+0.130
-0.032
CT
HC
+0.103
+0.552
C
O
-0.558
OS
-0.380
+0.008
CT
+0.065
HV
HV
HV
Mse
+0.255
H
N
-0.375
HC
HC
+0.083
CT
-0.159
-0.032
CT
HC
+0.103
+0.552
C
O
-0.558
+0.072
CT
HC
HC
HC
-0.042
CT
+0.010
HC
HC
-0.006
Nvl

Figure 3

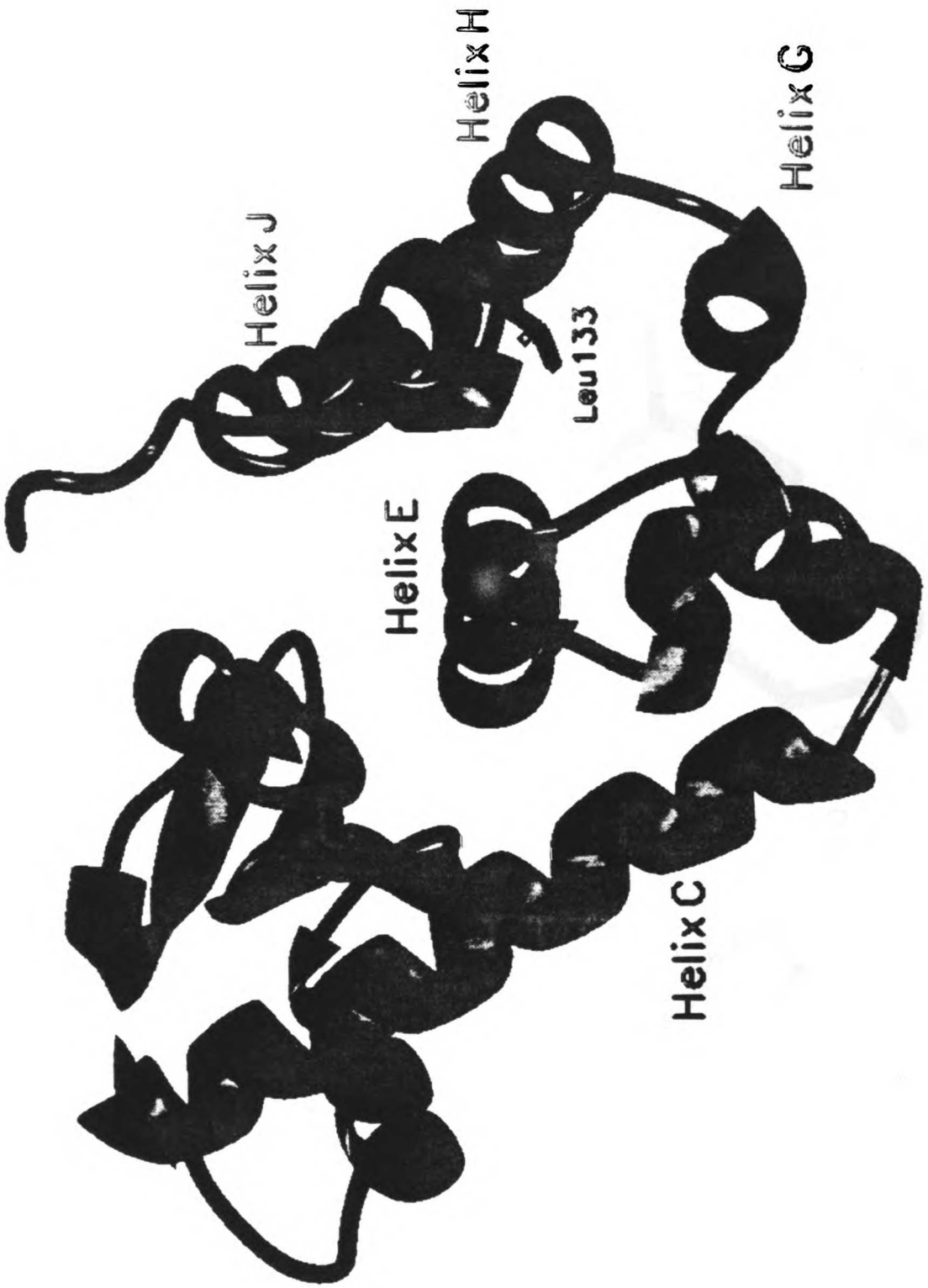
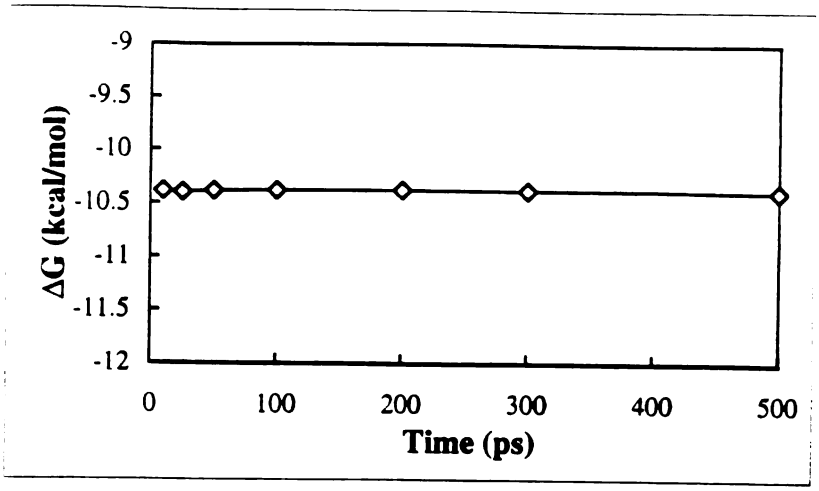



**Figure 4**

**Figure 5**


-9
-9.5
-10
-10.5
-11
-11.5
-12
ΔG (kcal/mol)
0
100
200
300
400
500
Time (ps)

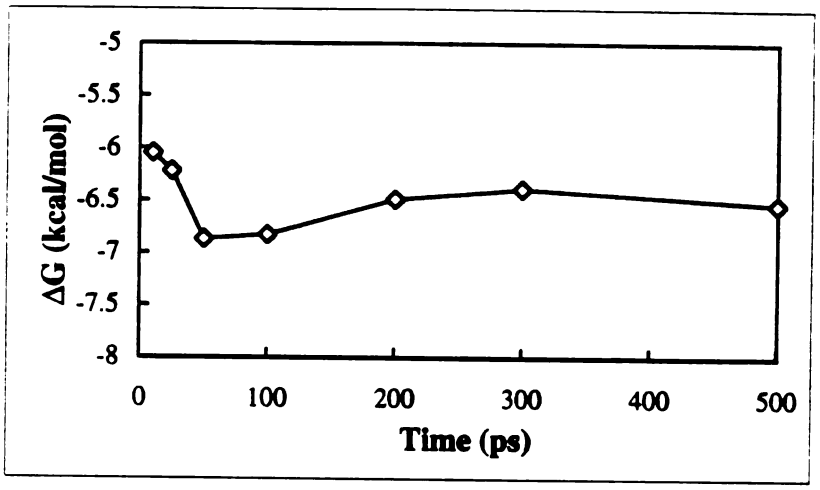
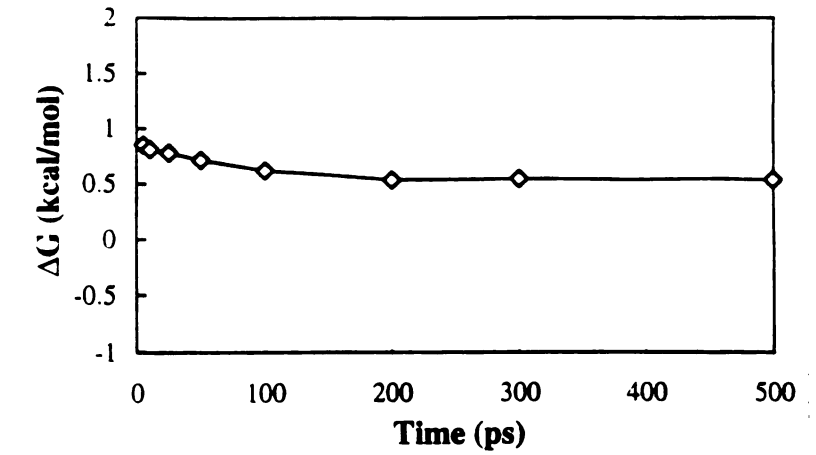
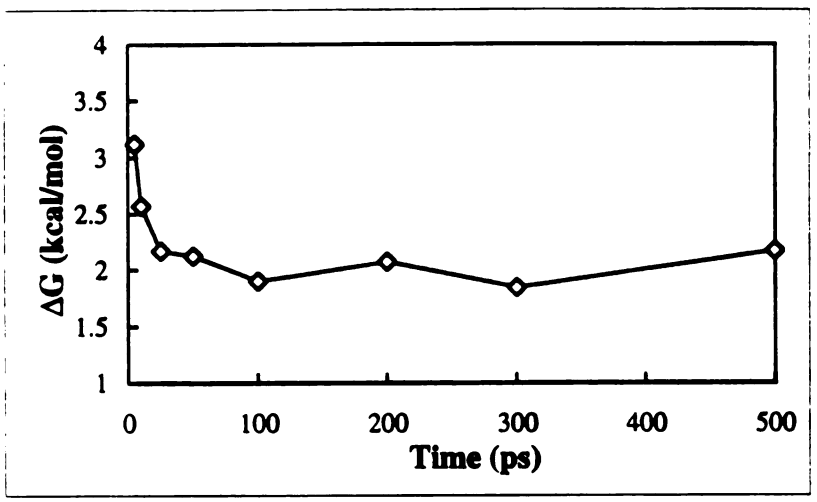

-5
-5.5
-6
-6.5
-7
-7.5
-8
ΔG (kcal/mol)
0
100
200
300
400
500
Time (ps)

**Figure 6a,b**


2
1.5
1
0.5
0
-0.5
-1
ΔG (kcal/mol)
0
100
200
300
400
500
Time (ps)


4
3.5
3
2.5
2
1.5
1
ΔG (kcal/mol)
0
100
200
300
400
500
Time (ps)

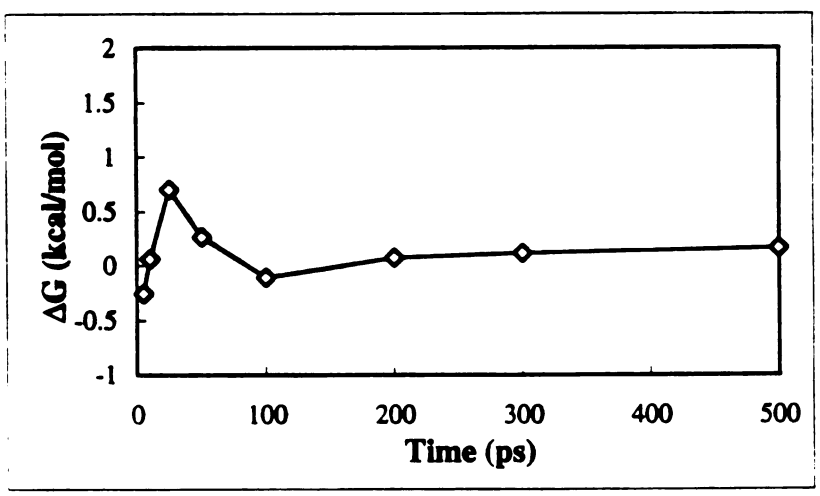
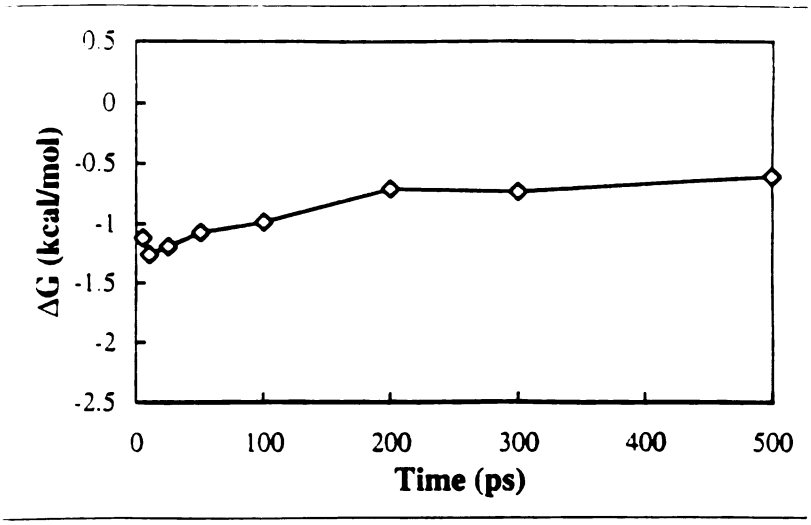
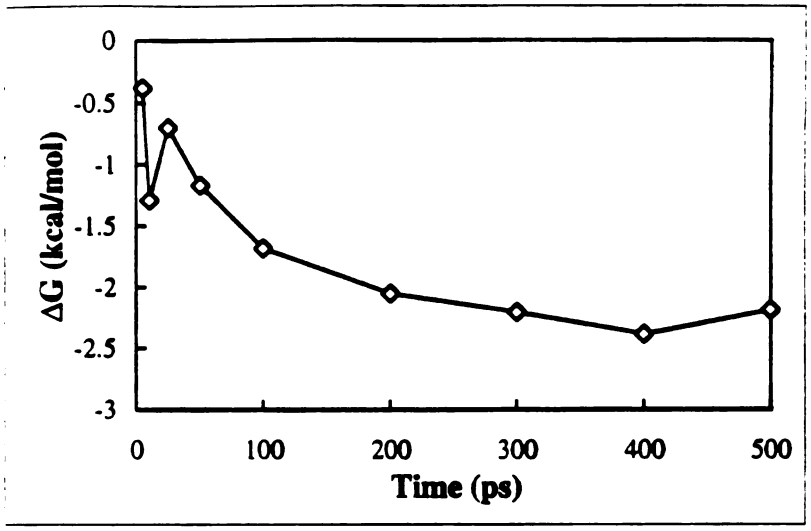

2
1.5
1
0.5
0
-0.5
-1
ΔG (kcal/mol)
0
100
200
300
400
500
Time (ps)

**Figure 6c,d,e**


0.5
0
-0.5
-1
-1.5
-2
-2.5
ΔG (kcal/mol)
0
100
200
300
400
500
Time (ps)


0
-0.5
-1
-1.5
-2
-2.5
-3
ΔG (kcal/mol)
0
100
200
300
400
500
Time (ps)

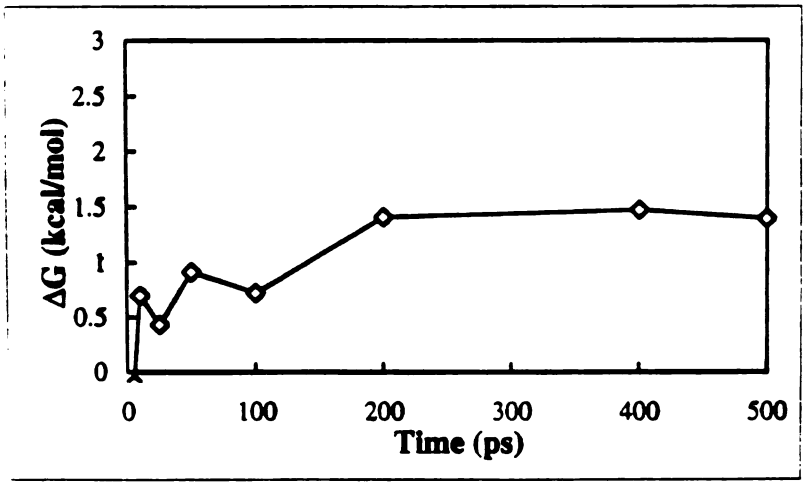
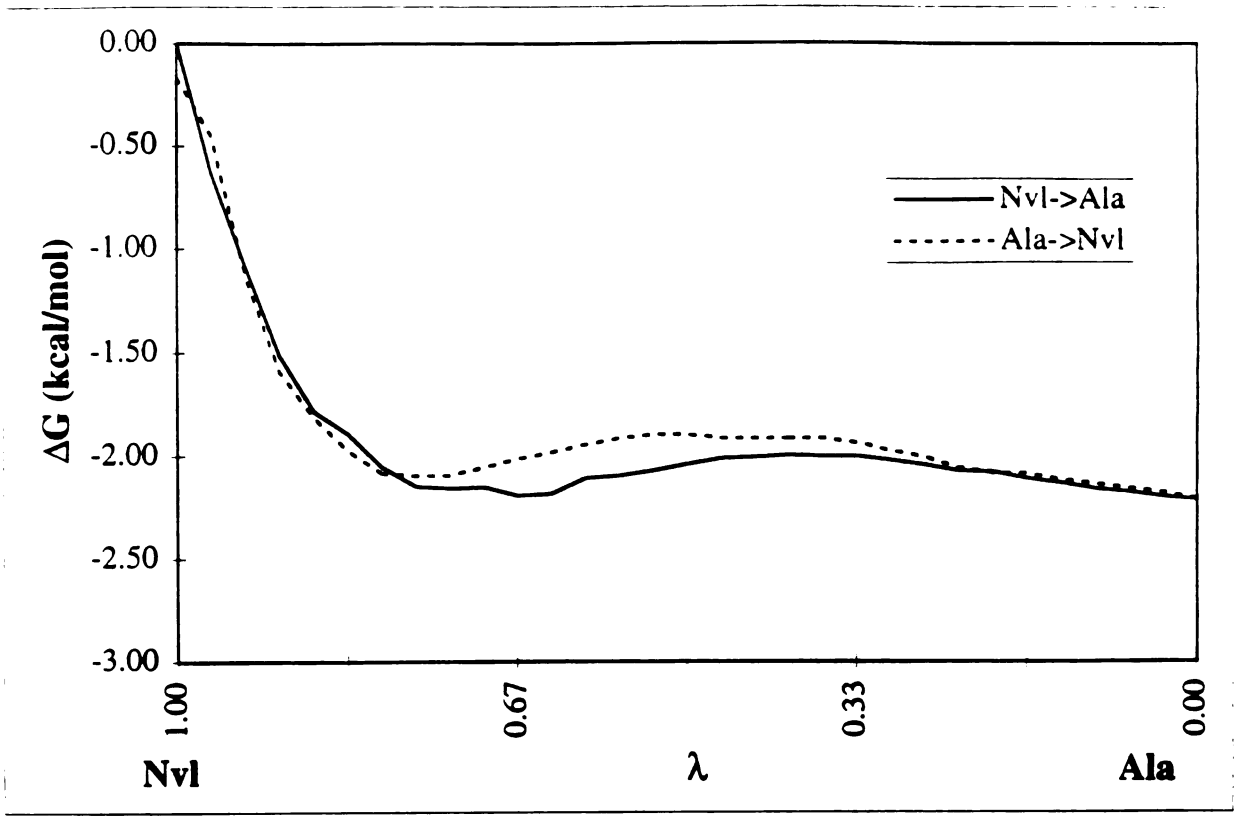

3
2.5
2
1.5
1
0.5
0
ΔG (kcal/mol)
0
100
200
300
400
500
Time (ps)

**Figure 6f,g,h**


0.00
-0.50
-1.00
-1.50
-2.00
-2.50
-3.00
ΔG (kcal/mol)
Nvl->Ala
Ala->Nvl
1.00
0.67
0.33
0.00
Nvl
λ
Ala

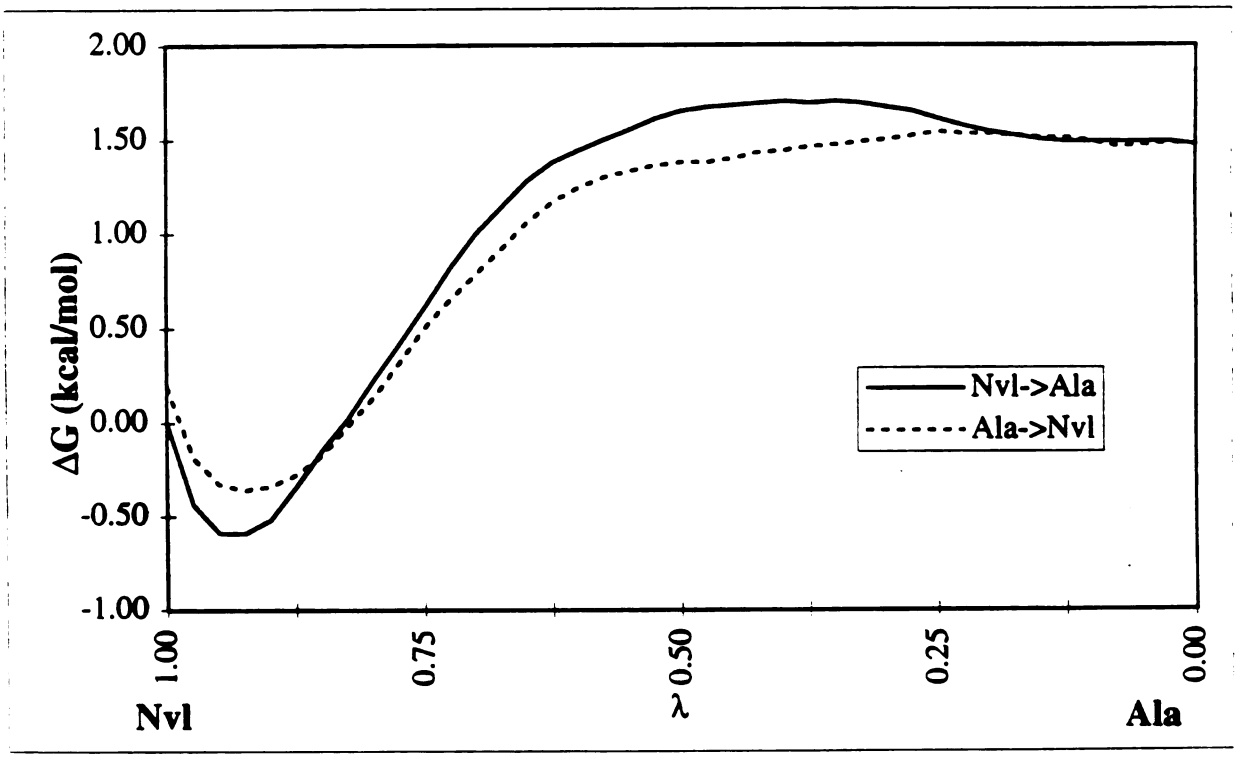
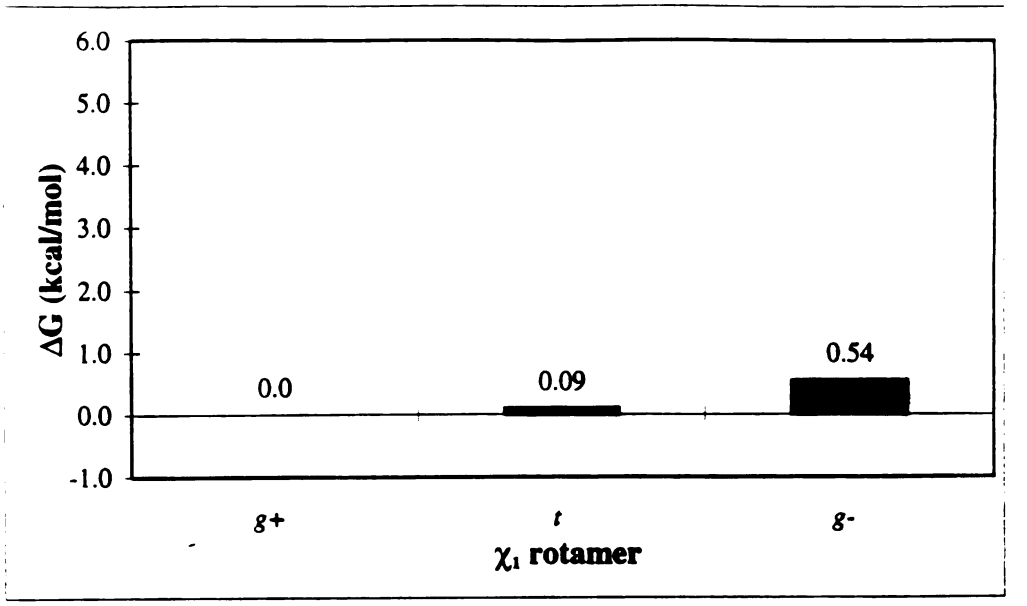

2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
ΔG (kcal/mol)
Nvl->Ala
Ala->Nvl
1.00
0.75
0.50
0.25
0.00
Nvl
λ
Ala

**Figure 7a,b**


6.0
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
ΔG (kcal/mol)
0.0
0.09
0.54
g+
t
g-
χ1 rotamer

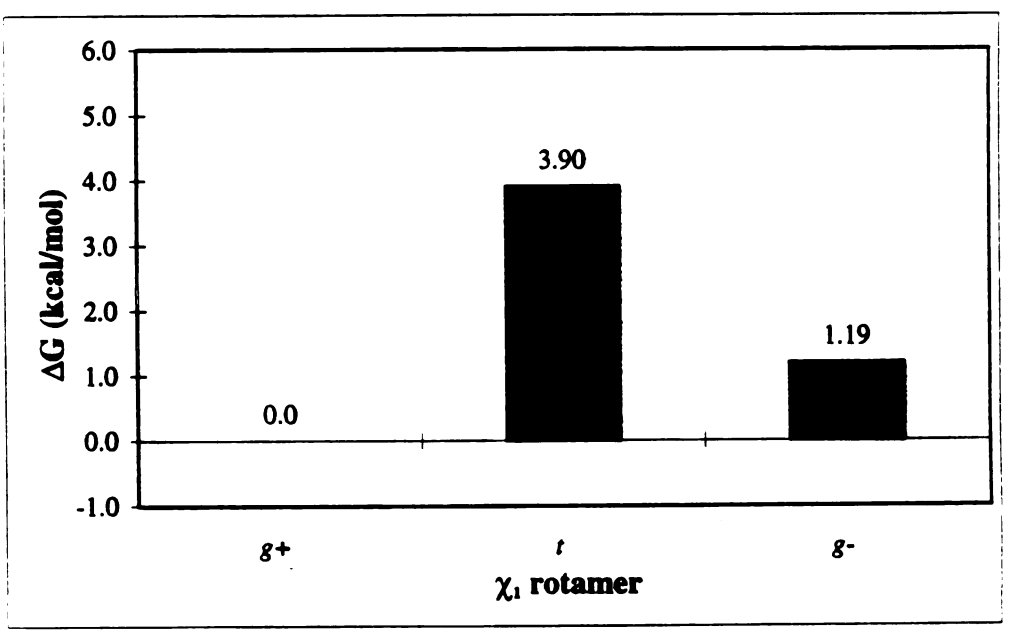

6.0
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
ΔG (kcal/mol)
0.0
3.90
1.19
g+
t
g-
χ1 rotamer

**Figure 8a,b**

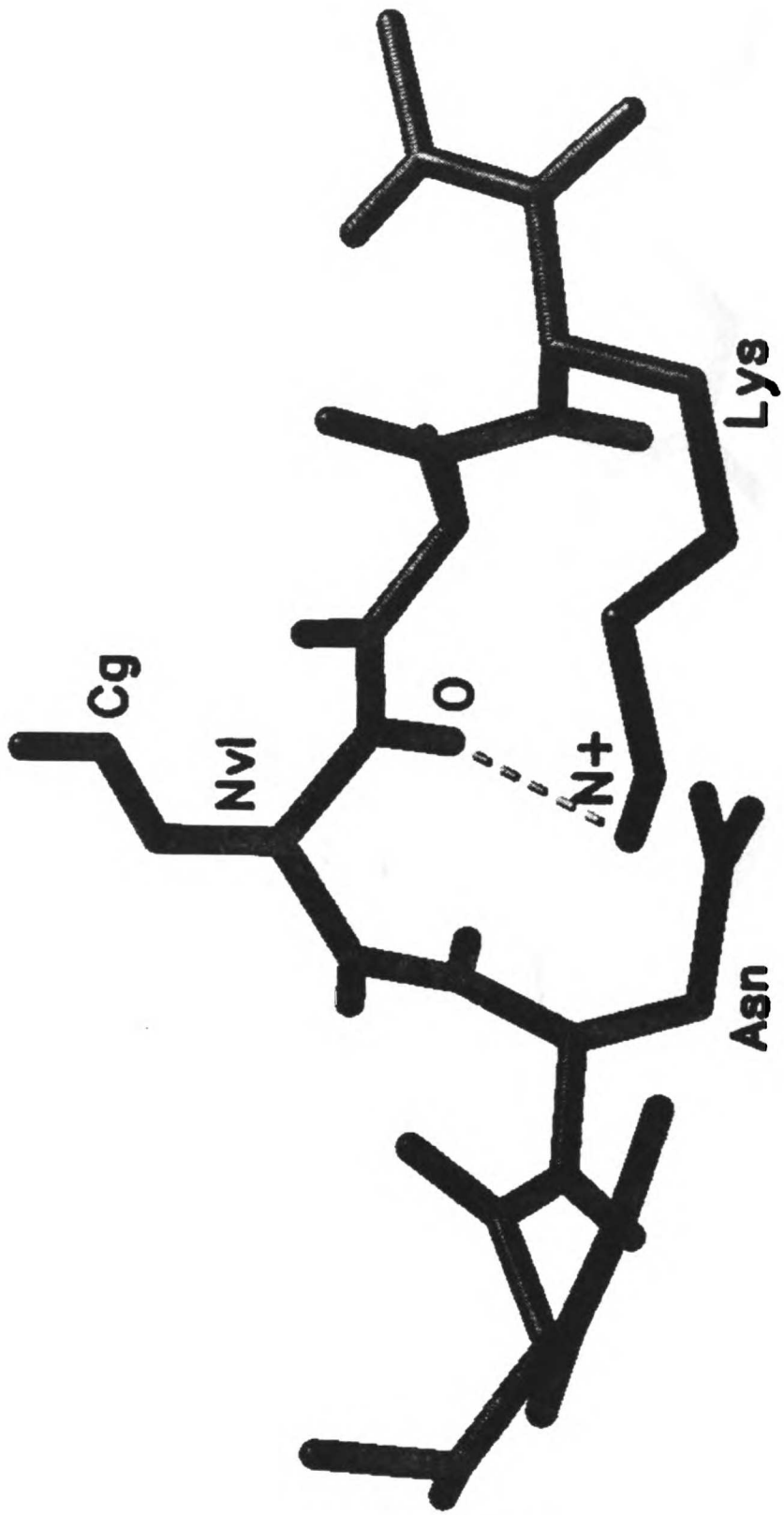



**Figure 9**

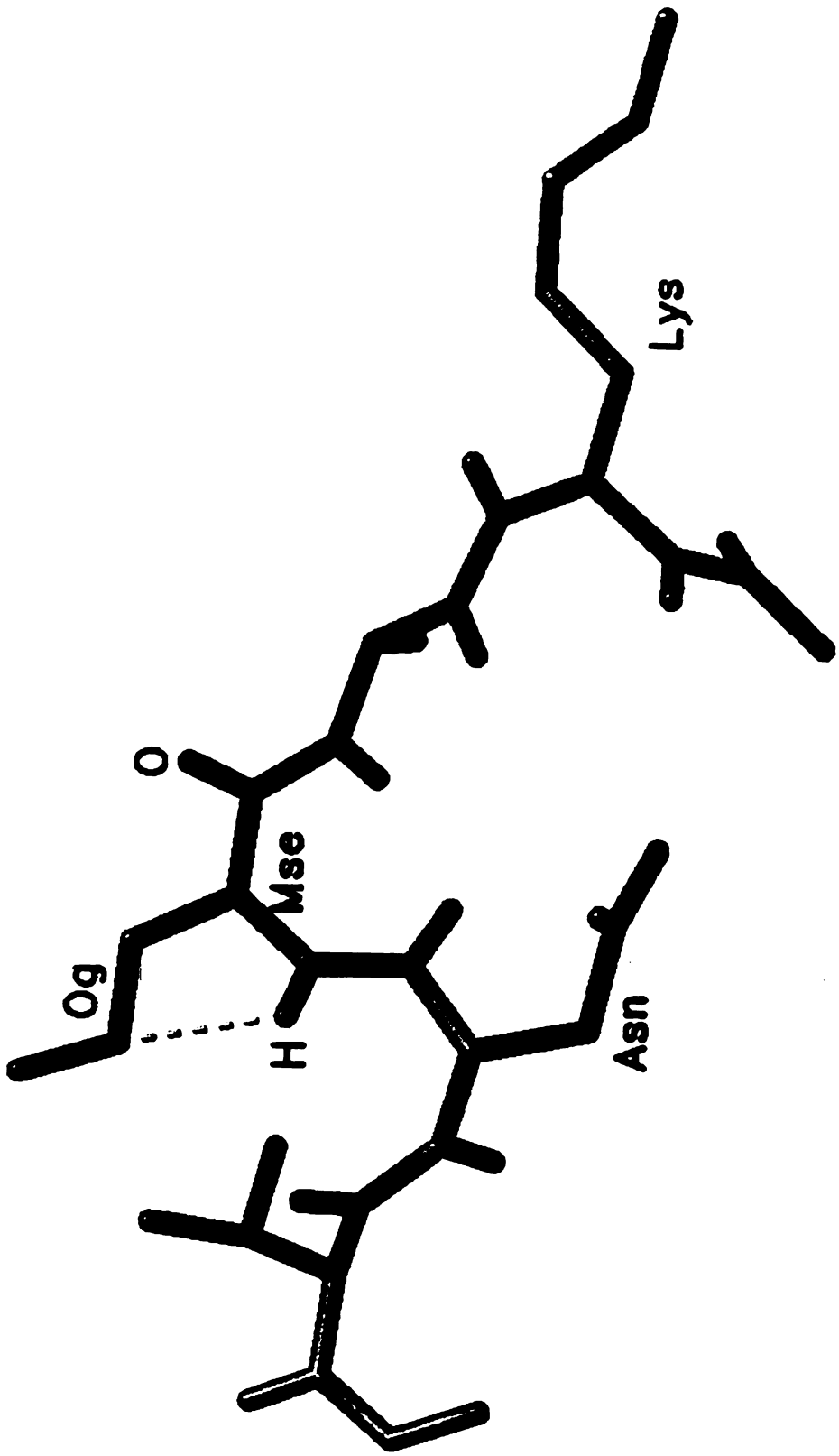



**Figure 10**

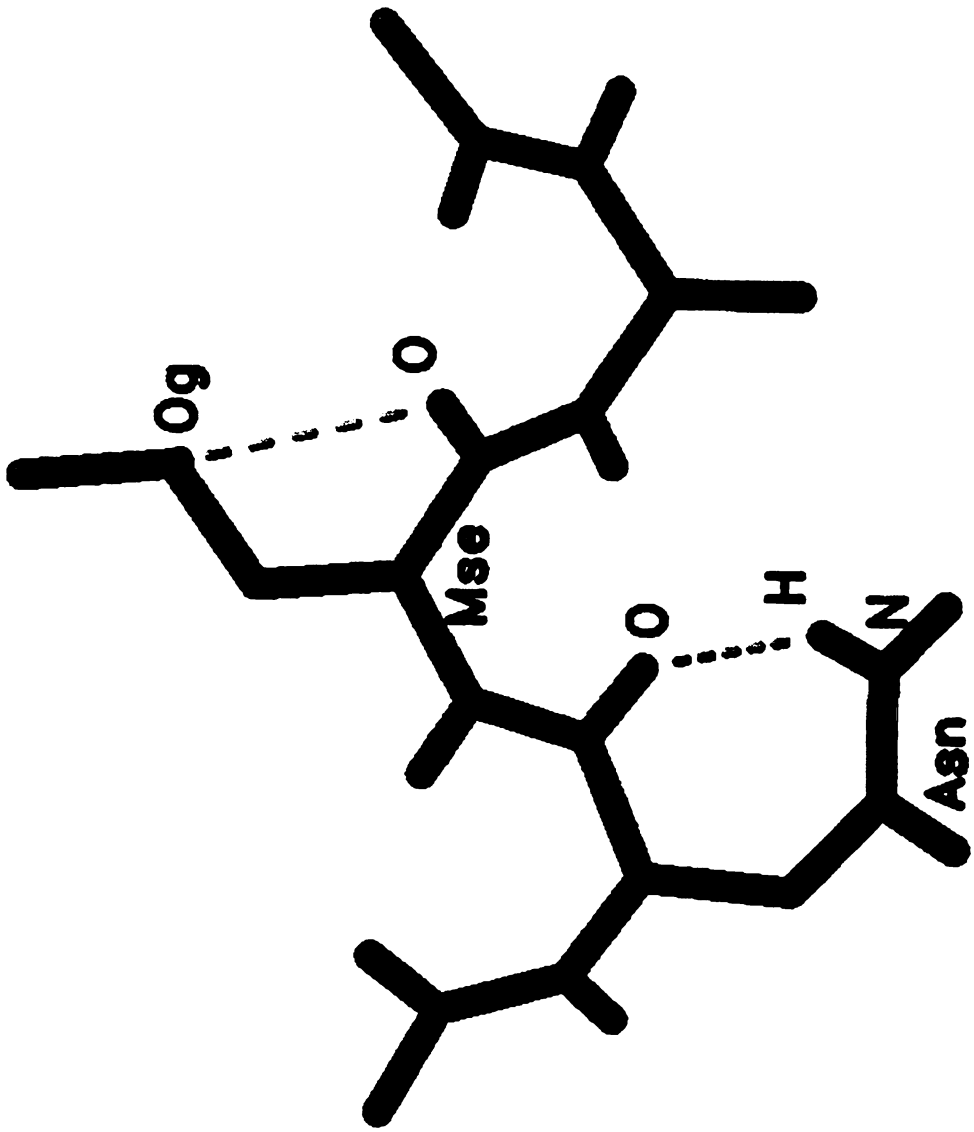



**Figure 11**

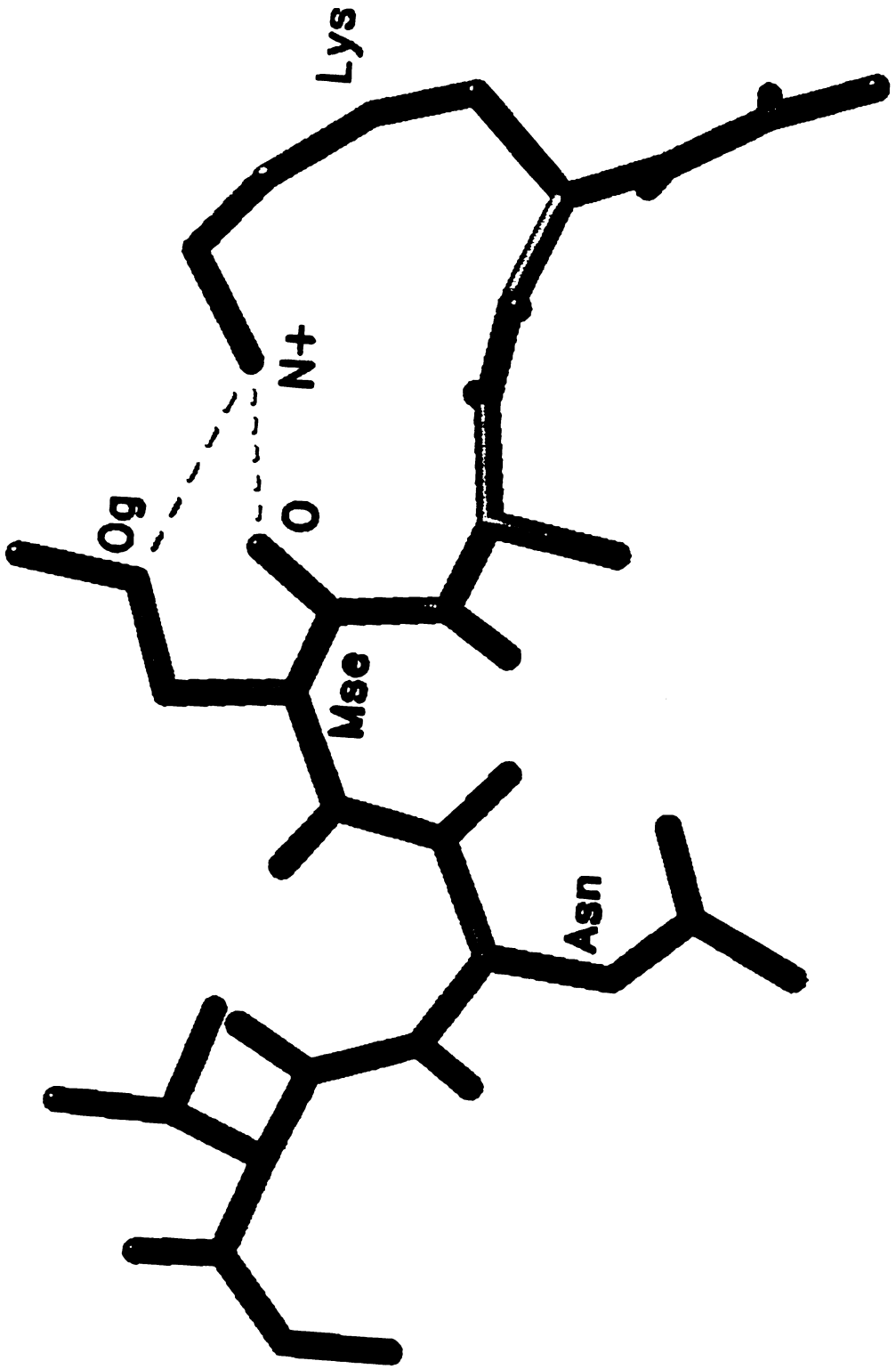



**Figure 12**

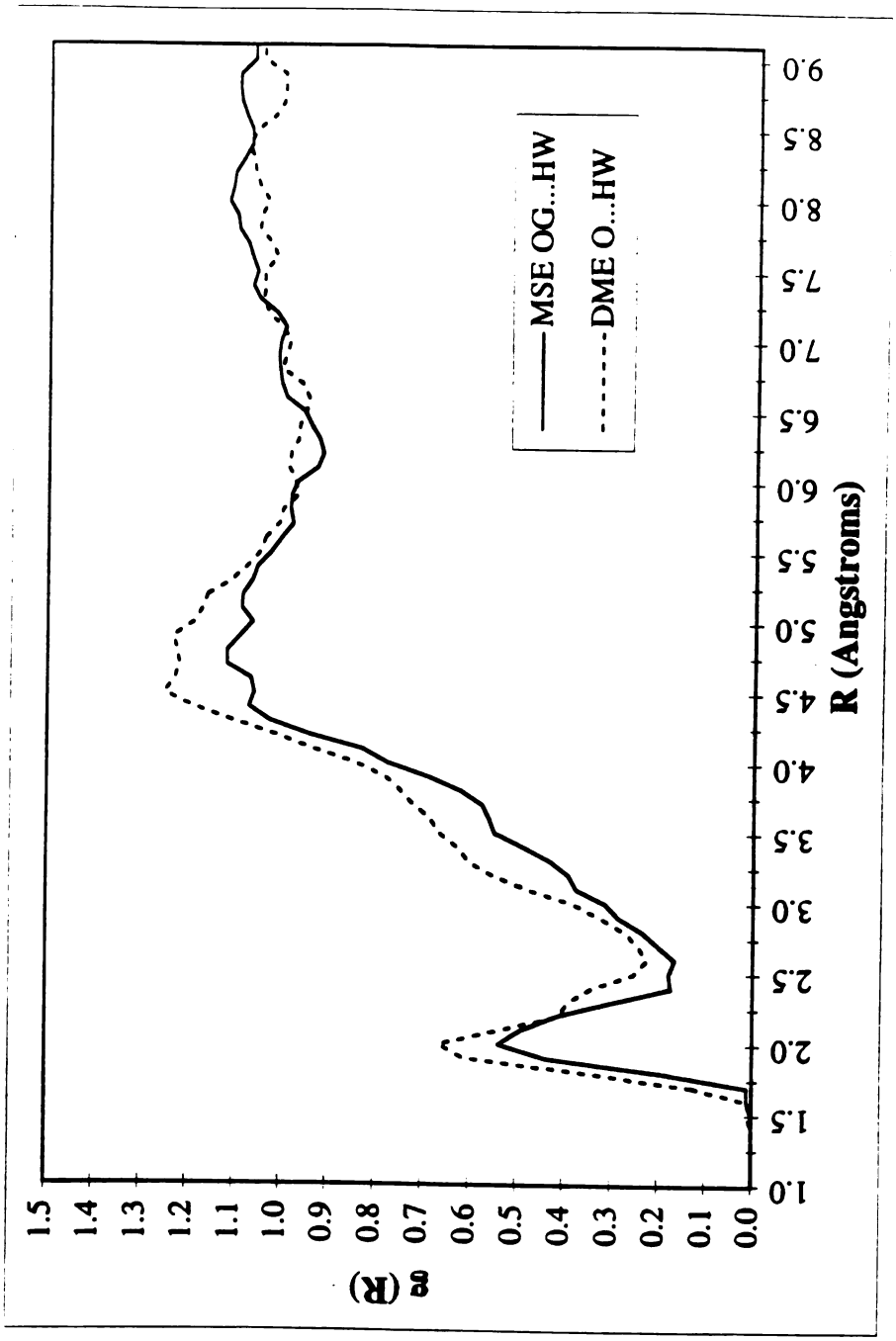

MSE OG...HW
DME O...HW
R (Angstroms)
g (R)
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5

**Figure 13**

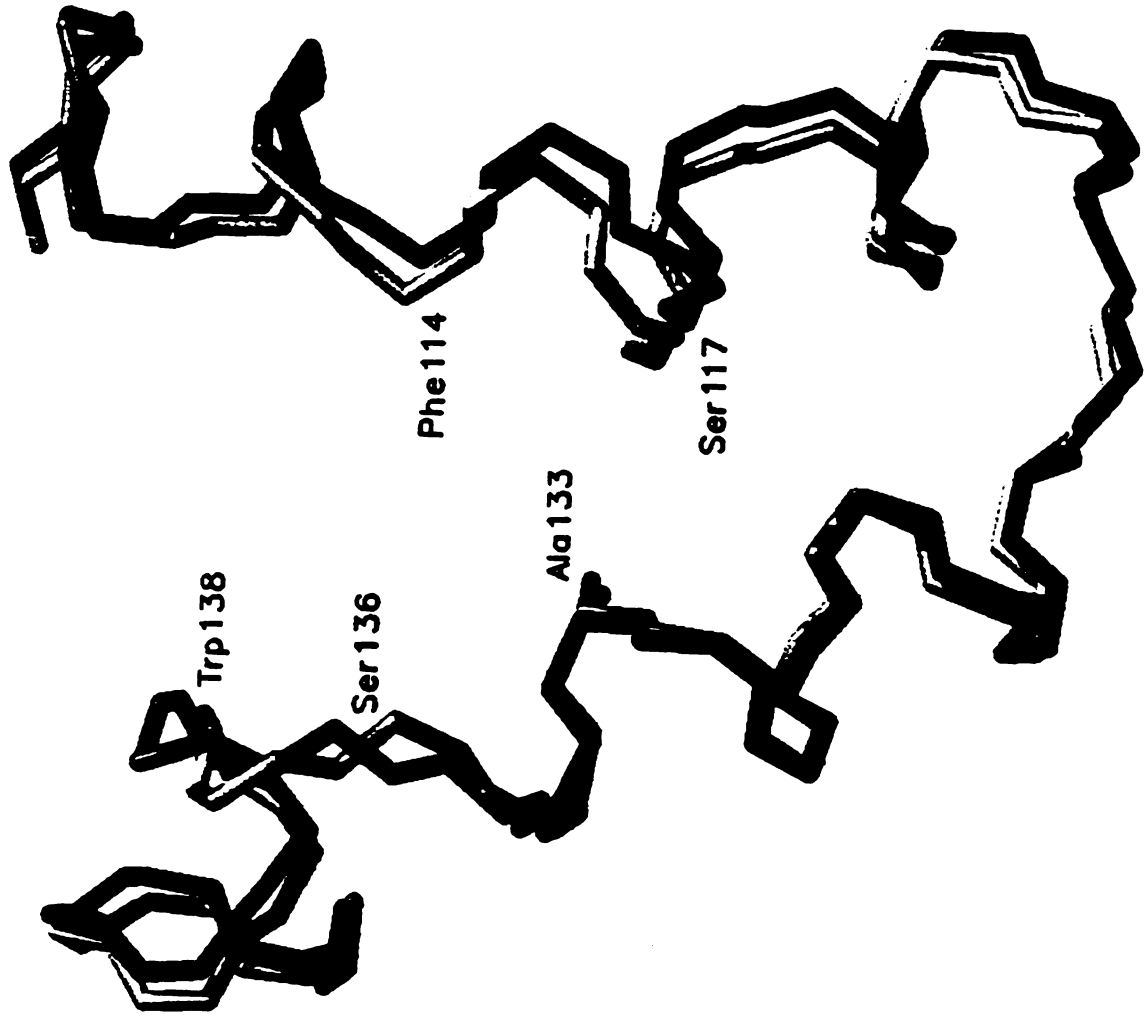
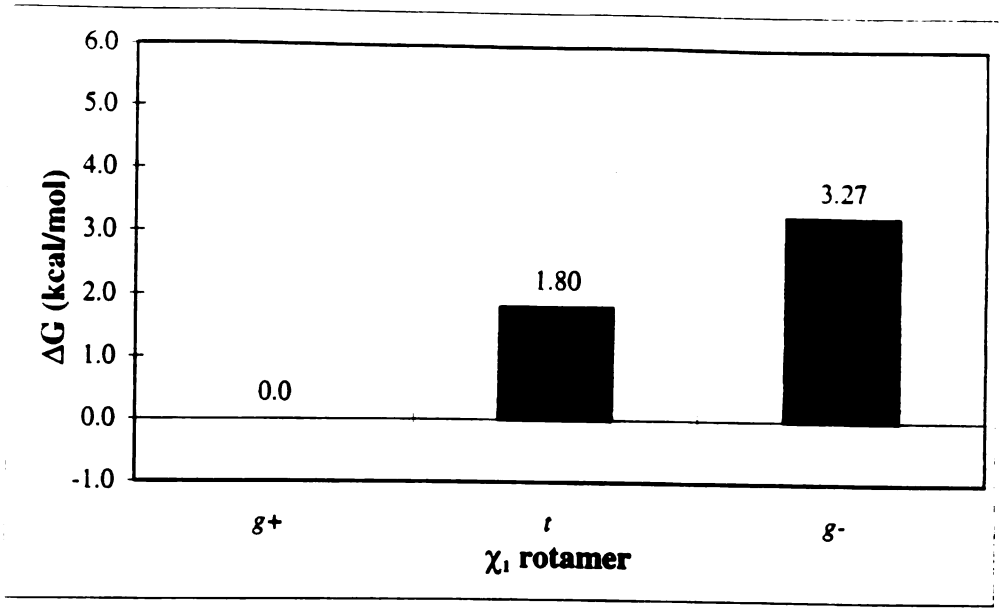



**Figure 14**


6.0
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
ΔG (kcal/mol)
0.0
1.80
3.27
g+
t
g-
χ1 rotamer

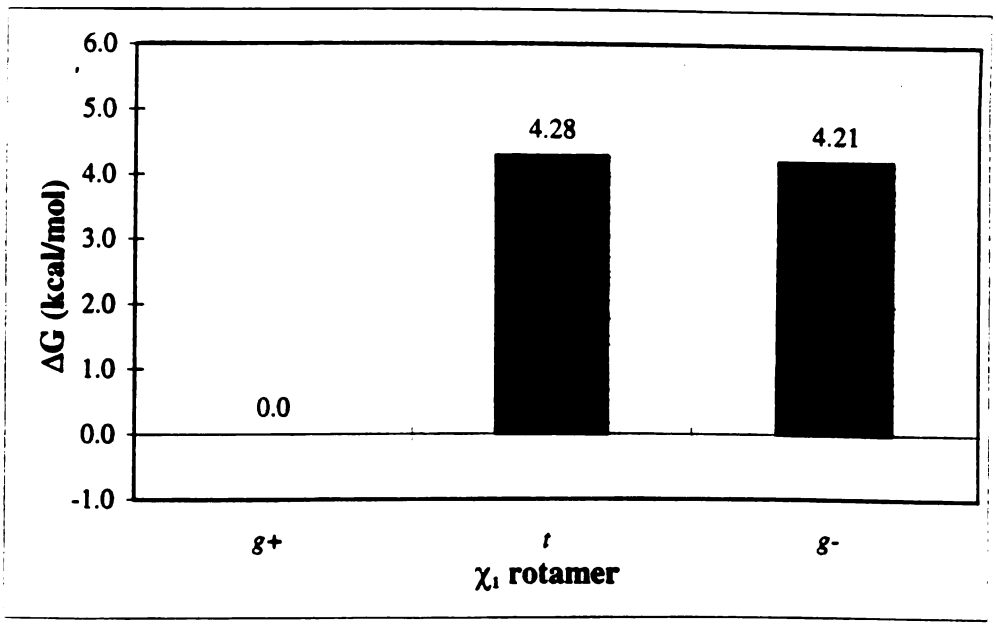

6.0
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
ΔG (kcal/mol)
0.0
4.28
4.21
g+
t
g-
χ1 rotamer

**Figure 15a,b**

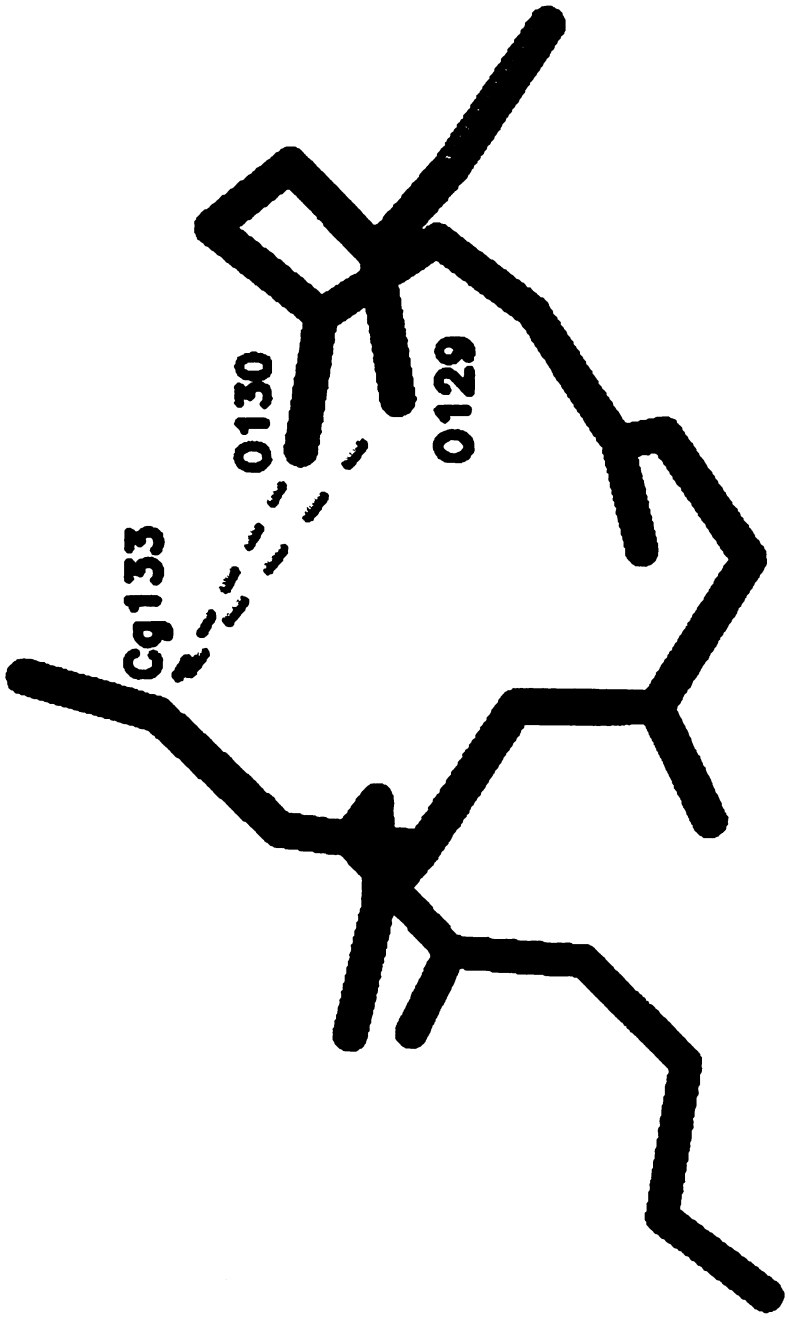



**Figure 16**

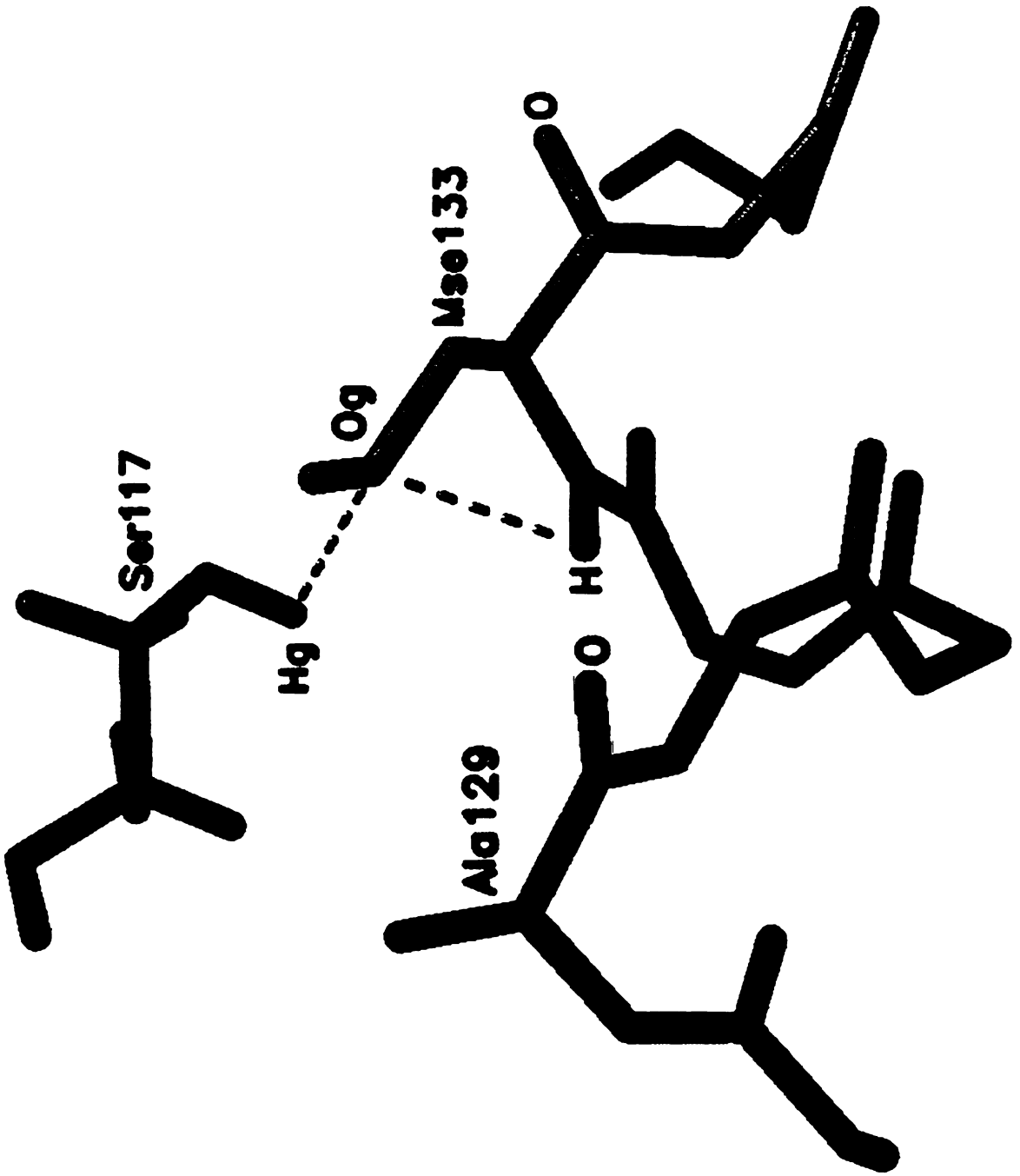



**Figure 17**

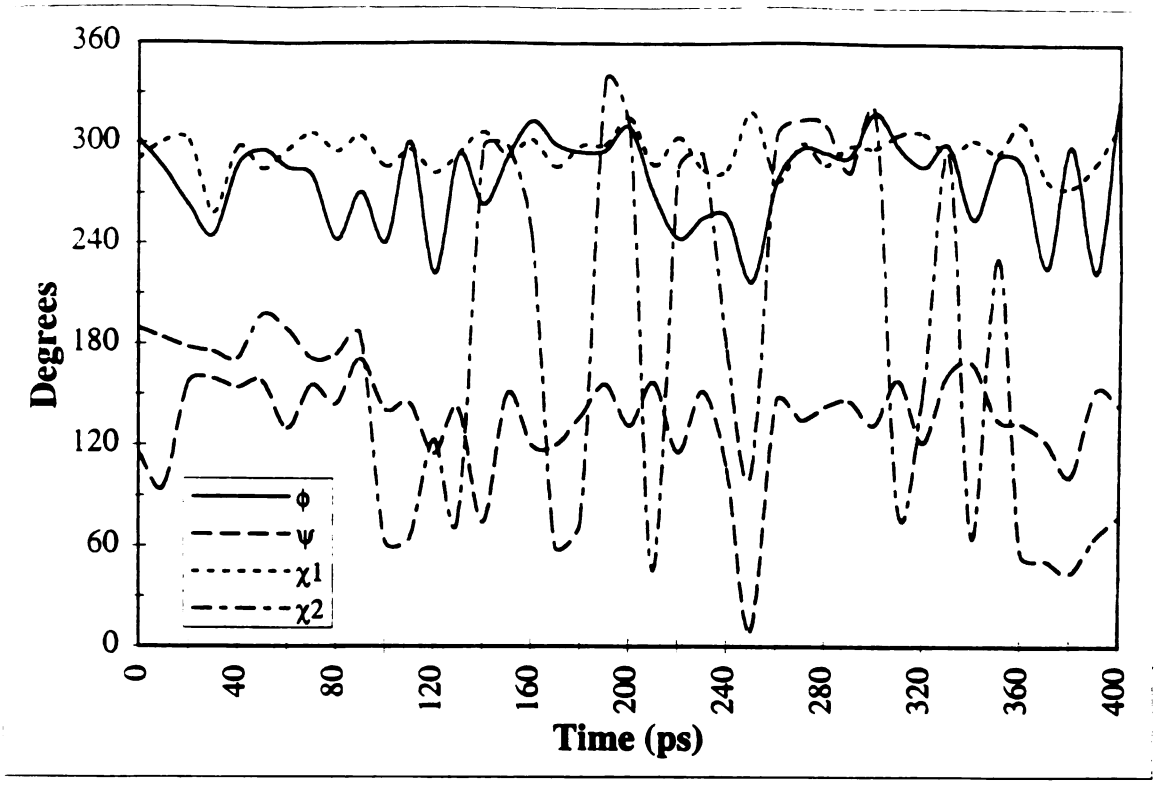

360
300
240
180
120
60
0
Degrees
0
40
80
120
160
200
240
280
320
360
400
Time (ps)
ϕ
ψ
χ1
χ2

**Figure 18a**

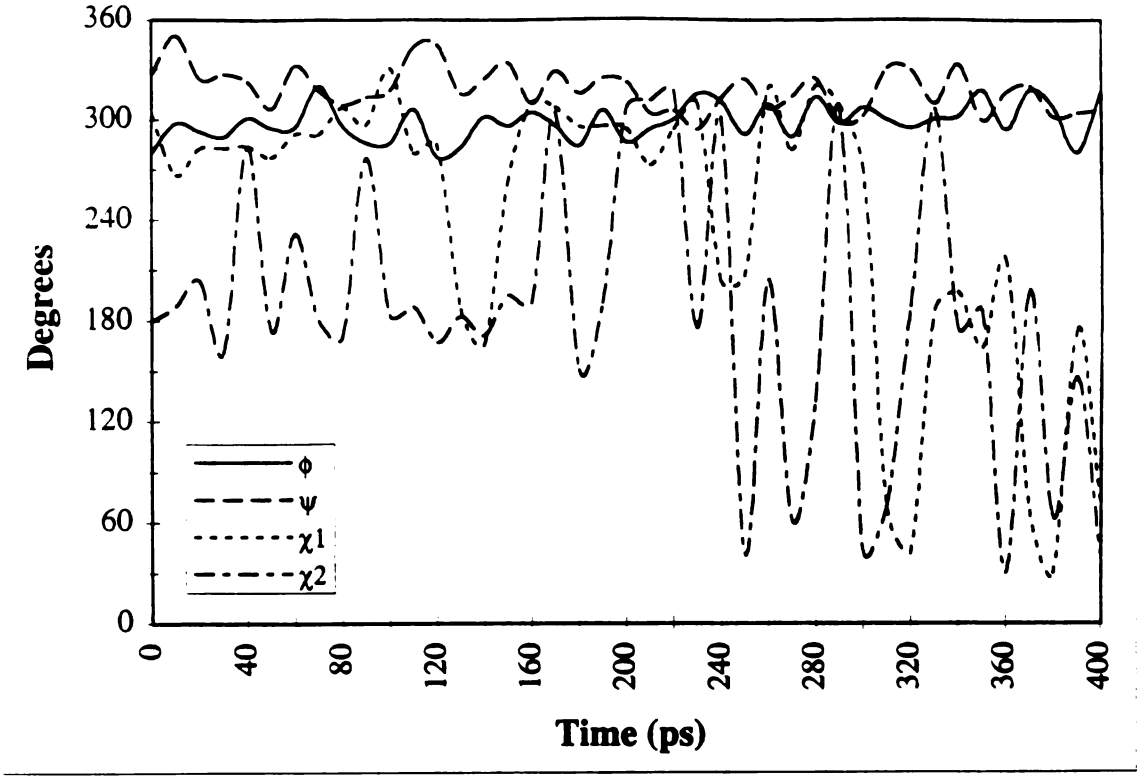

360
300
240
180
120
60
0
Degrees
0
40
80
120
160
200
240
280
320
360
400
Time (ps)
φ
ψ
χ1
χ2

**Figure 18b**

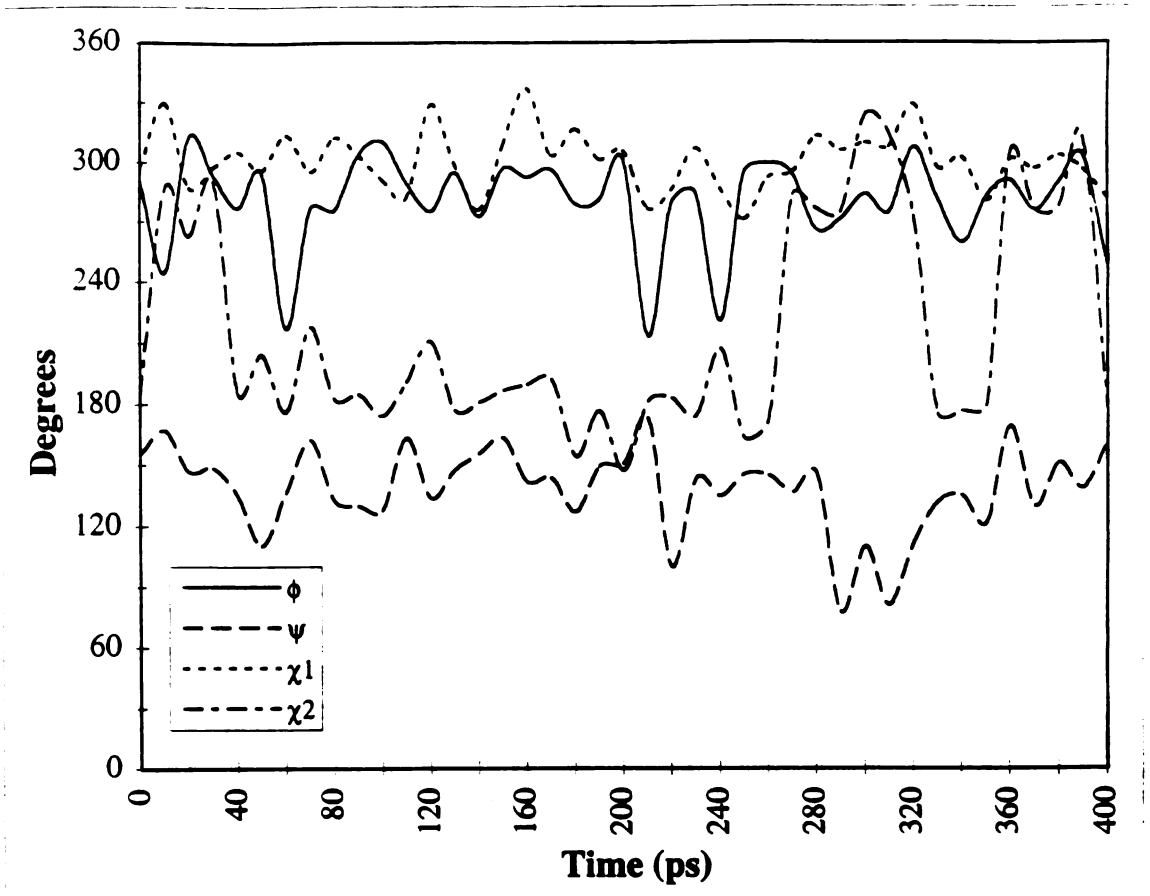

360
300
240
180
120
60
0
Degrees
0
40
80
120
160
200
240
280
320
360
400
Time (ps)
ϕ
ψ
χ1
χ2

**Figure 18c**

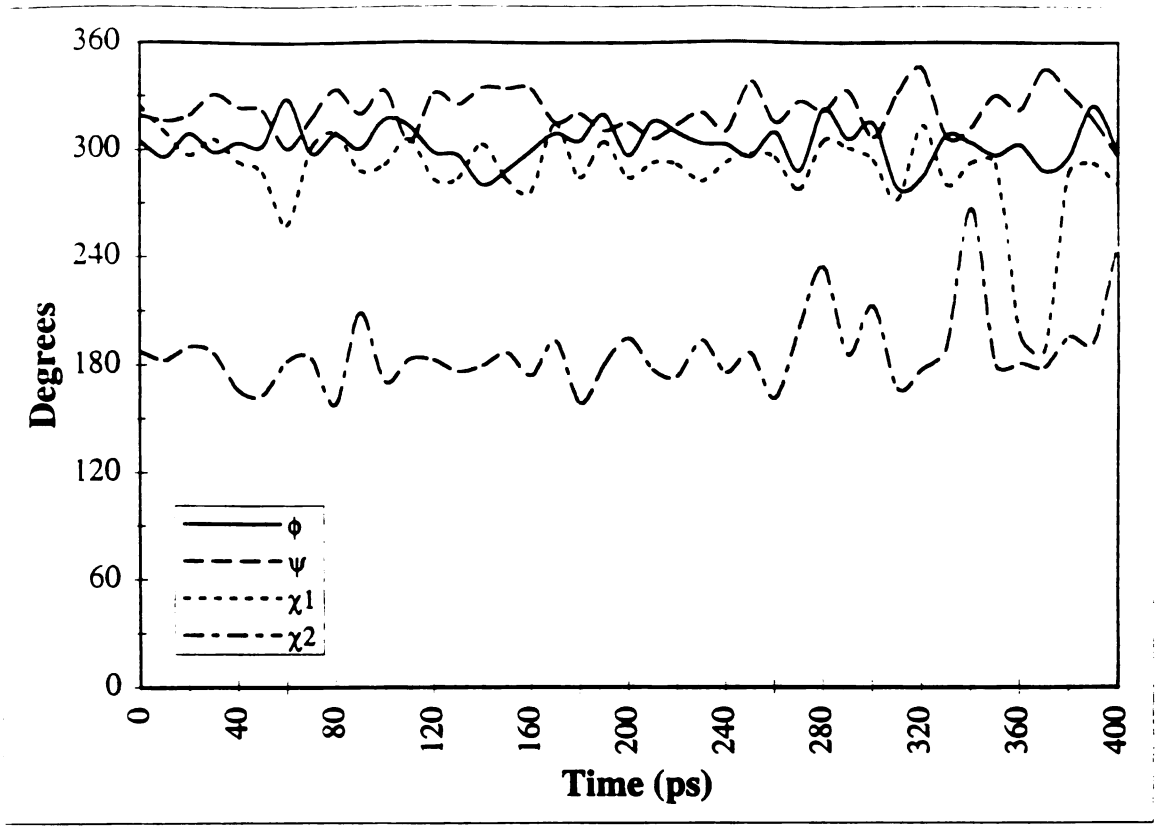

360
300
240
180
120
60
0
Degrees
0
40
80
120
160
200
240
280
320
360
400
Time (ps)
φ
ψ
χ1
χ2

**Figure 18d**

**Table I.** van der Waals parameters used in calculations.

| Molecule | Atom | Atom Type | r* | ε |
|---|---|---|---|---|
| $CH_3OCH_3$ | C | CT[a] | 1.908 | 0.1094 |
| | O | OS[b] | 1.650 | 0.1500 |
| | H | HV[c] | 1.387 | 0.0157 |
| $CH_3CH_2CH_3$ | C | CT[a] | 1.908 | 0.1094 |
| | H | HC[a] | 1.487 | 0.0157 |

a. Derived from liquid hydrocarbon studies (Sun *et al.*, 1992)

b. Obtained from Weiner *et al.* (1984) ribose ring oxygen; also uses 10-12 H-bonding terms

c. Derived from 6-31G* ab initio calculations (Veenstra *et al.*, 1992)

**Table II.** Unconstrained residues in constraint sets.[a]

| Constraint Set: | α-cons[b] | cavity-cons[c] |
|---|---|---|
| | 131-135 | 98, 99, 102 |
| | | 106, 111, 114 |
| | | 116-121 |
| | | 126-139 |
| | | 146-147 |
| | | 149-154 |

a. Two different constraint set were used for the Nvl->Ala mutation in the folded state.

b. Allows residues on either side of 133 in α-helix H to move.

c. Allows residues surrounding hydrophobic cavity to move.

**Table III.** Peptide Free Energy Changes (kcal/mol).

| | Solution | | | | *In Vacuo* | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\chi_1$ Rotamer: | *g*+ | *t* | *g*- | overall[a] | *g*+ | *t* | *g*- | overall[a] | $\Delta\Delta G_{unf-vac}$[b] |
| **Nvl→Ala**[c] | | | | | | | | | |
| Ace-**X**-Nme | -2.39 | -2.25 | -3.05 | -1.82 | -0.57 | -1.27 | -2.21 | -0.38 | -1.44 |
| Ace-Asn-**X**-Ala-Nme | -2.35 | -2.36 | -2.95 | -1.83 | -0.70 | -1.15 | -1.99 | -0.42 | -1.41 |
| Ace-Val-Asn-**X**-Ala-Lys-Nme | -2.71 | -3.11 | -3.07 | -2.28 | -0.88 | 0.01 | -0.44 | 0.33 | -2.61 |
| average: | -2.48 | -2.25 | -3.02 | **-1.98** | -0.72 | -0.80 | -1.547 | **-0.16** | **-1.82** |
| **Mse→Nvl**[d] | | | | | | | | | |
| Ace-**X**-Nme | 2.07 | -1.65 | 1.40 | 1.56 | 0.61 | -0.58 | 0.08 | 0.46 | 1.10 |
| Ace-Asn-**X**-Ala-Nme | 2.47 | -1.34 | 1.48 | 2.00 | 0.10 | -3.73 | -1.57 | -0.18 | 2.18 |
| Ace-Val-Asn-**X**-Ala-Lys-Nme | 2.31 | -1.59 | 2.01 | 2.05 | -0.41 | 1.64 | 0.98 | 1.41 | 0.64 |
| average: | 2.28 | -1.53 | 1.63 | **1.87** | 0.10 | -0.89 | -0.17 | **0.56** | **1.31** |

a. Calculated using equation (7).

b. Calculated free energy of solvation.

c. 400 ps simulations.

d. 200 ps simulations.

**Table IV.** Backbone conformations sampled during peptide perturbations.[a]

| | $\chi_1$ Rotamer | | |
|---|---|---|---|
| | *g*+ | *t* | *g*- |
| | $\phi,\psi$ | $\phi,\psi$ | $\phi,\psi$ |
| **Nvl→Ala** | | | |
| solution | | | |
| Ace-**X**-Nme | $P_{II}$ | $C_7^{eq}/C_5$ | $P_{II}$ |
| Ace-Asn-**X**-Ala-Nme | $C_7^{eq}$ | $C_7^{eq}/C_5$ | $C_7^{eq}$ |
| Ace-Val-Asn-**X**-Ala-Lys-Nme | $P_{II}/C_7^{eq}$ | $C_7^{eq}$ | $P_{II}$ |
| *in vacuo* | | | |
| Ace-**X**-Nme | $C_7^{eq}$ | $C_7^{eq}$ | $C_7^{eq}$ |
| Ace-Asn-**X**-Ala-Nme | $C_7^{eq}$ ($P_{II}$) | $C_7^{eq}$ ($P_{II}$) | $C_7^{eq}$ ($P_{II}$) |
| Ace-Val-Asn-**X**-Ala-Lys-Nme | $P_{II}$ | $P_{II}$ | $P_{II}$ |
| **Mse→Nvl** | | | |
| solution | | | |
| Ace-**X**-Nme | $C_7^{eq}$ ($P_{II}$) | $C_7^{eq}$ ($C_5$) | $P_{II}$ |
| Ace-Asn-**X**-Ala-Nme | $C_7^{eq}$ ($P_{II}$) | $C_7^{eq}$ ($P_{II}$) | $P_{II}$ |
| Ace-Val-Asn-**X**-Ala-Lys-Nme | $C_7^{eq}$ ($C_5$) | $P_{II}$ | $C_7^{eq}$ |
| *in vacuo* | | | |
| Ace-**X**-Nme | $C_7^{eq}$ | $C_5$ | $C_7^{eq}$ |
| Ace-Asn-**X**-Ala-Nme | $C_7^{eq}$ | $P_{II}$ | $P_{II}$ |
| Ace-Val-Asn-**X**-Ala-Lys-Nme | $P_{II}$ | $C_7^{eq}$ | $C_5$ |

a. $C_7^{eq}$ near $(\phi,\psi) = (-80°,70°)$; $C_5$ near $(\phi,\psi) = (-160°,170°)$; $P_{II}$ near $(\phi,\psi) = (-80°,150°)$. A slash (/) indicates both conformations sampled and parentheses indicate conformation sampled briefly.

**Table V.** Conformational free energies of $\chi_1$ rotamers relative to $\chi_1 = g+$ (kcal/mol).

| | Solution | | | | | *In Vacuo* | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\chi_1$ Rotamer: | *g*+ | *t* | *g*- | $q_{unc}$[a] | $\Delta G_t$[b] | *g*+ | *t* | *g*- | $q_{unc}$[a] | $\Delta G_t$[b] |
| **Nvl peptides** | | | | | | | | | | |
| Ace-Nvl-Nme | 0.0 | -0.14 | 0.66 | 2.60 | 0.57 | 0.0 | 0.70 | 1.64 | 1.38 | 0.19 |
| Ace-Asn-Nvl-Ala-Nme | 0.0 | 0.01 | 0.60 | 2.35 | 0.51 | 0.0 | 0.45 | 1.29 | 1.59 | 0.28 |
| Ace-Val-Asn-Nvl-Ala-Lys-Nme | 0.0 | 0.40 | 0.36 | 2.06 | 0.43 | 0.0 | -0.89 | -0.44 | 7.49 | 1.21 |
| average: | | 0.09 | 0.54 | 2.34 | 0.51 | | 0.09 | 0.83 | 3.49 | 0.75 |
| **Mse peptides** | | | | | | | | | | |
| Ace-Mse-Nme | 0.0 | 3.58 | 1.33 | 1.11 | 0.06 | 0.0 | 1.89 | 2.17 | 1.07 | 0.04 |
| Ace-Asn-Mse-Ala-Nme | 0.0 | 3.82 | 1.59 | 1.07 | 0.04 | 0.0 | 4.28 | 2.96 | 1.01 | 0.00 |
| Ace-Val-Asn-Mse-Ala-Lys-Nme | 0.0 | 4.30 | 0.66 | 1.33 | 0.17 | 0.0 | -2.94 | -1.83 | 156.0 | 3.03 |
| average: | | 3.90 | 1.19 | 1.17 | 0.09 | | 1.08 | 1.10 | 53.0 | 2.38 |
| **Nvl in protein**[c] | | | | | | | | | | |
| X-Nvl-X | 0.0 | 1.80 | 3.27 | 1.05 | 0.03 | | | | | |
| **Mse in protein**[d] | | | | | | | | | | |
| X-Mse-X | 0.0 | 4.28 | 4.21 | 1.00 | 0.00 | | | | | |

a. Calculated from eq. (5).

b. Calculated from eq. (6).

c. Calculation with cavity-cons (Table II).

d. Calculation with α-cons (Table II).

**Table VI.** Free energy changes in folded state (kcal/mol).

| | *g*+ | *t* | *g*- | overall[a] |
|---|---|---|---|---|
| Nvl →Ala, α-cons[b] | 1.86 | 0.46 | -0.97 | 1.92 |
| Nvl →Ala, cavity-cons[c] | 1.47 | -0.33 | -1.80 | **1.50** |
| $\Delta\Delta G_{packing}$[d] | -0.39 | -0.79 | -0.83 | -0.42 |
| Mse →Nvl, α-cons[b] | 0.07 | -2.41 | -0.87 | **0.03** |

a. Calculated using eq. (7). Bold values are those used to calculated $\Delta\Delta G_{unf\text{-}fold}$ (Table XI).

b. Residues 131-135 unrestrained (Table II).

c. Residues 131-135 and cavity residues unrestrained (Table II).

d. Free energy gained by allowing the backbone to adjust to mutation.

**Table VII.** Distances between Cβ of 133 and surrounding atoms for $\chi_1 = g+$ rotamer (Angstroms).[a]

| | 136 Cα | 138 Cα | 117 Cα | 114 Cα | 129 O | 130 O | 138 CZ3 |
|---|---|---|---|---|---|---|---|
| α-constraints | | | | | | | |
| NVL Cβ | 7.53 | 9.01 | 6.11 | 8.27 | 3.58 | 3.64 | 4.19 |
| ↓ | | | | | | | |
| ALA Cβ | 7.40 | 8.98 | 5.87 | 8.07 | 3.53 | 3.68 | 4.28 |
| cavity-constraints | | | | | | | |
| NVL Cβ | 6.70 | 7.60 | 6.45 | 7.90 | 4.23 | 3.37 | 4.73 |
| ↓ | | | | | | | |
| ALA Cβ | 6.44 | 7.10 | 5.47 | 7.25 | 3.66 | 3.50 | 4.02 |
| ↓ | | | | | | | |
| NVL Cβ | 6.61 | 7.33 | 6.26 | 7.37 | 4.27 | 3.40 | 4.26 |
| X-ray: L133 Cβ | 7.47 | 8.87 | 6.28 | 8.33 | 3.48 | 3.86 | 4.28 |
| A133 Cβ | 7.08 | 8.62 | 5.73 | 7.24 | 3.65 | 4.51 | 4.23 |
| $\Delta_{X\text{-}ray}$: | -0.39 | -0.25 | -0.55 | -1.09 | 0.17 | 0.65 | -0.05 |

a. The inter-residue distances between residue 133 and surrounding residues change little using the α-constraints (Table II). However, with the cavity-constraints (Table II), the inter-residue distances decrease on the same order of magnitude as seen in the difference between the Leu and Ala X-ray structures.

**Table VIII.** RMSD's of heavy atoms compared to L133 X-ray crystal structure.[a]

| Structure ($\chi_1 = g+$ ) | All Residues | Cavity Residues[b] | Backbone Atoms of Cavity Residues |
|---|---|---|---|
| α-constraints[c] | | | |
| Nvl | 1.08 | 1.02 | 0.33 |
| ↓ | | | |
| Ala | 1.27 | 1.39 | 0.40 |
| Mse | 1.09 | 1.01 | 0.32 |
| ↓ | | | |
| Nvl | 1.20 | 1.22 | 0.35 |
| cavity-constraints[d] | | | |
| Nvl | 1.24 | 1.38 | 0.68 |
| ↓ | | | |
| Ala | 1.40 | 1.73 | 1.05 |
| ↓ | | | |
| Nvl | 1.36 | 1.68 | 1.18 |

a. In Angstroms. Calculated for all non-hydrogen atoms from 20 ps data collection runs.
b. Residues used in cavity constraint set (Table II).
c. Simultaion with α-constraints (Table II).
d. Simulation with cavity-constraints (Table II).

**Table IX.** Folded state average dihedral angles of beginning and endpoint structures (degrees).[a]

| Structure | $\chi_1$ Rotamer | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *g+* | | | | *t* | | | | *g-* | | | |
| | $\phi$ | $\psi$ | $\chi_1$ | $\chi_2$ | $\phi$ | $\psi$ | $\chi_1$ | $\chi_2$ | $\phi$ | $\psi$ | $\chi_1$ | $\chi_2$ |
| **X-ray crystal structure** | | | | | | | | | | | | |
| Leu 133 | -65 | -30 | -80 | 161 | - | - | - | - | - | - | - | - |
| Ala 133 | -70 | -19 | - | - | - | - | - | - | - | - | - | - |
| **Nvl →Ala** | | | | | | | | | | | | |
| α-constraints | | | | | | | | | | | | |
| Nvl | -58 | -41 | -80 | 177 | -68 | -45 | -155 | 173 | -75 | -30 | 74 | -155 |
| ↓ | | | | | | | | | | | | |
| Ala | -56 | -35 | - | - | -66 | -37 | - | - | -82 | -14 | - | - |
| cavity-constraints | | | | | | | | | | | | |
| Nvl | -62 | -35 | -58 | 178 | -68 | -37 | -134 | -166 | -70 | -19 | 62 | -176 |
| ↓ | | | | | | | | | | | | |
| Ala | -58 | -42 | - | - | -72 | -41 | - | - | -72 | -36 | - | - |
| ↓ | | | | | | | | | | | | |
| Nvl | -54 | -52 | -62 | 175 | - | - | - | - | - | - | - | - |
| **Mse→Nvl** | | | | | | | | | | | | |
| α-constraints | | | | | | | | | | | | |
| Mse | -57 | -38 | -63 | 180 | -62 | -39 | -146 | 171 | -62 | -31 | 39 | -169 |
| ↓ | | | | | | | | | | | | |
| Nvl | -56 | -40 | -76 | 178 | -57 | -32 | -159 | 169 | -69 | -17 | 58 | -172 |

a. Calculated from 20 ps data collection runs.

**Table X.** Interatomic distances for $\chi_1$ rotamers of Nvl and Mse in folded state (Angstroms).

| | 133 HN | 133 O | 117 Hγ | 129 O | 130 O | 138 CZ3 |
|---|---|---|---|---|---|---|
| Side Chain Atom | | | | | | |
| X-ray: Leu Cγ ($\chi_1 = g+$ ) | - | 4.48 | - | 3.36 | 4.87 | 4.78 |
| Nvl Cγ[a] | | | | | | |
| $\chi_1 = g+$ | 2.88 | 4.55 | 3.98 | 3.68 | 4.01 | 5.32 |
| $\chi_1 = t$ | 3.97 | 3.34 | 4.58 | 4.62 | 5.49 | 3.81 |
| $\chi_1 = g-$ | 2.93 | 3.81 | 5.00 | 3.40 | 3.68 | 4.28 |
| Mse Oγ[b] | | | | | | |
| $\chi_1 = g+$ | 2.65 | 4.44 | 3.65 | 3.11 | 4.69 | 5.13 |
| $\chi_1 = t$ | 3.74 | 3.58 | 3.89 | 4.45 | 5.06 | 3.83 |
| $\chi_1 = g-$ | 2.41 | 4.42 | 4.58 | 3.11 | 3.85 | 5.09 |

a. Determined from 20 ps data collection run with cavity-constraints.

b. Determined from 20 ps data collection run with α-constrains.

**Table XI.** Calculated stability changes (kcal/mol).

| Mutation | $\Delta G_{vac}$ | $\Delta G_{unf}$ | $\Delta G_{fold}$ | $\Delta\Delta G_{unf\text{-}vac}$ [b] | $\Delta\Delta G_{fold\text{-}vac}$ [c] | $\Delta\Delta G_{unfolding}$ [a] | |
|---|---|---|---|---|---|---|---|
| | | | | | | $\Delta\Delta G_{calc,unf\ fold}$ [d] | $\Delta\Delta G_{expt}$ [e] |
| Nvl → Ala | -0.16 | -1.98 | 1.50 | -1.82 | 1.66 | **-3.48** | **-2.2 to -3.6** |
| Mse → Nvl | 0.56 | 1.87 | 0.03 | 1.31 | -0.53 | **1.84** | **1.56** |

a. Stability difference is relative to *unfolding*.

b. Calculated solvation free energy.

c. Difference gives effect of packing changes.

d. Calculated stability change.

e. Experimental stability change (Eriksson *et al.*, 1992; Medel *et al.*, 1992)

**Table XII.** $\Delta G_{conf}$ upon unfolding for $\chi_1$ (kcal/mol).

| Side Chain | $q_{unr.unf}$ [a] | $q_{unr.fold}$ [b] | $\Delta G_{conf}$ [c] |
|---|---|---|---|
| Nvl | 2.34 | 1.05 | -0.48 |
| Mse | 1.17 | 1.00 | -0.09 |

a. Partition function for $\chi_1$ in unfolded state, calculated from eq. (5) using average of peptides.

b. Partition function for $\chi_1$ in folded state, calculated from eq. (5).

c. Calculated from eq. (8), $\Delta G_{conf}$ represents the change in free energy due to the gain in conformational freedom of $\chi_1$ upon unfolding.

**Table XIII.** Contributions to $\Delta\Delta G$ for the $\chi_1 = g+$ rotamer (kcal/mol).[a]

| | Elec | vdW | Intra[b] | PMF[c] | Total |
|---|---|---|---|---|---|
| Nvl → Ala | | | | | |
| $\Delta\Delta G_{unf\text{-}vac}$ | -0.28 | -1.07 | -0.03 | -0.43 | -1.81 |
| $\Delta\Delta G_{fold\text{-}vac}$ | -0.58 | 3.09 | -0.05 | -0.41 | 2.05 |
| $\Delta\Delta G_{unf\text{-}fold}$ | 0.30 | -4.16 | 0.02 | -0.02 | -3.86 |
| Mse→Nvl | | | | | |
| $\Delta\Delta G_{unf\text{-}vac}$ | 0.03 | 1.59 | -0.21 | 0.08 | 1.49 |
| $\Delta\Delta G_{fold\text{-}vac}$ | -1.48 | 1.15 | -0.18 | 0.03 | -0.48 |
| $\Delta\Delta G_{unf\text{-}fold}$ | 1.51 | 0.44 | -0.03 | 0.05 | 1.97 |

a. Results for unfolded state are for dipeptide only.

b. Contributions from bonds, angles, and dihedrals.

c. Potential of mean force correction for changes in bond length (Pearlman and Kollman, 1991)

**Table XIV.** Free energy changes for simulations without side chain restraints (kcal/mol).[a]

| Mutation | $\Delta G_{unf}$[b] | $\Delta G_{fold}$ | $\Delta\Delta G_{unf\text{-}fold}$ |
|---|---|---|---|
| Nvl→ Ala | -2.36 | 1.00 | -3.36 |
| Mse→ Nvl | 1.74 | -0.01 | 1.75 |

a. Started in *g+, t* rotamer.

b. Results for dipeptide in solution.

# Chapter 5

# Conclusions and Future Work

## Conclusions on Modeling Protein Stability

We find ourselves at a perplexing time for molecular modeling of biological systems. Despite vast increases in computational power and experimental data, the systems that we can accurately model are in general only somewhat larger or more complex than a decade ago. The rate of progress in molecular modeling has been limited mainly by sampling issues, as well as the accuracy of force fields. It is clear that now is the time for careful analysis of the methodologies employed and improvement of them.

Encouraging progress has been made in the development of the AMBER force field in the past several years. The work presented in Chapter 2 on the refinement of vdW parameters for nonpolar hydrogens has contributed to the improvements seen with the Cornell *et al.* (1995) force field. The practice of reducing the r* hydrogen vdW parameter depending on the electronegativity of nearby atoms allows for a systematic assignment of vdW parameters on a sound physical basis. As seen in Chapter 4, the use of these parameters in our calculations has led to improved agreement with experiment.

The analysis of protein mutants has been a useful role for molecular dynamics simulations. The prediction of structural changes, though, and assessment of changes seen in simulations can be difficult. We have found that this task is simplified if there are a series of mutations which can be compared. In Chapter 3, correlations with the X-ray structures and comparison to the destabilizing F133 mutant aided the design of the unnatural amino acids. A comparison of different side chains was also useful in the analysis of β-branched unnatural amino acid side chains in T4 lysozyme in another collaborative project with the Schultz group (Cornish *et al.*, 1994). Systems in which there is experimental data for several mutants are thus ideal for study with simulation methods.

Calculations of protein stability, one of the greatest challenges for computational chemistry, have been presented in Chapter 4. Despite the negative evaluation of these types of calculations by Yun-yu *et al.* (1993), we have shown that these calculations are viable. We showed that simulations in proteins can be sampled to convergence, albeit in a fairly rigid system. We explored side conformational sampling and found that forced sampling did not change our final results much, but did allow us to determine the relative free energies of the $\chi_1$ rotamers and the contribution from side chain conformational freedom to protein stability. There are several areas of research which are quite interesting and deserve more attention. One is the solvation of peptide side chains. The interactions of the backbone, side chain, and solvent are complex, and can likely be elucidated with free energy calculations, which are ideal for studies such as this. The other area of importance is the reproduction of subtle structural changes seen in the protein upon mutation. Our results indicate that there are significant contributions to protein stability from these changes. Simulations of folded proteins in a periodic box of waters with the Cornell *et al.* (1995) force field and the Ewald electrostatic model show great promise in reproducing the dynamic properties of proteins (Fox and Kollman, 1996). An analysis of the free energy contribution from backbone changes using this model should prove fruitful. And finally, further modeling of the denatured state is warranted. Y.C. Sun *et al.* (1996) have already begun such studies with partially unfolded states of barnase.

In conclusion, the research presented in this thesis will enable more accurate and precise modeling of proteins and has shed new light on the roles of packing, solvation, and conformational freedom in protein stability. The future projects in modeling biological systems will become even more complex and challenging. It is the role of the scientists conducting them to insure the soundness of their results, and thus assure the contribution of molecular modeling to the greater understanding of nature.

## Proposed Research

*Protein Unfolding*

In NMR hydrogen exchange protein refolding experiments, Lu and Dahlquist (1992) found that 84 out of 164 backbone amides in T4 Lysozyme are protected enough in the native state to serve as probes of protein folding. The results indicated that an early intermediate is formed during refolding. The amides with the highest degree of protection are in residues 93-105 in helix E as well as residues in helix A, both of which are in the C-terminal structural domain. Helix E is the most buried helix in the protein and helix A packs against it. It is therefore likely that this region forms an initial framework for folding of the rest of the protein. Residues in the N-terminal domain also show significant protection in beta-sheet regions. While these experiments provide valuable data on the presence of secondary structure during folding, detailed structural information of these states is difficult to obtain experimentally due to the multitude of states which exist in low concentration; thus, analysis via X-ray crystallography and NMR is difficult.

The data of Lu and Dahlquist provide the opportunity to study the unfolding of T4 lysozyme with molecular dynamics simulations. No other native state has been characterized better in terms of its structural and energetic responses to mutation (Matthews, 1995). It will thus be interesting to compare simulations of the unfolding of the protein with the adaptability of its native state to mutation. Do the domains of T4 unfold concertedly? How does its unfolding compare to hen egg white lysozyme unfolding simulations? Is a stable intermediate found during unfolding, and does it match the intermediate found experimentally? How does hydrogen bonding of backbone amides compare to exchange protection (Alonso and Daggett, 1995)? These are some of the questions that will be addressed by these simulations.

I am also interested in conducting simulations on fragments of T4 lysozyme to enable a more detailed analysis of the interactions between secondary structure elements such as helices E and A. The motivation for these studies is two-fold. Firstly, efforts to decrease the size of technologically useful proteins have focused on finding smaller fragments which retain stability and function (Li *et al.*, 1995; Struthers *et al.*, 1996). Removing residues from a protein typically leads to destabilization of the remaining structure due to the exposure of hydrophobic residues and the loss of specific electrostatic interactions. Subsequent rounds of mutagenesis are then needed to reverse this destabilization. Simulations of the protein fragment at room temperature as well as at higher temperature should indicate which regions are less stable and the cause of their instability. Efforts could then be focused on these regions in experiments which diversify and select sequences (Matthews and Wells, 1993). The results should be generalizable since the disruptions introduced would be similar for other proteins. The second motivation for examining protein fragments is to investigate the role they may play as isolated, stable structural units in folding (Ruiz-Sanz *et al.*, 1995). The structures of helices A, C, and B have been studied by CD and NMR spectroscopy in SDS micelles and trifluoroethanol (TFE) solutions, respectively, which are known to stabilize alpha-helices (McKeish *et al.*, 1993; McLeish *et al.*, 1994; Najbar *et al.*, 1995). Simulations of the E and A helices in water as well as TFE would be of interest as it is these two helices that are proposed to form a stable nucleus in the folding of the protein.

In addition to the unfolding process mentioned above, efforts now underway in the Dahlquist group aim to experimentally determine the force required to unfold T4 lysozyme using an atomic force microscopy (AFM) (Dahlquist, 1995). It is hoped that the protein can be unfolded by attaching a modified residue of the protein to a Formica surface at one position and to the tip of the AFM at another site. The protein has been successfully linked to the Formica surface and efforts are now underway to develop the chemistry which will

enable attachment to the unreactive carbon tip. While this is an ambitious experiment, success in the study of biotin-streptavidin binding using AFM offers encouragement (Florin *et al.*, 1994; Moy *et al.*, 1994). Simulations of such an unfolding process would provide highly valuable structural information to compliment an exciting and novel approach to the investigation of protein structure (Berendsen, 1996; Grubmuller *et al.*, 1996).

*Backbone Response to Mutation*

Recently, great strides have been made in the prediction of side chain conformations using combinatorial searches of side chain rotamers and algorithm developments have sped the process of analyzing the possible combinations of side chain rotamers (Tuffery *et al.*, 1993; Lasters *et al.*, 1995). These techniques have applications in homology modeling, the design of protein-ligand interfaces, and can guide mutational experiments by indicating regions amenable to packing rearrangements.

However, these methods are limited in usefulness because they do not allow for movement of the protein backbone. It has been shown in experimental studies (Eriksson *et al.*, 1992), as well as in Chapter 4, that adjustment of the backbone in response to a perturbation is important energetically, and thus critical if the proper structure and function are to be selected. Lim *et al.* (1994) solved the crystal structure of a lambda repressor mutant that is more stable than wild-type despite three larger side chains in the protein core. Interestingly, the protein responds to this increase in packing volume via the outward movement of an alpha-helix rather than alterations of side chain rotamers. Not surprisingly, the Ponders and Richards (1987) side-chain packing algorithm rejects this variant. Lee and Levitt's algorithm (1991; Lee, 1995), which uses simulated annealing, accepts it, but only because of distortion of side chain torsions rather than backbone movement. In an analogous experiment, Baldwin *et al.* (1994) determined the crystal structures for eight variants of T4 lysozyme, each of which had three to five substitutions

in the C-terminal domain and were selected from a mutant pool based on enzyme function. The mutants were accommodated predominantly through backbone adjustments. The RMS shifts in backbone atoms for all mutants ranged from 0.17 to 0.63 Å, with individual alpha-carbon atoms moving up to 1.7 Å, yet all but one substituted residue retained the wild-type side chain rotamer within 20°. None of the variants were allowed according to the Ponders and Richards algorithm. It is clear from these studies that the assumption of a rigid backbone is not compatible with the dynamic nature of proteins.

A better fundamental understanding of the backbone conformations available within a fold class is needed for more accurate predictions of side chain packing. The studies of Baldwin and Lim suggest that certain backbone segments in protein cores are inherently more likely to adjust in response to changes in sequence. Thus, some low energy backbone conformations for one sequence may be compatible with a homologous sequence. In both the lambda repressor and T4 mutants, use of the mutant backbone structure greatly improved the performance of the packing algorithms. The question thus arises: are the backbone conformations of mutant sequences similar to states accessible to the native sequence?

I propose to generate sets of different backbone "microfolds" for the native fold of T4 lysozyme using molecular dynamics (MD) simulations to investigate this hypothesis. Molecular dynamics simulations at room temperature using the latest force field and the Particle Mesh Ewald method to model electrostatics produce RMS deviations of about 1.0 Å overall and 0.5-1.0 Å for buried residues (Fox and Kollman, 1996). These deviations in backbone structure are similar in magnitude to those seen in the studies of Lim and Baldwin, multiple crystal forms of T4 lysozyme mutants (Zhang *et al.*, 1995), and highly homologous proteins (Summers *et al.*, 1987). Thus the conformations sampled with MD should be similar to the native fold while allowing for appropriate deviations. In addition

to the backbone structures generated with MD, the many mutant structures of T4 lysozyme can be used as a database of alternate backbone conformations (Matthews, 1995).

The various backbone "microfolds" will be tested for compatibility with the eight sequences studied by Baldwin *et al.* These sequences varied from three to five substitutions at positions 121, 129, 133, 149, and 153. A side chain packing algorithm has been tested on 52 interior side chains in T4 lysozyme and offers encouraging results: 92% of $X_1$ and 88% of $X_2$ side-chain dihedrals were accurately reproduced within +/- 40° (Keller *et al.*, 1995). It is hoped that some if not all of the sequence changes observed will be allowed for certain microfolds. The backbone space sampled may only allow for limited changes; thus a protocol whereby small changes in sequence are made followed by relaxation of accepted structures and generation of a new set of microfolds using the new sequence may be optimal. The goal is to select those backbone changes which allow for the sequences and side chain packing observed experimentally; it is proposed that the protocols discussed above will yield these changes and shed light on the sequence and packing space available to proteins in general.

In conclusion, my goal is to conduct unfolding simulations of T4 lysozyme in order better understand the folding process, folding intermediates, and the key interactions which allow folding to occur. Molecular dynamics simulations can also be used to generate alternate backbone conformations for a protein fold which can then be used to account for protein backbone adjustments due to sequence differences.

## References


Alonso, D.O.V. and Daggett, V. (1995)*J Mol Biol,* **247**, 501-520.

Baldwin, E.P., Hajiseyedjavadi, O., Baase, W.A., and Matthews, B.W. (1993) *Science,* **262**, 1715-1718.

Berendsen, H.J.C. (1996) *Science,* **271**, 954-955.

Cornell,W.D., Cieplak,P., Bayly,C., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) *J. Am. Chem. Soc.*, **117**, 5179-5197.

Dahlquist, R. (1995) personal communication.

Eriksson, A.E., Baase, W.A., Zhang, X.-J., Heinz, D.W., Blaber, M., Baldwin, E.P., and Matthews, B.W. (1992)*Science*, **255**, 178-183.

Florin, E.L., Moy, V.T., and Gaub, H.E. (1994)*Science,* **264**, 415-417.

Fox, T. and Kollman, P.A. (1996) *Proteins*, **25**, 315-324.

Grubmuller, H., Heymann, B., and Tavan, P. (1996)*Science,* **271**, 997-999.

Keller, D.A., Shibata, M., Marcus, E., Ornstein, R.L., and Rein, R. (1995) *Protein Engineering,* **8**, 893-904.

Lasters, I., De Maeyer, M., and Desmet, J. (1995) *Protein Engineering,* **8**, 815-822.

Lee, C. and Levitt, M. (1991) *Nature,* **352**, 448-451.

Lee, C. (1995) *Folding and Design*, **1**, 1-12 (1995).

Li, B., Tom, J.Y.K., Oare, D., Yen, R., Fairbrother, W.J., Wells, J.A., and Cunningham, B.C. (1995) *Science,* **270**, 1657-1660.

Lim, W.A., Hodel, A., Sauer, R.T., and Richards, F.M. (1994) *PNAS*, **91**, 423-427.

Lu, J. and Dahlquist, F.W. (1992)*Biochemistry,* **31**, 4749-4756.

Matthews,B.W. (1995) *Adv. Prot. Chem.*, **46**, 249-278.

Matthews, D.J. and Wells, J.A. (1993)*Science,* **260**, 1113-1117.

McLeish, M.J., Nielsen, K.J., Wade, J.D., and Craik, D.J. (1993) *FEBS Letts*, **315**, 323-328 (1993).

McLeish, M.J., Nielsen, K.J., Najbar, L.V., Wade, J.D., Feng, L., Doughty, M.B., and Craik, D.J. (1994)*Biochemistry*, **33**, 11174-11183.

Moy, V.T., Florin, E.L., and Gaub, H.E. (1994)*Science*, **266**, 257-259.

Najbar, L.V., Craik, D.J., Wade, J.D., Lin, F., McLeish, M.J. (1995)*Biochim et Biophys Acta* , **1250**, 163-170.

Ponder, J.W. and Richards, F.M. (1987) *J Mol Biol*, **193**, 775-791.

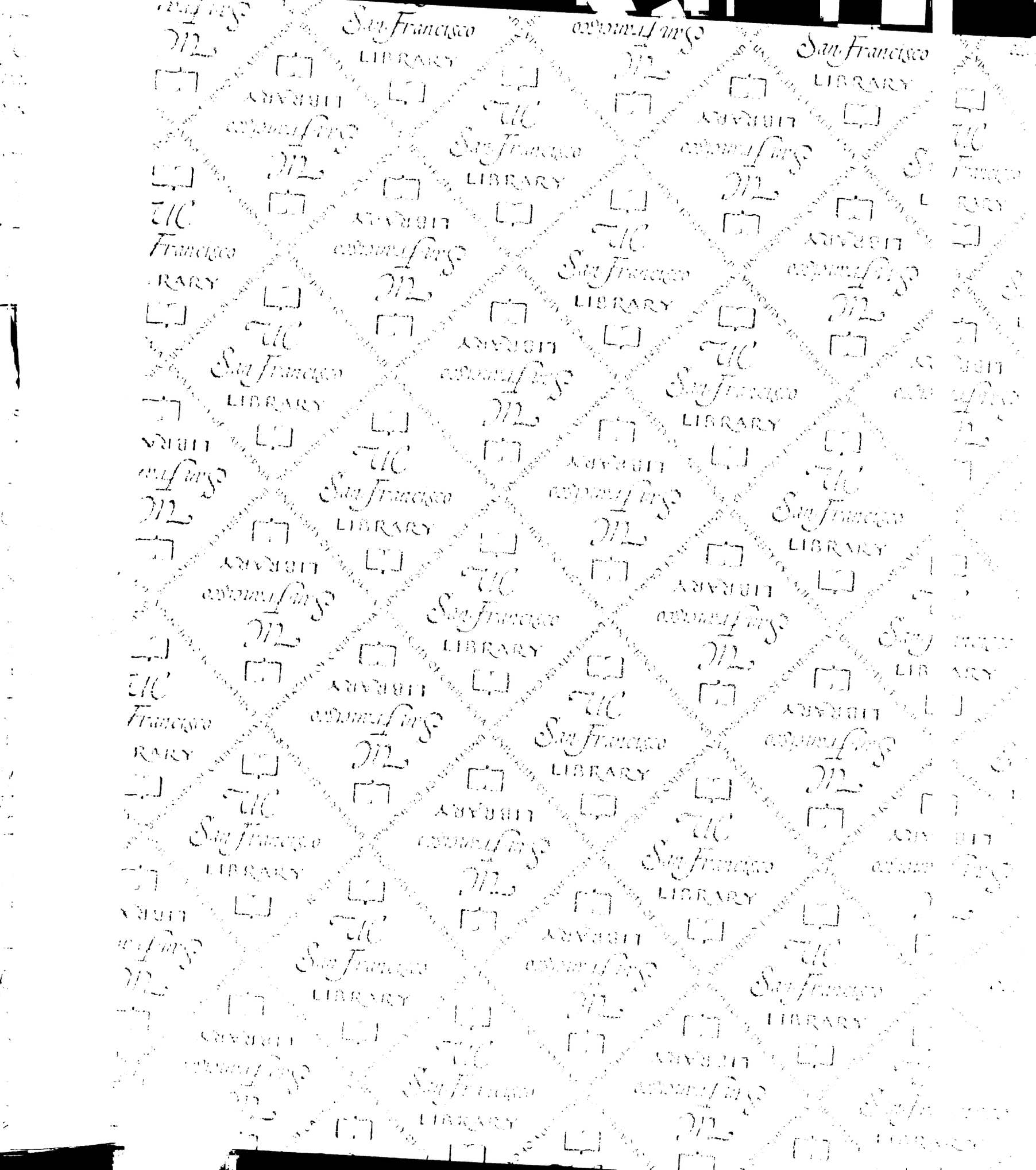
Ruiz-Sanz, J., de Prat Gay, G., Otzen, D.E., Fersht, A.R. (1995)*Biochemistry*, **34**, 1695-170.

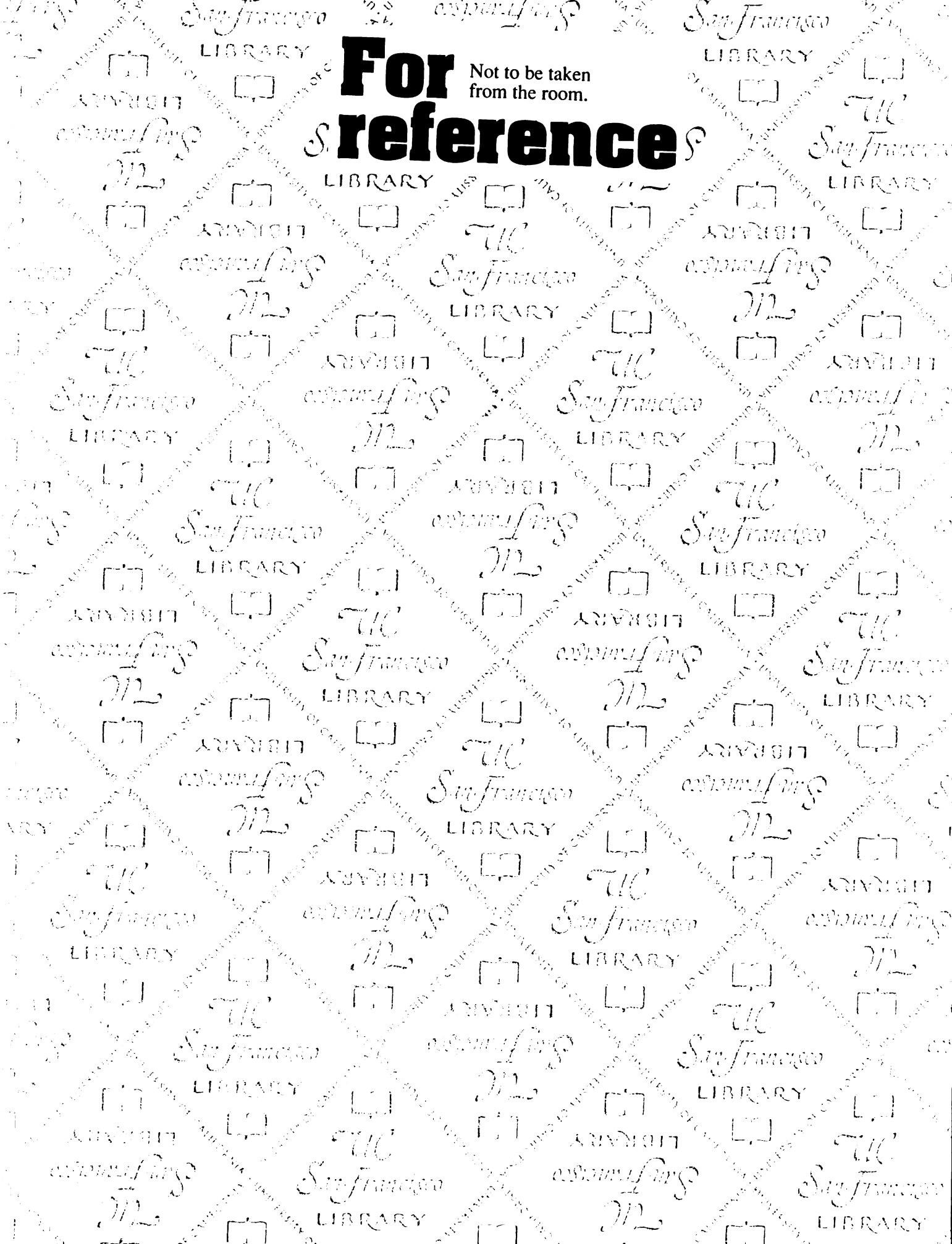
Struthers, M.D., Cheng R.P., and Imperiali, B. (1996) *Science*, **271**, 342-345.

Summers, N.L., Carson, W.D., and Karplus, M. (1987) *J. Mol. Biol.*, **196**, 175-198.

Tuffery, P., Etchebest, C., Hazout, S., and Lavery, R. (1993) *J. Comp. Chem.*, **7**, 790-798.

Zhang, X-j., Wozniak, J.A., and Matthew, B.M. (1995) *J. Mol. Biol.*, **250**, 527-552.


San Francisco
LIBRARY


For reference
Not to be taken from the room.
The San Francisco LIBRARY