

**COMPUTER SIMULATION OF FREE ENERGIES TO PREDICT
CIS/TRANS EQUILIBRIA OF PROLYL PEPTIDES
AND
SOLVATION FREE ENERGIES OF PHENYLALANYL PEPTIDES**

by

Tamaki Kurusu

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

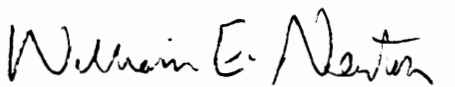
in

Biochemistry

APPROVED:



D. R. Bevan, Chairman



W. E. Newton



E. M. Gregory

August, 1996
Blacksburg, VA

Keywords: Molecular Modeling, Simulation, Free Energy, Proline, Solvation, Peptide

**COMPUTER SIMULATION OF FREE ENERGIES TO PREDICT
CIS/TRANS EQUILIBRIA OF PROLYL PEPTIDES
AND
SOLVATION FREE ENERGIES OF PHENYLALANYL PEPTIDES**

by

Tamaki Kurusu

David R. Bevan, Chairman

Department of Biochemistry

(ABSTRACT)

Two computer simulation studies were performed; one to help understand the structure-function relationships of prolyl peptides (Part I) and the other to help predict more efficient pharmaceutical drug delivery by molecular modification of small peptides (Part II).

In Part I, the free energy perturbation (FEP) method, using AMBER, was utilized to calculate the Gibbs free energy difference between *cis* and *trans* conformers of Ace-Tyr-Pro-NMe and Ace-Asn-Pro-NMe, from which the ratio of *cis* to *trans* conformers was obtained. Our simulation generated much lower %*cis* for both peptides as compared with experimental values and possible problems in our computational schemes are presented. However, our results were encouraging in that they predicted preference of *trans* conformers for both peptides and higher %*cis* for Ace-Tyr-Pro-NMe, compared to Ace-Asn-Pro-NMe, which agrees with experimental results.

Part II applied semiempirical (AMSOL) and microscopic simulation (POLARIS)

methods to obtain the solvation free energies of a series of phenylalanyl peptides with various degrees of methylation on their backbone nitrogens. It was clearly predicted that as a peptide length increased, so solvation free energy decreased, indicating less favorable permeability through the cell membrane system, in agreement with data in the literature. AMSOL also showed that solvation free energy change upon methylation was variable depending on the position of the substituted backbone nitrogen, which disagrees with the literature. However, non-systematic solvation free energy change of small amines upon methylation was successfully predicted by AMSOL, in good accord with experimental data.

ACKNOWLEDGMENTS

I would like to thank Dr. David Bevan for his patience, guidance, and helpfulness throughout my master's program. I would also like to extend my thanks to my committee members, Dr. William Newton and Dr. Eugene Gregory, for their patience and support. I recognize the faculty members and graduate students in the Department of Biochemistry for their assistance and encouragement. Special thanks are also due to Dr. Debabrota Ghosal for his helpful advice and friendship.

Departmental financial assistance during my early graduate program and the training opportunity at the Food Analysis Laboratory Control Center (FALCC) were greatly appreciated. Many thanks to Dr. Kent Stewart, Dr. Katherine Phillips, and all the personnel at FALCC for their direction and understanding.

Lastly, I would like to thank my family for their love and support.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	vii
LIST OF TABLES	viii
INTRODUCTION	1
PART I - <i>CIS/TRANS</i> EQUILIBRIA OF PROLYL DIPEPTIDES	3
LITERATURE REVIEW	4
METHODS	33
RESULTS	34
DISCUSSION	50
REFERENCES	58
PART II - SOLVATION FREE ENERGIES OF PHENYLALANYL PEPTIDES ...	66
LITERATURE REVIEW	67
METHODS	76
RESULTS	79
DISCUSSION	99
REFERENCES	103
PERSPECTIVES	106

APPENDIX A (PART I - Detailed Methods)	107
APPENDIX B (PART I - Additional Calculation Results)	122
APPENDIX C (PART II - Detailed Methods)	128
APPENDIX D (PART II - Additional Calculation Results)	138
VITA	150

LIST OF FIGURES

PART I - CIS/TRANS EQUILIBRIA OF PROLYL DIPEPTIDES

Figure 1	Free Energy Profile of Ace-Tyr-Pro-NMe by FEP with 100 windows <i>in vacuo</i> (<i>cis-trans</i>)	37
Figure 2	Free Energy Profile of Ace-Tyr-Pro-NMe by FEP with 500 windows <i>in vacuo</i> (<i>cis-trans</i>)	38
Figure 3	Free Energy Profile of Ace-Tyr-Pro-NMe by FEP with 1000 windows <i>in vacuo</i> (<i>cis-trans</i>)	39
Figure 4	Free Energy Profile of Ace-Tyr-Pro-NMe by FEP with 10000 windows <i>in vacuo</i> (<i>cis-trans</i>)	40
Figure 5	Gibbs Free Energy Profile of Ace-Tyr-Pro-NMe by FEP with 1000 windows in water (<i>cis-trans-cis</i>)	45
Figure 6	Gibbs Free Energy Profile of Ace-Asn-Pro-NMe by FEP with 1000 windows in water (<i>cis-trans-cis</i>)	46

PART II - SOLVATION FREE ENERGIES OF PHENYLALANYL PEPTIDES

Figure 1	Solvation Free Energy Change by Methylation (kcal/mol)	87
-----------------	--	----

LIST OF TABLES

PART I - CIS/TRANS EQUILIBRIA OF PROLYL DIPEPTIDES

Table 1	Summary of % <i>Cis</i> in Prolyl Dipeptides in the Literature	21
Table 2	Comparison of Free Energy Differenece (<i>cis</i> → <i>trans</i>) for Ace-Tyr-Pro-NMe <i>in vacuo</i> by FEP (AMBER 4.0)	41
Table 3	Gibbs Free Energy Change for Ace-Tyr-Pro-NMe and Ace-Asn-Pro-NMe in water by FEP (AMBER 4.1)	44
Table 4	Atomic Charges of Ace-Tyr-Pro-NMe by MOPAC 6.0	48

PART II - SOLVATION FREE ENERGIES OF PHENYLALANYL PEPTIDES

Table 1	Partition Coefficient (Log P) and Permeability Values in the Literature for D-Phenylalanyl Peptides	69
Table 2	D-Phenylalanyl Peptides Used for Solvation Free Energy Simulation	77
Table 3	Test Calculation Results by AMSOL and POLARIS	80
Table 4	Dipole Moments Comparison in MOPAC and AMSOL	81
Table 5	Solvation Free Energies Calculated by POLARIS with MNDO	83
Table 6	Solvation Free Energies Calculated by AMSOL with 1SCF Option	85
Table 7	AMSOL Test Calculation Results with 1SCF Option	86
Table 8	AMSOL Charge Comparison among Mono-D-Phe Peptides	88
Table 9	AMSOL Results - Dipole Moment and Surface Area	90
Table 10	AMSOL Results - Components of Free Energy of Solvation	91

Table 11	Comparison of Solvation Free Energies of Amides and Amines in the Literature	94
Table 12	Atomic Charge Comparison in the Literature	96
Table 13	Dipole Moments Comparison for Amines in the Literature	97
Table 14	AMSOL Results for Amines - Components of Free Energy of Solvation	98

INTRODUCTION

Along with the advancement of computer technology and the development of computational chemistry, molecular modeling has become a very important and powerful method for chemists and biochemists to predict the structure of molecules, characterize their properties, and understand chemical reactions and pathways at the atomic level.

Not only does molecular modeling allow us to study chemistry by running calculations on computers rather than by examining molecules and reactions experimentally, but also it is used to model short-lived, unstable intermediates and compounds that are impossible to study through observation. Moreover, it can help researchers quickly find molecules of enough interest from a large number of possible molecules to justify their preparation and examination, and predict whether they will be sufficiently stable to make experimental investigation practical.

In this study, molecular modeling is being tested for its ability to predict molecular properties of interest: *cis/trans* equilibria, by calculating Gibbs free energy differences between two conformational states, and permeability through lipid bilayer membranes by computing solvation free energies of the molecules of pharmaceutical interest. Examining the accuracy as well as practicality of molecular modeling will be the main focus of this project.

In Part I, *cis/trans* equilibria of two prolyl dipeptides, Ace-Tyr-Pro-NMe and Ace-Asn-Pro-NMe, are examined by first calculating the Gibbs free energies, which are then used to compute %*cis* of each peptide. The conformational characteristics of proline have been

implicated in various biological events as key factors of regulation. For example, proline imide bonds restrict their susceptibility to most peptidases by their stereochemistry and/or specific positional conditions. Also, N-terminal X-Pro sequence of biologically active peptides often regulates stability, maturity, and cellular locations of those peptides. Furthermore, the discovery of prolyl-peptidyl *cis/trans* isomerase and its ubiquity suggests its role in protein folding. Therefore, understanding *cis/trans* equilibria of prolyl peptide bonds in proteins and peptides may be the key to elucidate structure-activity relationships of many biologically active molecules in living cells.

In Part II, solvation free energies are computed for a series of phenylalanyl peptides with various degrees of methylation on their backbone nitrogens. These peptides were chosen because experimental results indicate that their partition coefficients do not correlate with their permeabilities through lipid bilayer membranes. Because passive absorption of peptides involves desolvation, the solvation free energy might be a very useful predictor of molecular permeability. Fast and yet accurate free energy predictions by computer simulation are very important for an efficient and reliable drug screening in the future.

PART I

***CIS/TRANS* EQUILIBRIA OF PROLYL DIPEPTIDES**

LITERATURE REVIEW

I. Characteristics of Proline in Biological Molecules

The amino acid proline holds a unique position among the 20 predominant amino acids in all proteins and peptides. The standard amino acids are called α -amino acids because they have a primary amino group and a carboxylic acid group substituent on the same carbon atom; on the other hand, proline has a secondary amino group due to its cyclic side chain and therefore is actually an α -imino acid.^[1] The cyclic nature of this amino acid gives two important conformational characteristics in a peptide bond: (1) the ϕ angle is constrained to values of -68° or -75° , depending on the ring-puckering positions (up or down, respectively),^[2] and (2) the peptide bond, X-Pro, gives both *cis* and *trans* conformations,^[3] whereas virtually all other peptide bonds found in proteins are in the *trans* conformation.^[4,5]

The following are important characteristics and possible biological roles of proline residues in proteins and peptides due to their unique conformational properties.

1. Conformational Preferences of Proline

Proline is almost never found at the carboxyl terminus of an α -helix and is rarely in the center of this structure, whereas it is observed at or near the amino terminus of helical segments of proteins.^[2] Under certain conditions, however, proline can appear in the center of a helix, which is bent at the proline residue.^[6] Therefore, prolyl residues perturb the structure of the preceding and following residues through steric interactions and by preventing

hydrogen-bond formation with a backbone carbonyl group.^[7] Also, proline has an unusually high probability of being found in a β -turn or a γ -turn, both in solution and the solid state, and severely constrains the backbone structure.^[8] Thus, inclusion of a proline residue in the sequence of a peptide or protein is expected to result in unique conformational preferences.^[2]

2. *Cis/Trans* Conformation at X-Pro Peptide Bonds

Due to the partial double bond nature of peptide bonds, they have a planar structure and assume either *cis* or *trans* conformation with a rotational barrier of approximately 20 kcal/mol.^[9] Generally they are in the *trans* conformation unless proline residues are involved.^[9] Very few non-proline *cis* peptide bonds have been recognized in native proteins by X-ray crystallography.^[10] In fact, analysis of structures in the Brookhaven Protein Data Bank revealed that only 0.05% of *cis* peptide bonds were found out of more than 31,000 amide bonds, whereas, 6.5% of 1,500 imide (X-Pro) bonds were shown to be in the *cis* conformation.^[10]

Because proteins are synthesized enzymatically and so in a stereospecific way, it is assumed that the translation process would generate only *trans* peptide bonds.^[10,11] Given the high barrier to *cis/trans* isomerization at room or physiological temperature, the existence of genetic control or intracellular catalysts of isomerization might be suspected. In fact, it was questioned if the four codons for proline, CCX where X = U, C, A, or G, determine the isomeric state of a proline imide bond in the folded protein and this possibility was investigated by inspecting proteins of known tertiary structure.^[12] The results show that there

is no significant correlation between the use of particular codons and the isomeric state of X-Pro bonds in folded proteins.^[12]

3. Prolyl-Peptidyl *Cis/Trans* Isomerases (PPIases)

In the late 1970s immunosuppressive drugs were introduced for clinical use, and they have had a profound effect on the success rate of organ transplant surgery and the treatment of autoimmune diseases.^[13] The immunosuppressive drugs include cyclosporin A (CsA), FK506, and rapamycin, which are known to bind to intracellular proteins, collectively called immunophilins.^[14] The CsA-binding immunophilins are called cyclophilins^[15,16] and the FK506- and rapamycin-binding ones are termed FK-binding proteins (FKBPs).^[17,18] It has been demonstrated that immunophilins display the prolyl-peptidyl *cis trans* isomerase (PPIase) activities,^[15,16] which are CsA-, FK506-, or rapamycin-sensitive.^[19]

Immunophilins are quite abundant (up to 0.4% of total cytoplasmic protein in various cells^[18,20]) and are phylogenetically highly conserved and present in almost all types of cells, ranging from those of mammals to bacteria.^[13,20] In addition, complexes of immunophilin (PPIase) and immunosuppressants are associated with the inhibition of T-cell activation upon antigen recognition and the subsequent proliferation of T cells^[21] by binding and inhibiting the cytoplasmic Ca^{2+} - and calmodulin-dependent phosphatase, namely calcineurin.^[22,23] This inhibition results in inability of the proper assembly and recognition of the appropriate promoter region on DNA of the transcription factor NF-AT (nuclear factor of activated T cells),^[13] consequently interfering with the normal transcriptional activation of interleukin-2

(IL-2)^[24] and so preventing T-cell activation. Therefore, immunosuppressive drugs prevent normal T-cell activation upon antigen recognition by interfering with Ca^{2+} -dependent, unique signal transduction pathways which regulate gene activities.^[13]

Recently, a new PPIase, parvulin, which is not a member of either the cyclophilin family or the FKBP family, was identified in *Escherichia coli*.^[25] All three families (cyclophilins, FKBP, and parvulins) include proteins containing only the basic PPIase domain and also proteins containing one or more PPIase domains as part of a large protein with N-terminal signal sequences to target them to the particular subcellular compartments (e.g., cytoplasm, endoplasmic reticulum, mitochondria, chloroplasts, and cell surface).^[26] PPIase domains of cyclophilin-, FKBP-, and parvulin families include approximately 165, 108, and 92 residues, respectively, and cyclophilins and parvulins have comparable PPIase activities, which are higher than that of FKBP.^[26] Although the nature of their catalysis remains to be completely elucidated, it seems that distortion-related events, which involves a transition state that is characterized by partial rotation about the C-N amide bond, are primary components in the mechanism of action.^[27] The active site of human cyclophilin (17.8 kD) was reported to have residues, Cys115, Cys62, Trp121, and His126, among which His126 appears to be of primary importance in prolyl isomerase activity.^[28]

In the absence of the immunosuppressive drugs, PPIases have been implicated in various cellular events, such as protein folding, protein translocation, and intracellular signaling, and it has even been suggested that they may be virulence factors in some infectious microbes.^[29] For example, it has been shown that CsA can inhibit the correctly folded state

of collagen,^[30] transferrin,^[31] and carbonic anhydrase,^[28] presumably via interference of PPIase activities. Also, *Nina A* (PPIase)-deficiency in *Drosophila* was reported to cause failing in folding and transporting of Rh1 rhodopsin from the rough endoplasmic reticulum to the rhabdomere (phototransduction organelle).^[29]

Therefore, ubiquitous PPIases may have multiple functions in supporting various cellular activities, suggesting the importance of proline-related protein conformations as key factors in those critical biological events.

4. Susceptibility of Proline Imide Bonds

The presence of proline in a polypeptide chain restricts its susceptibility to most peptidases,^[2] for some of which, there is an absolute requirement for X-Pro or Pro-X bond presence and/or specific positional conditions. For example, aminopeptidase P requires an N-terminal X-Pro bond and that bond must be in the *trans* conformation.^[33] Also, prolidase cleaves only the *trans* isomer of X-Pro bond in primarily dipeptides,^[34] and chymotrypsin makes its action only on the *trans* isomer of oligopeptides if proline occupies the second N-terminal position.^[35] For others, the requirement for the presence of a proline residue is less restrictive: proline-specific endopeptidase can cleave a Pro-Y bond in an -X-Pro-Y- sequence only when the preceding X-Pro bond is *trans*, but this enzyme also hydrolyzes Ala-Ala bonds at a much slower rate.^[36]

It can be said that *cis trans* isomerization at the prolyl peptide bond is generally required for the activity of many peptidases. Therefore, knowing that proline has the highest

specific positional preference (73%)^[10] of the residues located in bends and reverse turns,^[8] and bends and turns often appear on the surface of a protein, proline may serve as an important regulatory signal in determining the lifetime and degradation rates of biologically active proteins by controlling the susceptibility to most peptidases.^[2]

5. Influence of Proline in Biologically Active Peptides

It has been known that a number of biologically active polypeptides, including hormones, growth factors, neurotransmitters, coagulating proteins, toxins, and enzymes, contain an N-terminal X-Pro sequence.^[37] Some of them lose activity upon cleavage of the X-Pro bond; others remain active after such peptidolysis.^[2]

It is probable that removal of the X-Pro linkage at the N-terminus results in a major change in the stability, maturity, or cellular location of proteins. In another words, the remaining polypeptide may be susceptible to additional protease activities for further degradation or turnover, gain maturity, biological/enzymatic function, or be translocated to a different cellular compartment after specific X-Pro cleavage.

Because such an enzymatic cleavage of a prolyl peptide bond is stereo-specific, *cis/trans* isomerization may be involved in a regulation of the lifetime, biological activity, or processing of proteins or peptides which contain an N-terminal proline.^[38]

6. Role of Proline in Protein folding

Despite long-term, intensive studies, protein folding has been a major challenge and

is still not fully understood. A reaction as complex as protein folding must involve a variety of specific chemical processes, such as hydrogen bond or salt-bridge formation, solvent-induced hydrophobic associations,^[12] and shuffling of various interactions or bond formations among peptide residues and/or backbone functional groups until the correct three-dimensional structure of a particular peptide is achieved. It has been proposed that the essential information required for protein folding is contained within the primary amino acid sequence,^[39] however, some accessory proteins are now known to exist to assist a polypeptide to fold correctly and efficiently. Molecular chaperones are one type of accessory protein, and their predominant role appears to be preventing incorrect intermolecular association and aggregation of unfolded polypeptide chains.^[40] The major chaperone families which assist protein folding include HSP60 (Heat Shock Protein of approximately 60 kD), HSP70, HSP90, and SecB.^[41] Molecular catalysts are another type of accessory protein and these include protein disulfide isomerases (PDIsases) and PPIases. Because most proteins in eukaryotes are synthesized and processed in an oxidizing environment, disulfide bond formation is enhanced during the initial folding events and PDIsases help shuffling of various disulfide bonds until correct disulfide bonds are formed.^[42]

It has been proposed that *cis/trans* isomerization at the X-Pro bond in a peptide is the rate-determining step in protein folding.^[43] In fact, mutagenesis at the conserved proline residues in yeast cytochrome c resulted primarily in the elimination of the slow folding phases, leaving a greatly simplified kinetic mechanism of folding.^[12] Also, similar protein folding studies using ribonuclease T1 (RNase T1) from *Aspergillus oryzae*, which has two *cis* prolyl

peptide bonds at Pro39 and at Pro55, have shown that those proline residues are involved in the two slow folding steps.^[44] Mutagenesis at Pro39 resulted in a drastically destabilized protein and a new, very slow folding process, but replacement at Pro55 led to a much faster folding rate and a stable folded protein, indicating that the *cis* peptide bond at Pro39 is of utmost importance for stability and folding.^[45] In addition, a possible role of prolyl *cis/trans* isomerization in protein folding by PPIases was also suggested in the studies where the immunosuppressive drug CsA slowed down folding processes of collagen^[30] and transferrin,^[31] and where addition of PPIases accelerated the slow folding steps of carbonic anhydrase^[32] and barstar (an intracellular inhibitor of barnase, an extracellular RNase of *Bacillus amyloliquefaciens*).^[46]

It can be concluded that slow *cis/trans* isomerizations at the prolyl peptide bonds are ubiquitous phenomena in protein folding because most proteins contain prolyl residues,^[12,45] and that PPIases play an important role to help regulate those rate-determining steps in protein folding.

Therefore, proline residues in proteins and peptides confer significant regulatory signals on various biological events due to their unique conformational characteristics, and it is clear that understanding their conformational states, i.e., *cis trans* equilibria, in proteins and peptides and how they affect cellular activities may be the key to elucidating structure-activity relationships of many biologically active molecules in living cells.

II. Molecular Modeling Tools

There are two basic methods in molecular modeling: molecular mechanics and molecular quantum mechanics.

1. Molecular Mechanics

Molecular mechanics is probably the most popular method in molecular modeling and is designed to predict structures and energies of molecules.^[47] Each molecular mechanics method is characterized by its particular force field, which is defined by: 1) a set of equations to calculate the potential energy of a molecule, 2) a series of atom types, which depends on hybridization, charge, and the atomic environment, and 3) a parameter set, which fits the equations and atom types to experimental data.^[48]

One of the fundamental assumptions of molecular mechanics is that the total energy of a molecule can be divided into several parts.^[49] Thus the total energy, E_{total} , of a molecule is expressed as a sum of the potentials such as bond-stretching (E_{bond}), angle-bending (E_{angle}), torsional interactions (E_{tor}), van der Waals interactions (E_{vdw}), and lastly electrostatic interactions (E_{elec}):

$$E_{\text{total}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{tor}} + E_{\text{vdw}} + E_{\text{elec}}$$

Each part of the energy consists of simple, analytical functions to describe the energy of the system based on classical-mechanical models of molecular structure.^[50,51] Those functions include Hooke's law-type harmonic functions, truncated Fourier series functions, Lennard-

Jones functions, and so on.^[50,51] The parameters used in the above functions are determined by fitting to experimental data on various reference molecules.^[52] For example, equilibrium bond lengths and bond angles are taken from microwave and X-ray data on some model compounds and their force constants are obtained from normal-mode calculations or linear interpolation models.^[53,54] On the other hand, atomic charges for the electrostatic energies are given by quantum mechanical calculations.^[53]

In molecular mechanics, electrons are not explicitly included in calculations because the electronic and nuclear motions can be uncoupled from one another and considered separately by applying the Born-Oppenheimer approximation.^[49] Therefore, molecular mechanics approaches chemical problems based on the nuclear structure of the molecules and the electronic system is implicitly taken into account via parameterization in those potential functions mentioned previously.^[48,49]

Molecular mechanics assumes that the parameters of one molecule can be applied to a similar structural unit in another molecule.^[49] This transferability of parameters is the fundamental concept in molecular mechanics.^[47,49] However, parameters are in general not transferable from one force field to another because different forms of equations are used in each force field and its parameters have been specifically tuned within those equations to minimize errors.^[49]

Those approximations make the molecular mechanics method quite inexpensive computationally and, therefore, allow it to handle very large systems including enzymes, which may contain many thousands of atoms.^[48] Some of molecular mechanics force fields

include MM2 (Molecular Mechanics 2),^[55] which is designed primarily for small organic molecules and can be extended to peptides, AMBER (Assisted Model Building and Energy Refinement),^[53] CHARMM (Chemistry at HARvard Macromolecular Mechanics),^[54] and OPLS (Optimized Potentials for Liquid Simulations),^[56] which are all developed for proteins and nucleic acids and the last one was especially designed for solvated systems.

However, molecular mechanics also carries some limitations. The quality of calculated results depends on the extent of parameterization for different classes of molecules.^[48] In addition, due to the neglect of electrons, molecular mechanics cannot handle problems which involve bond formation or bond breaking processes, and it cannot predict certain molecular properties, such as reaction energies, multipole moments, or polarizabilities, which depend on molecular orbital interactions.^[48]

2. Molecular Quantum Mechanics

There are two major classes of molecular quantum mechanics methods: *ab initio* and semiempirical methods. They fundamentally differ from molecular mechanics in that their computations are based on the laws of quantum mechanics, not on classical mechanics, and that they treat electrons explicitly in the molecular system.

In molecular quantum mechanics methods, the solution to the Schrödinger equation, $H\Psi = E\Psi$, is sought to elucidate chemical characteristics and properties, where H is the Hamiltonian operator, Ψ is the wavefunction of the system, and E is the energy of the system. However, many approximations and assumptions must be employed to solve the Schrödinger

equation for most molecules because its exact solution is only possible for the hydrogen atom and other hydrogen-like atoms such as He^+ and Li^{2+} which have only one electron in each system.^[57]

The approximations and assumptions include the following: the molecular wavefunction is a function of only spatial coordinates not of time, the electrons have the same mass regardless of how fast they move, ignoring the relativistic effects, and the electron distribution within a molecular system depends on the positions of the nuclei, and not on their velocities (Born-Oppenheimer approximation).^[47,48]

In addition, we use the linear combination of atomic orbital (LCAO) approximation. The atomic wavefunction is represented as a combination of one-electron functions (orbitals) and then the molecular wavefunction is described by combination of those atomic orbitals.^[47,58]

(i) *Ab Initio* method

As its name implies, the *ab initio* ("from the beginning" in Latin) method uses the complete Hamiltonian, which describes the kinetic and potential energies of the system, and does not use any experimental data other than the values of the fundamental physical constants: the speed of light, the masses and charges of electrons and nuclei, and Planck's constant.^[48,57]

Ab initio methods employ Hartree-Fock theory. In this theory the molecular wavefunction is given by a Slater determinant which is expressed by the products of spin

functions and molecular orbitals which in turn are sums of the products of molecular orbital expansion coefficients and basis functions.^[48] The basis functions are in the form of exponential functions, called Slater-type orbitals (STOs) or Gaussian-type orbitals (GTOs), and express atomic spatial orbitals such as s, p, d, etc.^[57,59] To solve the Schrödinger equation, Hartree-Fock theory utilizes the variational method in that we determine the values of the molecular orbital expansion coefficients in such a way that those coefficients minimize the energy of the resultant wavefunction.^[48,58] To achieve this goal, the Roothaan-Hall equation is used. In this method, the Schrödinger equation is represented by the Hartree-Fock wavefunction, the orbital energy matrix, the matrix of overlap integrals between the basis functions, and the Fock matrix with terms consisting of the single-electron Hamiltonian, two-electron repulsion integrals, and the one-electron density matrix.^[47,58] The solution to the Roothaan-Hall equation depends on the molecular orbital expansion coefficients and so does the Fock matrix through the density matrix, therefore, it must be solved in an iterative fashion.^[48,58] Initially the density matrix, which is composed of molecular orbital expansion coefficients, is guessed to form the Fock matrix, and the Roothaan-Hall equation is solved to give an initial energy matrix, which is then used to compute an improved density matrix.^[57,60] Because this procedure continues until no further improvement in molecular orbital expansion coefficients and energies occurs from one cycle to the next, it is called the self-consistent-field (SCF) method.^[57]

The main deficiency of Hartree-Fock theory is its incomplete description of the electron correlation in the molecular system.^[48,58] It is ideal for the wavefunction to include

all possible alternative electron configurations using virtual orbitals as well as occupied ones and it is called full configuration interaction (CI).^[47,58] Full CI is best, but computationally very expensive and thus not practical.^[2] An alternative approach to the electron correlation treatment is Møller-Plesset (MP) perturbation theory.^[48,58] In MP theory, a correction corresponding to exciting two electrons to higher energy molecular orbitals is added to the Hartree-Fock wavefunction by using the full CI ground-state wavefunction and energy which are expanded as a power series function but truncated at the different orders.^[50,58] The MP method is not variational in that it may give the total energy which is not the upper limit of the true energy, but it is much faster to compute compared with the full CI method.^[47,50]

(ii) Semiempirical methods

Semiempirical quantum mechanical calculations differ from *ab initio* methods in that there are additional extensive approximations in the semiempirical approach, i.e., neglecting some of the terms or replacing them with simpler forms or with empirically determined parameters.^[47]

Firstly, semiempirical methods consider only the valence electrons of atoms and the inner-shell electrons are treated as part of an unpolarizable core.^[47,61] Secondly, some of the electron-electron interactions are neglected in semiempirical calculations. The Hückel theory, which is one of the earliest semiempirical methods, ignores all electron-electron repulsion terms.^[51] Neglect of Differential Overlap (NDO) methods neglect some electron-electron interactions.^[51] CNDO (Complete NDO) theory neglects the overlap distribution of any two

different atomic orbitals for the same electron (one-electron integrals) and of any four different atomic orbitals for two electrons (two-electron integrals), whereas INDO (Intermediate NDO) and MINDO (Modified INDO) methods retain the two-electron integrals where atomic orbitals are on the same atom.^[61,62,63] On the other hand, Neglect of Diatomic Differential Overlap (NDDO) approximation considers all interactions except those by diatomic differential overlaps.^[64] Modified Neglect of Diatomic Overlap (MNDO), Austin Model 1 (AM1), and Parametric Method Number 3 (PM3 or MNDO-PM3), where PM3 has been developed most recently, are all based on the NDDO method and use different parametric approximations.^[65,66,67] Lastly, semiempirical methods employ extensive empirical parameterizations to offset their approximations to solve the Schrödinger equation. Parameterizations are performed for repulsion and resonance integral terms, ionization potential values for s and p orbitals, and atomic orbital exponents, and so on by optimizing or simply using experimental data from atomic spectroscopy- and X-ray crystallography-type information on model molecules so that they can reproduce molecular properties such as geometries, heats of formation, dipole moments, and ionization potentials.^[60,63]

Whenever molecular modeling methods are considered to study any molecular system of interest, we always need to make a compromise between the rigor of the mathematical treatment of that system and the computer time that is required for the calculation.^[49] For *ab initio* and semiempirical methods, computing time and disk storage requirement go up approximately with l^4 and m^2 , respectively, where l is the number of basis functions and m is the number of valence orbitals.^[47] On the other hand, the time required in molecular

mechanics is proportional to approximately n^2 , where n corresponds to the number of atoms.^[49] Obviously, *ab initio* calculations are of limited utility in studying biological molecules and so semiempirical methods are more applicable to those systems than *ab initio* methods.^[52] For much larger molecules of biological interest, molecular mechanics is therefore the only currently viable computational approach available for molecular modeling.^[49]

III. Prediction of Prolyl *Cis/Trans* Equilibria by Molecular Modeling

The *cis/trans* equilibrium of prolyl peptide bonds can be described either by the equilibrium constant K or by the Gibbs free energy, ΔG , in the following equation:

$$\Delta G (cis \rightarrow trans) = - RT \ln K \quad (1)$$

where R = gas constant ($= 0.001987 \text{ kcal mol}^{-1} \text{ K}^{-1} = 8.3144 \text{ J mol}^{-1} \text{ K}^{-1}$),

T = temperature in Kelvin

and $K = [trans]/[cis]$.

In experiment, the equilibrium constant K is sought to obtain the *cis/trans* ratio directly. On the other hand, molecular modeling approaches first calculate ΔG and then the equilibrium constant K using formula (1).

1. Experimental Data

The equilibrium between the *cis* and *trans* conformations of X-Pro peptide bonds in

short peptides can readily be measured by the relative resonance intensities in ^{13}C NMR,^[68,69,70] ^1H NMR,^[69,71] ^{15}N NMR,^[70] and ^{17}O NMR.^[72]

Isomeric states at the X-Pro peptide bonds can also be determined by using proline-specific peptidases (aminopeptidase P, prolidase, etc.), where the *cis trans* ratio in the sample is estimated by comparing amplitudes of the fast and slow phases of proteolysis (Isomeric Specific Proteolysis or ISP).^[33] (See **Table 1.**) In the ISP method, fast and slow phases in proteolysis correspond to direct hydrolysis of the *trans* peptide and the *cis*-to-*trans* isomerization and subsequent bond cleavage, respectively.^[73] This technique has some advantages over NMR in that it helps investigate *cis/trans* equilibria at prolyl peptide bonds in larger proteins, and ISP also gives *cis trans* isomerization kinetics.

Even though the NMR technique has proven to be a very powerful tool to investigate *cis/trans* equilibria at prolyl residues, its use has been limited to small peptides due to problems caused by many overlapping resonances in larger molecules.^[73] In addition, ISP can be applied to larger proteins, but it is still difficult to study native proteins because initial denaturation or chain-trimming cleavage of the sample is necessary prior to the stereo-specific cleavage in order to make X-Pro bonds accessible to the prolyl proteases.^[73,74] Therefore, there are no experimentally feasible methods to examine *cis/trans* equilibria at X-Pro bond in native proteins yet.

Table 1 Summary of %*Cis* in Prolyl Dipeptides in the Literature

Peptide	Solvent	pD/pH	Temp(°C)	% <i>cis</i>	Method	Ref. #
H-Gly-Pro-OH	D ₂ O	6.5	25	37	¹³ C	68
	D ₂ O	7.5	25	37	¹³ C	71
	H ₂ O	7.9	23.5	39	ISP	34
H-Ala-Pro-OH	D ₂ O	6.5	25	40	¹³ C	68
	D ₂ O	7.5	25	39	¹³ C	71
H-Leu-Pro-OH	D ₂ O	6.5	25	49	¹³ C	68
H-Phe-Pro-OH	D ₂ O	6.5	25	76	¹³ C	68
	D ₂ O	7.0	25	77	¹³ C	71
	H ₂ O	7.9	23.5	73	ISP	34
H-Tyr-Pro-OH	D ₂ O	7.5	25	78	¹³ C	71
Ace-Asn-Pro-NMe	D ₂ O	6.6	25	13	¹³ C	69
	D ₂ O	6.6	25	15	¹ H	69
Ace-Tyr-Pro-NMe	D ₂ O	6.6	25	36	¹³ C	69
	D ₂ O	6.6	25	35	¹ H	69

Note: Ace = acetyl group
NMe = methylamide group
¹³C and ¹H = ¹³C MNR and ¹H NMR, respectively
ISP = Isomeric Specific Proteolysis

2. Molecular Modeling Methods

The molecular modeling tool available for Gibbs free energy calculations of biological molecules is termed the free energy perturbation (FEP) method, which computes ΔG of two different states, based on the statistical mechanical definition of free energy, either by composition or by conformation.

(i) Theoretical Background

Statistical Mechanics

The operational elements of statistical mechanics in the determination of free energy are configurational integrals based on a set of constant quantities such as T , V , and N (canonical ensemble for Helmholtz free energy), or T , P , and N (isothermal-isobaric ensemble for Gibbs free energy), where T is the absolute temperature, V is volume, P is pressure, and N is the number of particles in the system.^[75] In this discussion, the latter ensemble system is considered to focus on the Gibbs free energy calculations.

The configurational partition function is defined as:

$$Z_N = \int \dots \int \exp[-\beta E(X^N)] dX^N, \quad (2)$$

where $E(X^N)$ is the configurational energy, $\beta = (kT)^{-1}$, k is Boltzmann's constant, T is the absolute temperature, and the configurational integration extends over all space dX^N of the N -particle system.^[75] The Boltzmann probability function is:

$$P(X^N) = \exp[-\beta E(X^N)]/Z_N, \quad (3)$$

for a configuration corresponding to a particular energy, $E(X^N)$.^[75] The average or mean

energy expression is:

$$\langle E(X^N) \rangle = \int \dots \int E(X^N) P(X^N) dX^N, \quad (4)$$

which is equivalent to the internal energy (U), and the free energy is expressed as:

$$G = -kT \ln \langle \exp[-\beta E(X^N)] \rangle. [76] \quad (5)$$

Molecular Dynamics (MD) and Monte Carlo (MC) Methods

MD solves the classical mechanics of motion of all atoms of a system to allow the calculation of dynamical molecular properties, in which Newton's equation of motion is written as:

$$m_i (d^2 \mathbf{r}_i / dt^2) = -\nabla_i [U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)] \quad (6)$$

where $i = 1, \dots, N$ and

$$\nabla = \mathbf{i}(\partial/\partial x) + \mathbf{j}(\partial/\partial y) + \mathbf{k}(\partial/\partial z)$$

where m_i and $\mathbf{r}_i$ represent the mass and position of particle i , and U is the potential-energy surface, which depends on the positions of the N particles in the system.^[77,78] From the solution of these equations, the atomic positions and velocities as a function of time are obtained.^[78]

In MC simulations, the atoms are moved randomly according to probabilistic rules which lead to the generation of over a million structures for given thermodynamic conditions (temperature, volume, etc.).^[79] Although this approach has been popular in the study of liquids and small organic molecules, it has seldom been used to study models of proteins or nucleic acids.^[79] Dynamical simulations have been favored primarily because they directly

provide the time evolution of atomic motion.^[79] Also, molecular dynamics is far more efficient than traditional MC methods in providing equilibrium properties of large molecules.^[79]

ΔG is derived as:

$$\Delta G = G_b - G_a = -kT \ln \langle \exp[-\beta \Delta E(X^N)] \rangle_a \quad (7)$$

where $\Delta E = E_b - E_a$ is the difference between the potential energy functions describing states b and a .^[80] $\langle \dots \rangle_a$ means the ensemble average of the quantity within the brackets is to be calculated, by generating an equilibrium ensemble of conformations using the potential energy function corresponding to state a .^[80] MD or MC methods are used to sample the conformational space to compute the average quantities within the broken brackets, $\langle \dots \rangle_a$.^[81,82]

It is critical that the states a and b should not be too disparate so that the ensemble generated by E_a will contain enough conformations where E_b is low in energy.^[83] In other words, if the two states a and b are very different in that a particular conformation corresponds to a minimum energy for state b but a high energy for state a , then this conformation, which is important to b , will be rarely sampled since we are generating an ensemble of conformations using the potential for state a .^[84]

Coupling Parameter λ

In order to calculate the free energy difference ΔG , the coupling parameter, λ , is generally introduced and it defines a transition path between initial and final states, 0 and 1.^[85]

With this technique, we can calculate the ΔG between two disparate states a and b as a sum of ΔG differences between similar states.^[84] Thus, ΔG between λ_i and λ_{i+1} is given as:

$$\Delta G = G(\lambda_{i+1}) - G(\lambda_i) = -kT \ln \langle \exp[-\beta(E_{\lambda_{i+1}} - E_{\lambda_i})] \rangle_{\lambda_i} \quad (8)$$

and the total ΔG change can be obtained by summing the ΔG 's between the intermediate states as:

$$\Delta G = \sum_{i=0}^{k-1} \Delta G(\lambda_i \rightarrow \lambda_{i+1}) \quad (9)$$

where the interval $\lambda=0$ to $\lambda=1$ has been divided up into k -subintervals (windows).^[85]

It is important to note that nonphysical paths, i.e., 'mutational' processes in which a functional group, subunit, or even substrate in the initial state 0 is being changed into another in the final state 1, are admissible in the case where the quantity of interest is a state function like ΔG , which of course is independent of the path.^[85] Mutational processes are often used to examine effects of mutations on the structure or properties of the molecule of interest. In addition to the mutational processes, other changes along the λ -coordinate include a conformational transition (e.g., *cis/trans* isomerization), a structural transition (e.g., α -helix to random coil), and a simple reaction (e.g., bond formation or breaking).^[85]

Therefore, from an ensemble average of a potential energy difference, we can derive the free energy difference between two states.^[80]

(ii) Methodologies

The four main approaches to free energy determination via molecular simulation are

thermodynamic integration, the perturbation method, slow growth, and the potential of mean force calculations.^[75,86]

Thermodynamic Integration (TI)

In the TI method, ΔG between the initial state $\lambda=0$ and the final state $\lambda=1$ is calculated by the definition of an integral in calculus:

$$\Delta G = \int_0^1 [\delta G(\lambda)/\delta \lambda] d\lambda \quad (10)$$

and it is the same as:

$$\Delta G = \int_0^1 \langle \delta E(X^N, \lambda)/\delta \lambda \rangle_\lambda d\lambda \quad (11)$$

after some rearrangements.^[75,85] The simulation has never properly equilibrated unless λ is infinitesimally small and any change from $\lambda=0$ to $\lambda=1$ must be carried out on the picosecond time scale.^[85]

Perturbation Method

This method addresses the free energy calculation by employing the following equations:

$$\Delta G = -kT \ln \langle \exp[-\beta(E_1(X^N) - E_0(X^N))] \rangle_0 \quad (12)$$

or since two states 0 and 1 are interchangeable,

$$\begin{aligned} \Delta G &= -kT \ln \langle \exp[-\beta(E_0(X^N) - E_1(X^N))] \rangle_1 \\ &= -kT \ln \langle \exp[+\beta(E_1(X^N) - E_0(X^N))] \rangle_1 \end{aligned} \quad (13)$$

where the configurational averaging is performed with respect to the probability function representative of the final state of the system.^[85]

The two equations, (12) and (13), are referred to as running 'forwards' and 'backwards', respectively, on the λ -coordinate^[85] and this technique is called 'double-ended' or 'double-wide' sampling.^[81] The discrepancy between the two estimates of ΔG is an estimate of the statistical uncertainty in the result due to inaccuracies in the numerical integrations of (12) and (13) and arise from conformations sampled in $P_0(X^N)$ but not in $P_1(X^N)$ and vice versa.^[85]

The perturbation approach is numerically accurate only when the initial and final states of the system differ by a relatively small amount, i.e., ΔG differences should be less than 2 kT or ≈ 1.5 kcal/mol between the initial and final states.^[85] Therefore, for the simulation in which ΔG is considerably larger than 2 kT, the perturbation method can be used successfully by defining the intermediate states on the λ -coordinate which differ successively by no more than 2 kT.^[85]

Slow Growth

This method handles free energy calculations by changing λ infinitesimally over each step of the simulation so that any two sequential λ states are extremely similar and so we assume that the ensemble average used to calculate any ΔG_i can be approximated by a single, instantaneous value as:

$$\begin{aligned}\Delta G_i &= -kT \ln \langle \exp[-\beta(E_{\lambda+1}-E_{\lambda})] \rangle_{\lambda} \approx -kT \ln \exp[-\beta(E_{\lambda+1}-E_{\lambda})] \\ &= E_{\lambda+1} - E_{\lambda}\end{aligned}\quad (14)$$

therefore,

$$\Delta G = \sum_{\lambda=0}^{\lambda=1} (E_{\lambda+1} - E_{\lambda}).^{[80,86]}\quad (15)$$

This technique can be applied to either the TI or the perturbation method.^[86] However, the validity of this approach has not been proven yet.^[83]

Potential of Mean Force (PMF)

The PMF method deals with the radial distribution function for a distance (R) of any two particles, which is defined by:

$$g(R) = [N(N-1)/\rho] \exp[-\beta E(X^N|R)] dX^{N-2}\quad (17)$$

where ρ is the number density, and $G(R)$ is expressed as:

$$G(R) = -kT \ln[g(R)] + \text{constant}\quad (18)$$

where $G(R)$ is known as a potential of mean force and the derivative of $G(R)$ with respect to R is the force acting between the two particle.^[85] In a special case where two of the N -particles are fixed in space at a distance, R , and the free energy is expressed as:

$$G(R) = -kT \ln\{\exp[-\beta E(X^N|R)]\} dX^{N-2}\quad (16)$$

where $(X^N|R)$ denotes a conformation conditional on the two fixed particles being R , and the quantity R can be a simple distance, an internal coordinate representing a torsional degree, or a generalized structural alteration in the system.^[85] Therefore, PMF calculates free energy

differences between two states which differ by conformation. It is useful for the study of ΔG due to conformational changes, molecular associations, and the free energy profile for chemical reactions in condensed phases.^[85]

For all of the methods above, the energy function, E , which was originally a Hamiltonian function, is practically presumed to be the same as a potential function in molecular mechanics because of the assumption that kinetic energy contributions to ΔG between the states 0 and 1 are cancelled out.^[86]

It is true that an exact evaluation of the ensemble average of the quantity in brackets will exactly yield the free energy difference, however, for any complex system it is impossible to determine whether the ensemble average has converged with absolute certainty and this indeterminacy of convergence is the underlying problem in FEP calculations.^[80] This means that for any reasonably complex potential surface, we can never know absolutely that we have sampled all accessible conformations that would contribute significantly to the ensemble average,^[80] and one obtains only estimates of the true, infinite-order ensemble averages of the energy from the finite-order computer simulation.^[75,87] Running a simulation for a longer period of time should yield a more accurate evaluation of the free energy, and simulations run for different lengths of time should yield approximately the same value if the shorter simulation has converged.^[80] Since obtaining a seemingly 'good' value from a free energy calculation is frequently fortuitous, additional testing is required to ensure the value has legitimately converged.^[80]

(iii) Published Results

FEP approaches have been applied to the studies of solvation (see the next section), molecular association,^[88,89,90] macromolecular stabilities,^[89] and enzyme catalysis,^[86] however, very few cases have been found for conformational ΔG calculations.

The barrier to isomerization of *N,N*-dimethylacetamide (DMA) was calculated with the OPLS force field with additional parameters developed for the anti and syn transition states (with the nitrogen lone pair anti and syn to the oxygen) by *ab initio* calculations. $\Delta G^\ddagger$ for the anti and syn transition states from the planar ground state in water were 16.7 kcal/mol and 16.9 kcal/mol, respectively.^[91] This simulation was done by generating 1M (1×10^6) conformations of equilibration and 3M conformations of data collection in each window (4.5°/window) in the NPT ensemble (constant Number of particles, Pressure, and Temperature in the system) at 25 °C and 1 atm.^[91] In addition, Monte Carlo simulation using combined quantum and molecular mechanical potentials (QM/MM),^[92] where part of the system (the solute molecule or the active site region of an enzyme) is treated quantum mechanically and the environment or solvent molecules are approximated by molecular mechanical potentials, has been applied to calculate the barriers to isomerization of *N,N*-dimethylformamide (DMF) in water.^[93] In this approach, the semiempirical AM1 method, which was applied to DMF, was modified to reproduce barrier heights obtained from *ab initio* and experimental results and the empirical TIP3P model for water approximated the solvent molecules.^[93] The calculated $\Delta G^\ddagger$ for the anti and syn transition forms from the planar ground state were 22.5 ± 0.2 kcal/mol and 20.8 ± 0.1 kcal/mol, respectively, where at least 500K

conformations were generated for equilibration followed by 1-1.5M conformations for data collection in each window (about 8°/window) at 25 °C and 1 atm with the NPT ensemble. This simulation required about eight months with two SUN SLC computers.^[93] Experimental values for the free energy of activation for isomerization of DMA and DMF in water have not been reported.

Hodel et al. studied the *cis/trans* conformational ΔG difference using nine-residue peptides from staphylococcal nuclease by the TI method with a thermodynamic cycle technique.^[94] In this work the difference between the isomerization free energies of the two peptides with mutations K¹¹⁶A and K¹¹⁶G at the X¹¹⁶-Pro¹¹⁷ peptide bond were examined without explicit solvent by mutating A to G in each isomeric state as:

$$\begin{aligned}\Delta \Delta G &= \Delta G_A(cis \rightarrow trans) - \Delta G_G(cis \rightarrow trans) \\ &= \Delta G_{cis}(A \rightarrow G) - \Delta G_{trans}(A \rightarrow G)^{[94]}\end{aligned}\tag{19}$$

$\Delta \Delta G = 2.67$ kcal/mol had a fair agreement with the experimental value of 2.4 kcal/mol, and it was concluded that the multiple minima problem^[95] in the conformational sampling becomes a major source of error as more complex systems are attempted for free energy calculations.^[94]

The *cis/trans* isomerisation of the N-methylated peptide bond of a bacterial collagenase peptide inhibitor (HS-CH₂-CH₂-CO-Pro-NMe-Ala) in water was investigated by the TI method.^[96] Thirteen independent molecular dynamics runs at 300 K were performed at fixed λ -values which were not uniformly distributed between 0 and 1 in that a larger number of λ -values were taken around the energy barrier, totally in 300 ps from $\lambda=0$ to 1.^[96] The resulting free-energy difference between the *trans* and *cis* isomers was 0.95 ± 0.5

kcal/mol in favor of the *trans* form, compared with 1.5 ± 0.5 kcal/mol by experiment, and it was concluded that this free-energy difference came mainly from the differences of solute-water electrostatic interactions around the carbonyl groups.^[96]

There has not shown any direct calculation of ΔG for *cis trans* isomerization of prolyl peptide bonds, but it is suggested that the FEP simulation approach will be useful in the prediction of ΔG for such conformational differences as well as in the rationalization and design of drugs and enzymes.^[82] However, much work remains to be done to overcome convergence problems before the computational capabilities and accuracy of these methods are fully developed and are applicable on a routine basis.^[82]

In this research project, the *cis trans* equilibria in water for the two prolyl dipeptides, Ace-Tyr-Pro-NMe and Ace-Asn-Pro-NMe (see **Table 1**), are predicted by calculating Gibbs free energy differences between the *cis trans* conformers using the combined perturbation and PMF method.

METHODS

The overall approach in these free energy simulations was the following. First, preliminary *in vacuo* free energy calculations (*cis* $\rightarrow$ *trans* direction only) were done, using Ace-Tyr-Pro-NMe, in AMBER version 4.0 (University of California, San Francisco) to estimate the number of calculation steps at each window and the overall window size that should be applied for the simulation with explicit water molecules. Then the Gibbs free energies (*cis* $\leftrightarrow$ *trans*) were calculated for the two dipeptides, Ace-Tyr-Pro-NMe and Ace-Asn-Pro-NMe, in a water-filled box using AMBER version 4.1 (University of California, San Francisco). All the calculations were run on a Silicon Graphics Personal Iris Model 4D/35. See Appendix A for a detailed description of the explicit water simulation.

Dipeptides, Ace-Tyr-Pro-NMe and Ace-Asn-Pro-NMe, were built in the QUANTA version 3.3.1 (Molecular Simulations, Inc.) Sequence Builder facility. After setting Tyr-Pro and Asn-Pro peptide bonds at the *cis* conformation using the constraint option, a conformational search for 500 conformers was performed for these peptides using the Random Sampling method, where each starting conformer was obtained by choosing the torsion angles in the peptide in a random fashion. Each of the generated conformers was then energy minimized in CHARMM using the Adopted-Basis Newton-Raphson optimizer for 1000 steps or until the energy gradient became lower than its tolerance of 0.001. The lowest-energy *cis* conformer was used as a starting structure for the FEP calculations for each of the dipeptides.

The *in vacuo* FEP calculations were prepared and run in AMBER (4.0) as follows: the peptide, Ace-Tyr-Pro-NMe, was first energy-minimized for 50 steps by a steepest descent method then for 450 steps by a conjugate gradient method in the MINMD module. A constant dielectric of 1 for the *in vacuo* simulations was employed in MINMD and in all the subsequent modules of AMBER. After minimization, the peptides were gradually warmed from 10 K to 300 K in 5 psec and equilibrated at 300 K for another 5 psec in the SANDER module. The SHAKE option, where all bond lengths are constrained, and a time step of 0.002 psec were utilized. Free energies were calculated in the GIBBS module with 100 ($\Delta\lambda = 0.01$), 500 ($\Delta\lambda = 0.002$), 1000 ($\Delta\lambda = 0.001$), and 10000 ($\Delta\lambda = 0.0001$) windows with different numbers of equilibration and data collection steps. A default non-bonded cutoff of 8 Å for the intermolecular interactions was used, and the non-bonded pair list was updated every 10 steps. For all runs, double-wide sampling was applied so that at each window, free energy was calculated in both the forward and the backward directions along the path. The temperature was kept constant at 300 K.

Explicit water simulations were run in AMBER (4.1) as follows: first, each peptide was placed in a TIP3P water-filled box (618 water molecules for Ace-Tyr-Pro-NMe and 621 for Ace-Asn-Pro-NMe), and energy-minimization was done only for water molecules to eliminate high-energy spots in the box. Then the whole system was energy-minimized in the same manner as before in the SANDER module. Before the slow-warming step, the whole box was equilibrated by a SANDER dynamics run for 10 psec at 10 K. Then it was slowly heated up to 300 K in 10 psec and subsequently equilibrated again at 300 K for 110 psec to

stabilize pressure in the box. Gibbs free energies were calculated in three steps: *cis* → *trans* run, equilibration at *trans*, and *trans* → *cis* reverse run. The *cis* → *trans* and *trans* → *cis* runs were done with 1000 windows (2 psec per window) and at each window 500 steps of equilibration and another 500 steps of data collection were performed. Between the *cis* → *trans* and *trans* → *cis* runs the system was equilibrated at the *trans* conformation for 2 psec without data collection steps. Pressure was kept constant at 1 bar and the 'fast water' algorithm was used to shorten the computing time. In addition all the options in the preliminary *in vacuo* FEP runs were applied here.

RESULTS

In Vacuo Simulation

Preliminary *in vacuo* free energy calculations were done in AMBER 4.0 in the *cis* → *trans* direction using Ace-Tyr-Pro-NMe to estimate the number of MD steps at each window and the overall window size which would be used for the explicit water simulation. **Figures 1 through 4** show *in vacuo* free energy profiles of Ace-Tyr-Pro-NMe with different numbers of steps of equilibration and data collection and with different numbers of windows. A summary of the graph data is presented in **Table 2**.

One of the common trends seen in **Figures 1 through 3** (100, 500, and 1000 windows, respectively) is that the forward paths with the (500/5000), (5000/500), and (5000/5000) options agreed well within the same window option, but the (500/500) option had different forward profiles compared to the others. The similar trend was also found in the backward paths. This could imply that (500/500) runs were unable to converge as well as the other runs with larger numbers of steps. It is interesting to see that a ten-times increase in equilibration steps (5000/500) had an almost equivalent effect on the free energy profile as a ten-times increase in data collection steps (500/5000) with 100, 500, and 1000 windows. No runs with numbers of steps larger than 500 were performed with 10000 windows (**Figure 4**) since it would not be feasible to run subsequent calculations with explicit waters.

Figure 1 (100 windows) shows no apparent convergence in free energy in any of the step options. With 500 and 1000 windows (**Figures 2 and 3**), however, improvement is

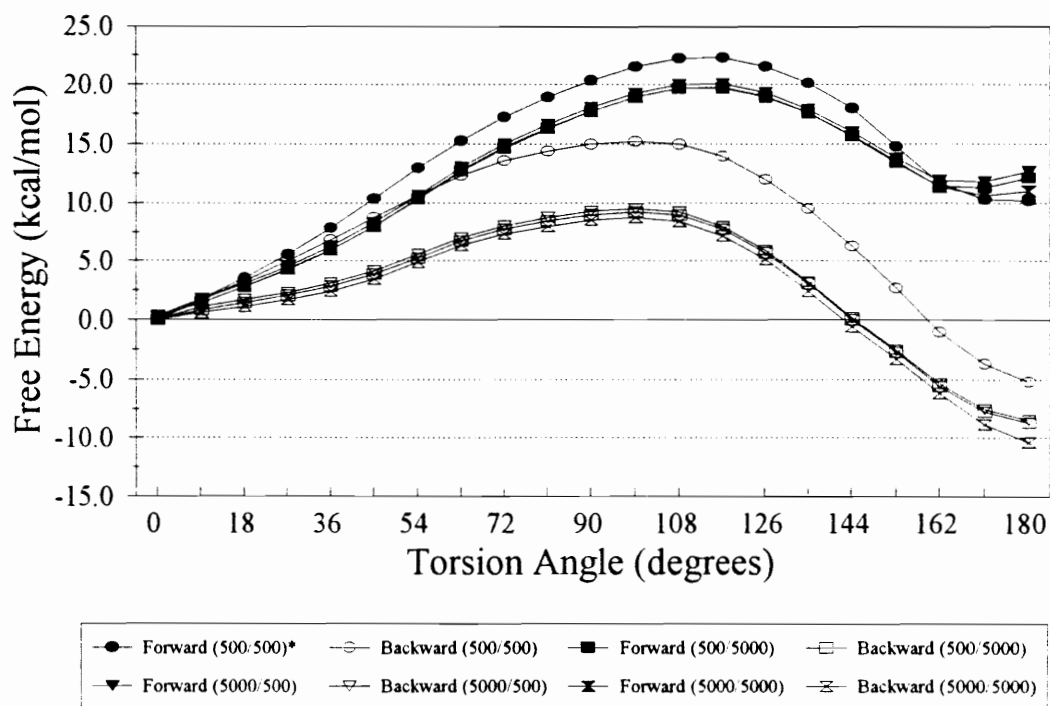


Figure 1 Free Energy Profile of Ace-Tyr-Pro-NMe by FEP with 100 windows *in vacuo* (*cis-trans*)
 * (number of MD steps in equilibration/number of MD steps in data collection) per window

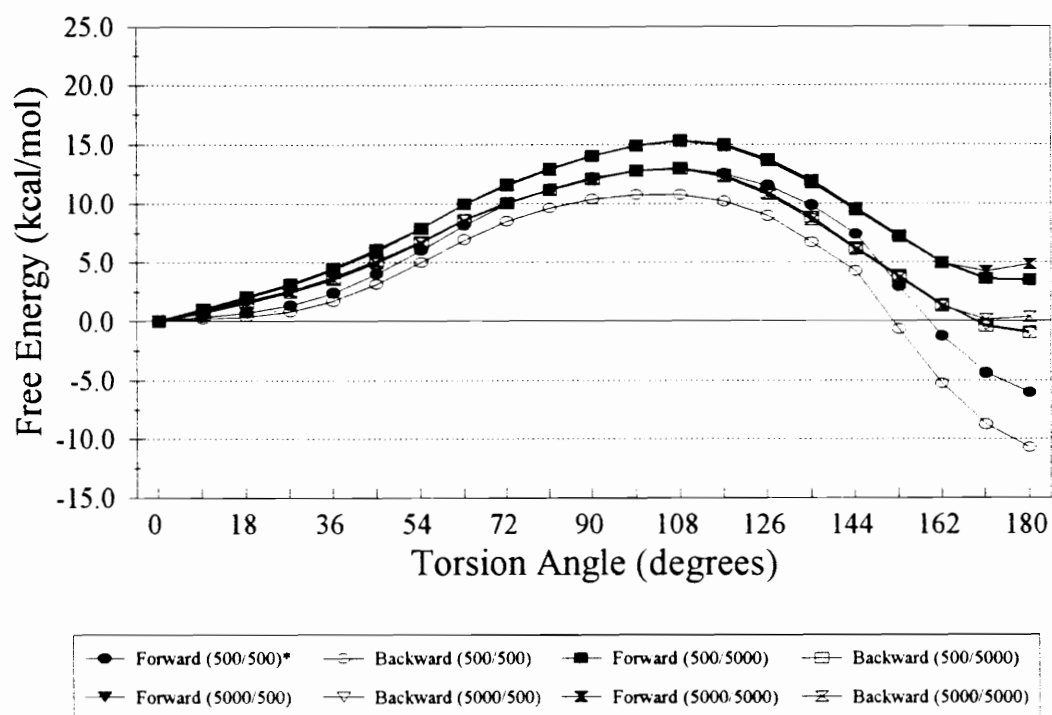


Figure 2 Free Energy Profile of Ace-Tyr-Pro-NMe by FEP with 500 windows *in vacuo* (*cis-trans*)
 * (number of MD steps in equilibration/number of MD steps in data collection) per window

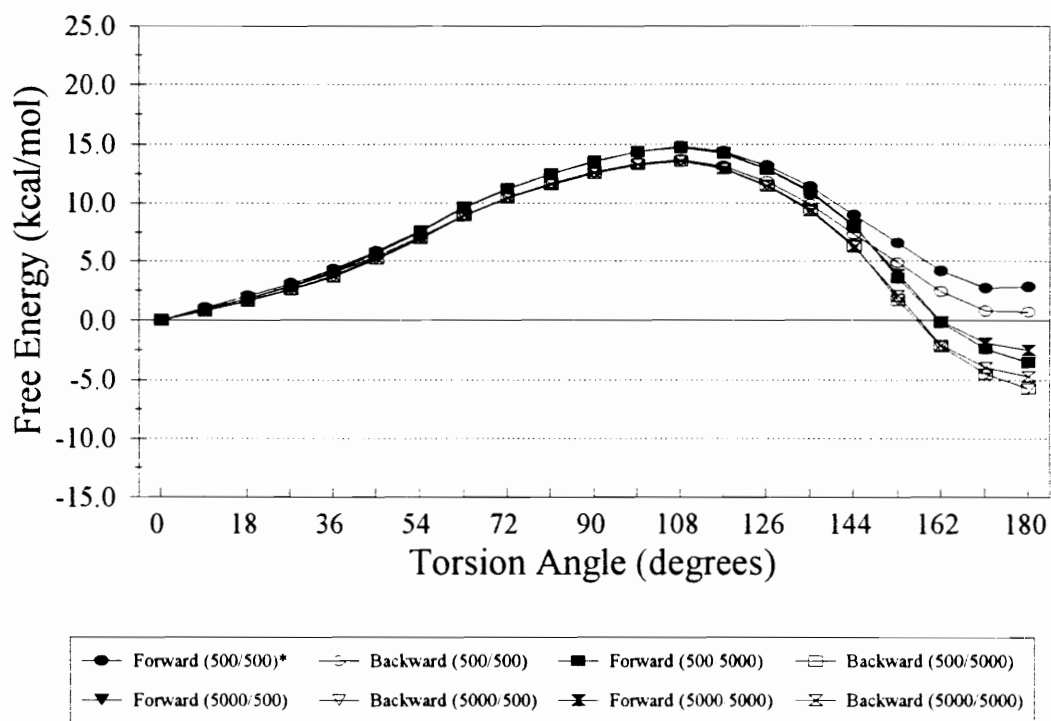


Figure 3 Free Energy Profile of Ace-Tyr-Pro-NMe by FEP with 1000 windows *in vacuo* (*cis-trans*)
 * (number of MD steps in equilibration/number of MD steps in data collection) per window

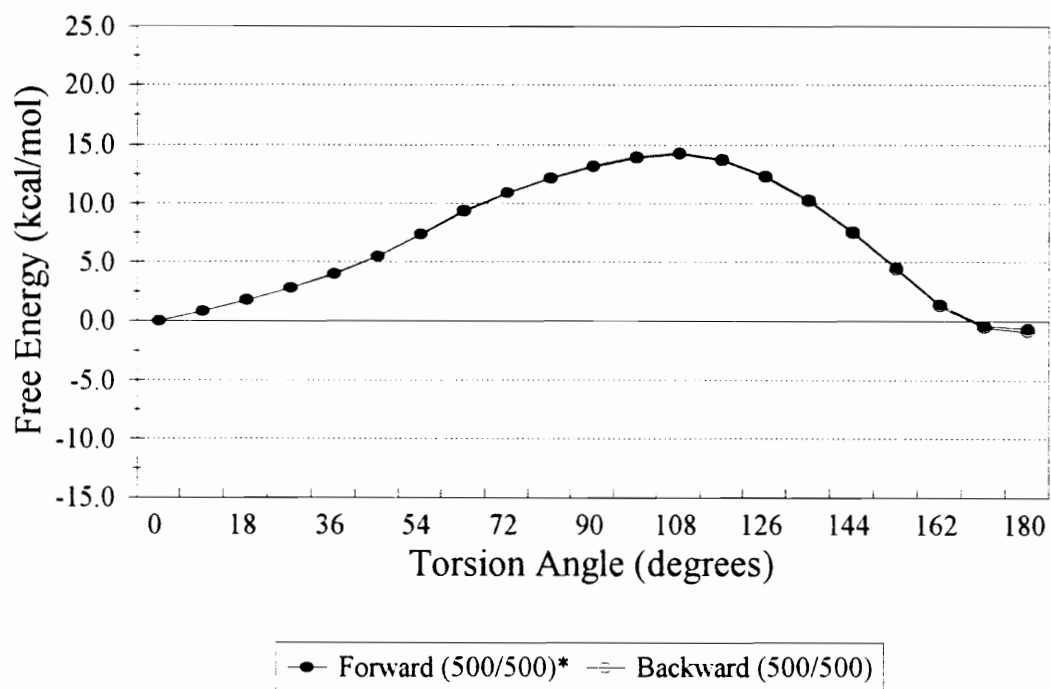


Figure 4 Free Energy Profile of Ace-Tyr-Pro-NMe by FEP with 10000 windows *in vacuo* (*cis-trans*)
 * (number of MD steps in equilibration/number of MD steps in data collection) per window

Table 2 Comparison of Free Energy Difference (*cis* $\rightarrow$ *trans*) for Ace-Tyr-Pro-NMe *in vacuo* by FEP (AMBER 4.0) with different numbers of windows, equilibration steps and data collection steps (Experimental = -0.34 kcal/mol)

		Free Energy Difference (kcal/mol)			RSD [#] (%)
# of Windows		Forward	Backward	Mean	
100	(500/500)*	10.20	-5.20	2.50	435.2
	(500/5000)	12.16	-8.70	1.73	852.6
	(5000/500)	12.68	-8.46	2.11	708.5
	(5000/5000)	11.00	-10.38	0.31	4908.9
500	(500/500)	-6.05	-10.78	-8.41	39.8
	(500/5000)	3.39	-0.97	1.21	255.1
	(5000/500)	3.52	-0.87	1.32	234.5
	(5000/5000)	4.81	0.34	2.57	123.0
1000	(500/500)	2.85	0.74	1.80	83.1
	(500/5000)	-3.54	-5.76	-4.65	33.8
	(5000/500)	-3.53	-5.75	-4.64	33.9
	(5000/5000)	-2.47	-4.70	-3.58	43.8
10000	(500/500)	-0.62	-0.83	-0.73	20.6

* (# of MD steps in equilibration / # of MD steps in data collection) per window

[#] Relative Standard Deviation

obvious, and relative standard deviations (RSDs) were much smaller than those in 100 windows (**Table 2**), which was further narrowed in the 10000 window simulation (**Figure 4**). On the other hand, overall free energies of activation ($\Delta G^\ddagger$) did not seem to vary much among different window options: the averages of $\Delta G^\ddagger$ from both *cis* and *trans* ground state to the excited state in both forward and backward directions were 14.7, 14.0, 15.5, and 14.6 kcal/mol for the 100, 500, 1000, and 10000 window options, respectively.

Comparing the free energy values from 1000 windows with the (5000/5000) option and 10000 windows with the (500/500) option, in which 10,000,000 overall steps were taken in each run, the 10000-window option seemed to provide much better convergence in that it had lower RSD (20.6% vs 43.8%) and better free energy difference (-0.73 kcal/mol vs -3.58 kcal/mol; compared to the experimental value of -0.34 kcal/mol). The same can be said with 100 window (5000/5000) and 1000 window (500/500) options (1,000,000 overall steps in each run). This suggests that increasing the number of windows might be more critical than increasing the number of steps in each window in terms of free energy convergence and accuracy.

From initial studies, we knew that an explicit water simulation in the *cis* - *trans* direction with Ace-Tyr-Pro-NMe in AMBER 4.0 would take approximately two months to complete with the (500/500) step option in 1000 windows. Our preliminary *in vacuo* runs also indicated that increasing the number of steps for equilibration or data collection ten times could result in an approximately five-times increase in run time. Considering the available computer resources, our test simulation results, and time restrictions, the option of 1000

windows with (500/500) steps was chosen for our subsequent FEP runs with explicit water molecules. Therefore, the total FEP run time for our two peptides in both *cis* $\rightarrow$ *trans* and *trans* $\rightarrow$ *cis* directions was expected to be up to approximately eight months.

Explicit Water Simulations

The explicit water runs using AMBER 4.1 are summarized in **Table 3**, and Gibbs free energy profiles for each of the dipeptides are shown in **Figures 5** and **6**. Simulations were performed with 1000 windows where 500 equilibration and 500 data collection steps were taken in each window for Ace-Tyr-Pro-NMe and Ace-Asn-Pro-NMe. In both cases, the calculated Gibbs free energy differences were much lower than the experimental values, resulting in very low %*cis* values. Also, overall *cis* $\rightarrow$ *trans* $\rightarrow$ *cis* runs had Gibbs free energy differences of 0.93 kcal/mol for Ace-Tyr-Pro-NMe and 0.22 kcal/mol for Ace-Asn-Pro-NMe, indicating possible hysteresis since they were expected to be 0 kcal/mol. In addition, overall free energies of activation ($\Delta G^\ddagger$) were much lower than those from the previous *in vacuo* runs. The averages of $\Delta G^\ddagger$ were 8.9 and 8.6 kcal/mol for Ace-Tyr-Pro-NMe and Ace-Asn-Pro-NMe, respectively. According to the ^1H NMR study with N-ethyl-N-methyl amides (which often are used to model peptide bonds in proteins), liquid-phase values of $\Delta G^\ddagger$ are expected to be approximately 16-21 kcal/mol and to be higher than the gas-phase values by 1.4-3.1 kcal/mol.^[97] Also, $\Delta G^\ddagger$ for a peptide bond in proteins has been suggested to be ca. 20 kcal/mol.^[9] Although our gas-phase and aqueous-phase values of $\Delta G^\ddagger$ cannot directly be compared with each other, it seems that our simulation could not reproduce experimental

Table 3 Gibbs Free Energy Change for Ace-Tyr-Pro-NMe and Ace-Asn-Pro-NMe in water by FEP (AMBER 4.1) with 1000 windows (500 equilibration and 500 data collection steps in each window). All the values are shown as *cis* $\rightarrow$ *trans* change.

Ace-Tyr-Pro-NMe (Experimental = -0.34 kcal/mol; %*cis* = 36)^[69]

Direction	Gibbs Free Energy Change (kcal/mol)			RSD (%)	% <i>cis</i>
	Forward	Backward	Mean		
<i>cis</i> $\rightarrow$ <i>trans</i>	-3.26	-3.37	-3.32	2.5	0.37
<i>trans</i> $\rightarrow$ <i>cis</i>	-4.28	-4.23	-4.25	0.8	0.08
Overall <i>cis</i> $\rightleftharpoons$ <i>trans</i>			-3.78	14.3	0.17

Ace-Asn-Pro-NMe (Experimental = -1.07 kcal/mol; %*cis* = 14)^[69]

Direction	Gibbs Free Energy Change (kcal/mol)			RSD (%)	% <i>cis</i>
	Forward	Backward	Mean		
<i>cis</i> $\rightarrow$ <i>trans</i>	-3.74	-3.76	-3.75	0.4	0.18
<i>trans</i> $\rightarrow$ <i>cis</i>	-4.01	-3.92	-3.97	1.6	0.13
Overall <i>cis</i> $\rightleftharpoons$ <i>trans</i>			-3.86	3.4	0.15

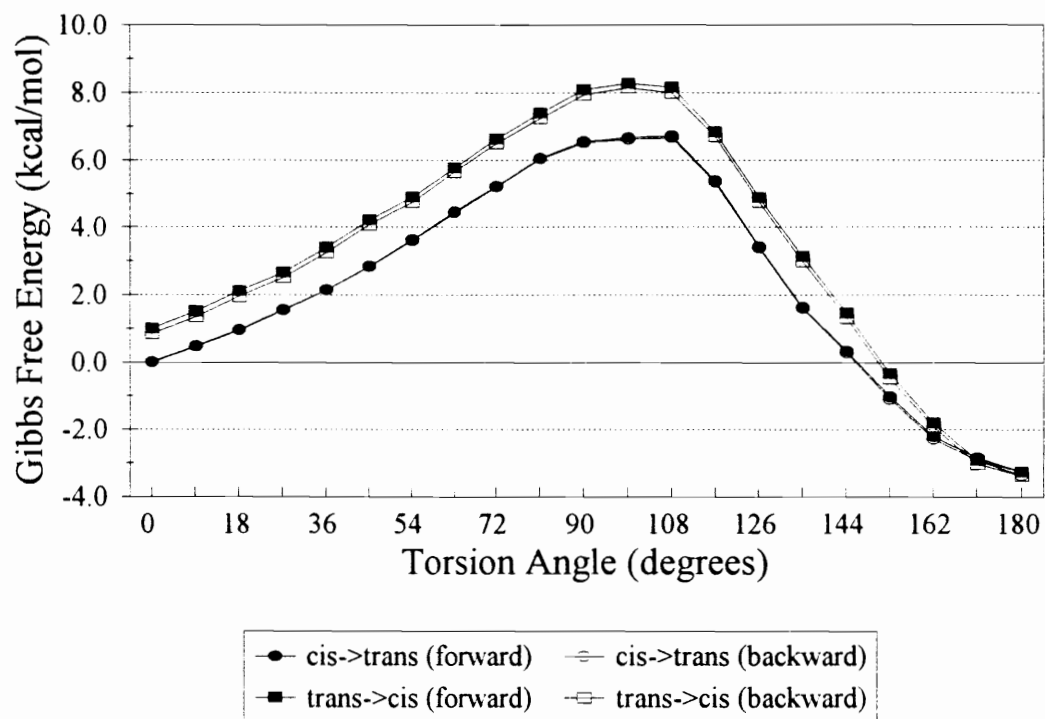


Figure 5 Gibbs Free Energy Profile of Ace-Tyr-Pro-NMe by FEP (AMBER 4.1) with 1000 windows in water (*cis-trans-cis*) 500 equilibration and 500 data collection steps in each window.

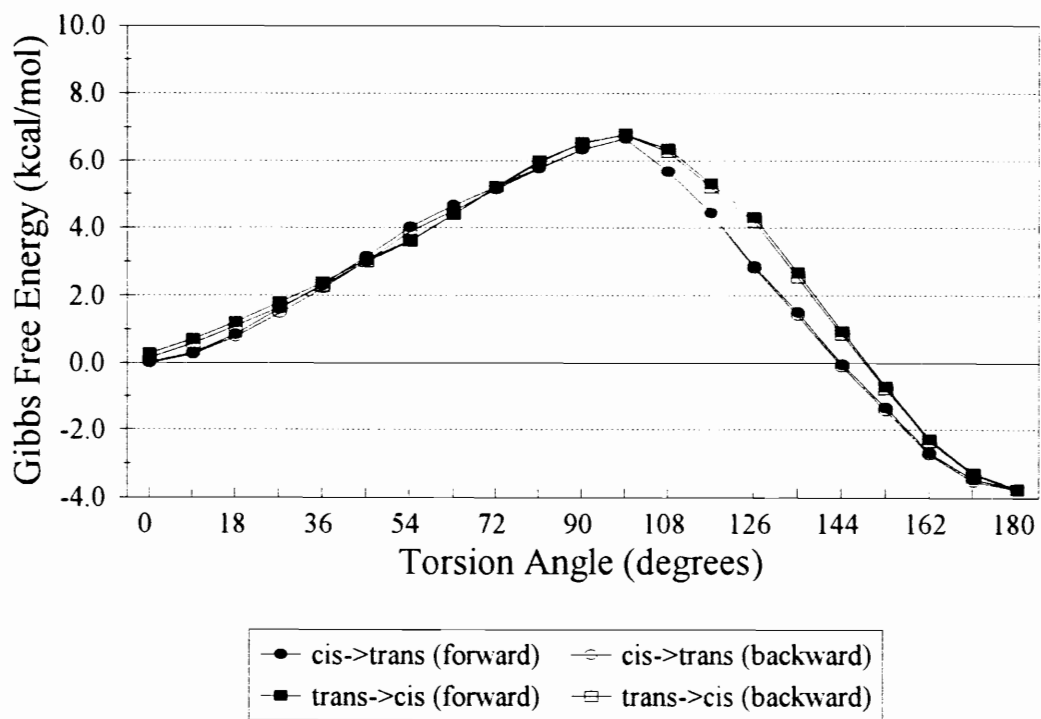


Figure 6 Gibbs Free Energy Profile of Ace-Asn-Pro-NMe by FEP (AMBER 4.1) with 1000 windows in water (*cis-trans-cis*) 500 equilibration and 500 data collection steps in each window.

values for the aqueous-phase $\Delta G^\ddagger$. However, precision between forward and backward values in either direction was excellent (very low RSDs) and our simulation clearly showed preference of the *trans* conformer in both dipeptides (*trans* conformers have lower free energies than *cis*) and that %*cis* in overall average is higher in Ace-Tyr-Pro-NMe than in Ace-Asn-Pro-NMe, which agrees with the literature.^[69]

Atomic Charge Differences among Conformers of Ace-Tyr-Pro-NMe

Because in AMBER 4.1 the Hamiltonian of our peptides were not altered as the torsion angle was changed, atomic charges of the peptides were calculated by MOPAC 6.0 with the ESP method (see PART II and APPENDIX C for detailed method) to see how parameters like atomic charges could differ depending on the torsion angle of the peptide bond.

RSD values in **Table 4** indicate that many of the atoms of the dipeptide do have quite different levels of atomic charges which are dependent on the peptide torsion angle of the dipeptide. It is interesting that atoms at the *N*-methyl group were much more affected by the isomerization (higher RSDs) than those at the acetyl group and that overall charge differences in the prolyl residue were greater than those in the tyrosyl residue, indicating unique influence of the prolyl residue over molecular properties upon conformational change.

See APPENDIX B for additional calculation results.

Table 4 Atomic Charges of Ace-TYR-PRO-NMe by MOPAC 6.0

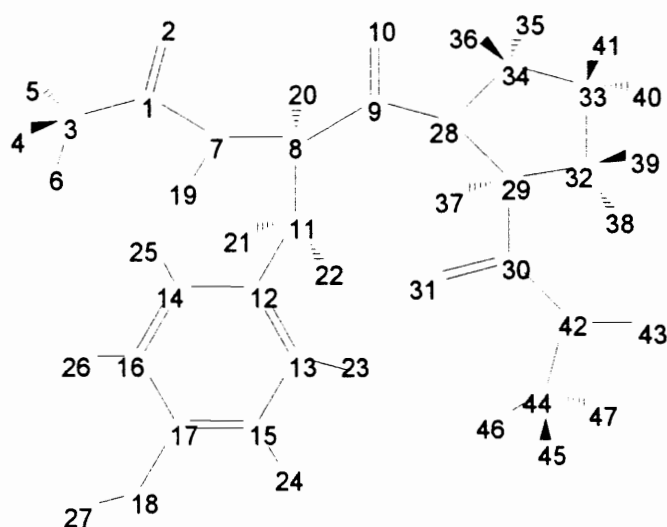
Atom	Location**	Residue	Atomic Charge by MOPAC6.0*			RSD (%)
			Torsion Angle			
			<i>cis</i>	90°	<i>trans</i>	
H	1	ACE	0.1712	0.1756	0.1785	2.10
C	2	ACE	-0.5218	-0.5320	-0.5426	1.95
H	3	ACE	0.1704	0.1719	0.1721	0.54
H	4	ACE	0.1676	0.1744	0.1773	2.88
C	5	ACE	0.4576	0.4459	0.4462	1.48
O	6	ACE	-0.3742	-0.3562	-0.3716	2.65
N	7	TYR	-0.6201	-0.6033	-0.5603	5.19
H	8	TYR	0.3621	0.3505	0.3472	2.22
C	9	TYR	0.2761	0.2705	0.2693	1.33
H	10	TYR	0.1646	0.0888	0.1188	30.77
C	11	TYR	-0.6611	-0.5847	-0.6382	6.24
H	12	TYR	0.2054	0.1889	0.1994	4.22
H	13	TYR	0.2169	0.1909	0.2233	8.16
C	14	TYR	0.2520	0.1998	0.2505	12.69
C	15	TYR	-0.1998	-0.2414	-0.2992	20.23
H	16	TYR	0.2009	0.2040	0.2324	8.17
C	17	TYR	-0.3652	-0.2378	-0.2672	23.00
H	18	TYR	0.2137	0.2223	0.2002	5.25
C	19	TYR	0.1510	0.1162	0.0866	27.33
O	20	TYR	-0.2863	-0.2822	-0.2769	1.67
H	21	TYR	0.2633	0.2716	0.2642	1.71
C	22	TYR	-0.2564	-0.3270	-0.2004	24.28
H	23	TYR	0.2377	0.2157	0.2142	5.91
C	24	TYR	-0.3089	-0.2355	-0.2905	13.72
H	25	TYR	0.2408	0.2072	0.2118	8.28
C	26	TYR	0.0472	0.1456	-0.0014	117.39
O	27	TYR	-0.2814	-0.2374	-0.3041	12.36
N	28	PRO	-0.1453	-0.1628	-0.0328	62.08
C	29	PRO	-0.2057	-0.2258	-0.1239	29.15
H	30	PRO	0.1074	0.1153	0.1025	5.96
H	31	PRO	0.1335	0.1210	0.1381	6.76
C	32	PRO	-0.0945	-0.0975	-0.2458	59.27
H	33	PRO	0.0832	0.0886	0.1270	23.98
H	34	PRO	0.0887	0.0902	0.1091	11.84

Table 4 - continuation

Atom	Location**	Residue	Atomic Charge by MOPAC6.0			RSD (%)
			Torsion Angle			
			<i>cis</i>	90°	<i>trans</i>	
C	35	PRO	-0.3842	-0.3816	-0.1849	36.08
H	36	PRO	0.1575	0.1303	0.1167	15.41
H	37	PRO	0.1328	0.1710	0.0952	28.50
C	38	PRO	0.3303	0.1755	0.0380	80.67
H	39	PRO	0.0531	0.0975	0.1015	31.97
C	40	PRO	-0.0437	0.1034	0.1441	145.43
O	41	PRO	-0.2663	-0.3429	-0.2694	14.80
N	42	NME	-0.2462	-0.1236	-0.3830	51.71
H	43	NME	0.2673	0.1943	0.3240	24.83
C	44	NME	-0.3254	-0.4142	-0.2323	28.08
H	45	NME	0.1499	0.1322	0.0934	23.09
H	46	NME	0.1732	0.1538	0.1014	26.01
H	47	NME	0.1108	0.1733	0.1411	22.05

* Atomic charges were calculated using the ESP method. See PART II for details.

** See the figure below.



DISCUSSION

Our simulation failed to obtain ΔG values close to the experimental values for both dipeptides examined. Some of the possible problems in our modeling schemes, which could have caused the discrepancies, are discussed below.

Hysteresis

Although the free energy perturbation procedure is exact when simulation averages are taken over runs of infinite length, for a finite sample size, a systematic error could result in hysteresis.^[98] One type of hysteresis error is that two opposite runs (e.g., *cis* $\rightarrow$ *trans* and *trans* $\rightarrow$ *cis*) follow different paths when the structural relaxation time is longer than the run time in one window (time-lag hysteresis). The other error type occurs when the accessed conformational substates do not sufficiently represent the full state during the window run (sample-size hysteresis).^[98] In our calculation, the possibility of having time-lag hysteresis is very slight because the window-type runs are known to have no time-lag hysteresis as long as the number of equilibration steps is large enough.^[98]

On the other hand, sample-size hysteresis is very common. Even though the ergodic hypothesis states that time averages and ensemble averages are identical for a system in thermodynamic equilibrium, it is very difficult to sample all the conformational space (ergodic convergence) because the time required for an N-particle system to cover all points in the conformational space scales as $\exp(N)$.^[99] A quantitative measure of this ergodic

convergence, which is based on statistical symmetry in that all particles of a given type have identical average characteristics,^[99] has been suggested and applied for liquids,^[99] a peptide, and an enzyme/product complex.^[100,101] However, this is a necessary but not a sufficient condition for ergodicity, because convergence might be happening only in one of the conformational substates,^[99] and there has not been any measure which ensures true ergodic convergence for a given system for the FEP simulation. Nevertheless, it has been speculated that 100 ns may be the upper limit for the convergence time for proteins.^[102] Considering the current computational time limit (1 ns of total data sampling time for 1000-window simulation), even though our peptides are considerably smaller, it might be very difficult to achieve an ergodic convergence in our system.

To circumvent this sampling problem, the umbrella sampling technique^[103] was developed. Umbrella sampling utilizes an artificial biasing potential (umbrella potential), which biases the sampling toward the favorable regions of the conformational space, and the probability distribution resulting from the simulation can be corrected afterwards to yield the probability distribution of the corresponding unbiased system.^[104] This technique has been often chosen because it could overcome large energy barriers, which prevent efficient sampling with the potential of mean force option,^[104] resulting in less chance to have sample-size hysteresis. However, this option is not available in the current version of AMBER.

Nonbonded Cutoff

To reduce the computational demand of the FEP method, MD simulations for free

energy calculations are often carried out using some drastic simplifications. One of them is that the long-range Coulomb interactions are truncated at shorter distances to decrease the number of nonbonded interacting pairs to be calculated.^[105] Our simulation employed 8 Å-cutoff, which is commonly used and also was utilized for developing the TIP3P (Transferable Intermolecular Potential function, 3-Point model)^[106] water potential model.^[107] Although the cutoff distance must be no greater than 1/2 of the simulation box side length,^[108] it is well known that energetic effects or conformational stability could be dependent on the cutoff size of the simulation.^[109,110] Those cutoff dependencies were found in much larger systems (protease-inhibitor complex^[109] and a 17-residue model peptide^[110]) than ours, but yet peptides are polar molecules and so the ignored long-range Coulomb interactions could be important for the accuracy of the simulation.

Some methods have been developed in which the long-range non-bonded interactions are considered without a significant increase in computation time. The twin cutoff method utilizes a regular short-range non-bonded cutoff as well as a long-range electrostatic correction every time the non-bonded list is updated,^[111] which was proved to be important for stabilizing a protein structure over a long simulation.^[112] Similarly, a combination of a long-range cutoff for the interaction energy of the perturbed group with any other residue and a short-range cutoff for all other interactions has been successfully used for an enzyme-peptide complex with charged residues.^[109] On the other hand, the Ewald method^[113] and the PPPM (particle-particle, particle-mesh) method^[114] both split the electrostatic potential into the sum of a short-range and a long-range function, where the Ewald method handles the

long-range interaction by the Fourier Transform analytically but the PPPM method uses the numerical Fast Fourier Transform techniques, which was shown to be more computationally efficient.^[115] A modified PPPM method has also been suggested, where the long-range interactions are estimated by using a high dielectric continuum to mimic the surrounding explicit system.^[116] Furthermore, the PPPC (particle-particle and particle-cell) method creates cubic cells of atoms which interact with near atoms as in usual methods with cutoff but with distant cells through their total charge and dipole moment within the cell, and the size of cells increases as a function of the distance from the atom.^[105] These alternative methods to conventional cutoffs should be worthwhile to consider for future FEP computations.

Bond-Stretching Correction

Another possible problem, which could influence our results, is that our simulation ignored the bond-stretching contribution of the rotated peptide bond. At 90° and 270° , the peptide bond between TYR (or ASN) and PRO is orthogonal and, therefore, the double-bond character due to the resonance interactions of the peptide bond should be much weaker. Thus, the peptide C-N bond should be longer and the C=O bond should be shorter than at 0° or 180° . In fact, a peptide C-N bond is $0.14 \text{ \AA}$ shorter than an N-C $_{\alpha}$ single bond and a peptide C=O bond is $0.02 \text{ \AA}$ longer than that of aldehydes and ketones.^[1] However, our simulation utilized the SHAKE option, in which all the bonds are constrained, to reduce the computation time. Therefore, the above mentioned peptide-bond stretching effect during the simulation was not incorporated into our free energy calculations. In fact, bond-stretching

contributions to the free energy difference were shown not to be negligible, particularly when the equilibrium bond lengths are being changed.^[117]

Pearlman and Kollman circumvented this problem by performing a mutation-type simulation (methane $\rightarrow$ neopentane) with bond constraints, followed by another complete simulation where only the bond lengths of interest were changed from their initial to final states by the potential of mean force method.^[117] In our case, possibly further dividing the runs at 90° and 270° and performing bond-stretching corrections at those angles (bond-lengthening) as well as at 180° and 360° (bond-shortening) could help improve our results and yet might not increase overall computing time considerably.

Path Dependency

Although the free energy difference is in principle independent of the pathway chosen, in practice it has been shown that the choice of pathway could affect the accuracy of the free energy difference.^[118] In our runs, the overall Gibbs free energy differences for the *cis* $\rightarrow$ *trans* $\rightarrow$ *cis* runs for both peptides were much larger than the average free energy differences between forward and backward runs; 0.93 vs. 0.08 kcal/mol and 0.22 vs. 0.06 kcal/mol for Ace-Tyr-Pro-NMe and Ace-Asn-Pro-NMe, respectively (see **Table 3**). Because rotation of the peptide bond was done for 360° during the simulation, the pathways *cis* $\rightarrow$ *trans* and *trans* $\rightarrow$ *cis* did actually take different pathways.

Table 3 shows that within individual runs RSDs between forward and backward directions were relatively small (i.e., 0.4% - 2.5%) for both peptides but according to **Figure**

5 the ΔG and $\Delta G^\ddagger$ values were quite different between two runs (i.e., ca. 1 kcal/mol) for Ace-Tyr-Pro-NMe and that Ace-Asn-Pro-NMe seems to have taken slightly different profiles between forward and backward runs (see **Figure 6**). Even though principally the Gibbs free energy differences should be the same for both *cis* $\rightarrow$ *trans* and *trans* $\rightarrow$ *cis* pathways, there might be some biases in our simulation possibly due to some of the conformations which occurred during the simulation in one of the pathways and required longer equilibration or data collection steps than the other conformations. This could have happened around the *trans* conformation of Ace-Tyr-Pro-NMe in **Figure 5**, where two curves started to deviate from each other around 165° but kept the similar shapes thereafter. In fact, it has been suggested that by choosing a pathway that minimizes the relaxation time of the system, it is possible to significantly reduce the time required to calculate the free energy differences between two states.^[118] In our system, this possibility of pathway bias has not been sought yet since AMBER 4.1 automatically chooses a 360°-rotational pathway.

Use of a Single Hamiltonian

Lastly, the most critical problem in our FEP simulation of dipeptides might be the fact that during the runs the Hamiltonian of our dipeptides was never changed along with λ . In usual mutation-type FEP simulations, the Hamiltonian of the molecule of interest is changed at each window with λ by interpolating its parameters as $[\lambda * P_I + (1 - \lambda) * P_F]$, where P_I = parameter of the initial state and P_F = parameter of the final state (single molecular topology method).^[186] On the other hand, the dual molecular topology method uses the equation, $H(\lambda)$

$= \lambda H_B + (1 - \lambda)H_A$, where H = Hamiltonian.^[86] However, in our PMF-type modeling, the initial state Hamiltonian was being used along with λ all the way to the final state. In other words, the program assumed that the Hamiltonian of the molecule would not change along with λ . In fact, the parameters for the proline residue from AMBER 4.1 are developed by averaging both *cis* and *trans* conformers, each of which has two different backbone conformations.^[119]

Currently, there is no option to do parameter interpolation with λ using the initial and final state Hamiltonians in AMBER 4.1 when torsion-PMF-type simulations are done. Considering that atomic charges of peptides could change substantially with torsion angle (**Table 4**) and that electrostatic energy occupies the largest part of the total energy of the molecule in our runs, the use of a single Hamiltonian throughout the PMF simulation could cause considerable bias in the free energy values. A method like combined quantum mechanical and molecular mechanical potential (QM/MM),^[93] where a molecule or a part of the molecule of interest is described quantum-mechanically and its environment is handled by molecular mechanics, should certainly be considered to overcome this problem. However, this approach could cost substantial amount of CPU time and might not be practical. Updating parameters in the molecular mechanics potential for a molecule of interest semiempirically at the beginning of each window might be an alternative to this approach.

In conclusion, our explicit water FEP simulation using two dipeptides, Ace-Tyr-Pro-NMe and Ace-Asn-Pro-NMe, could not reproduce experimental %*cis* values. Possible

problems in current FEP calculations were considered, which include sample-size hysteresis, nonbonded cutoff, bond-stretching correction, path dependency, and the use of a single Hamiltonian. However, it is encouraging that precision between forward and backward Gibbs free energy values in either directions was excellent and that our simulation clearly showed preference of the *trans* conformer within both peptides as well as higher *cis*-conformer preferences in Ace-Tyr-Pro-NMe than Ace-Asn-Pro-NMe, which agrees with the experimental results. As computer power increases and innovative techniques to solve/overcome the above-mentioned problems emerge, the FEP method will be more accurate and useful in predicting molecular states like *cis/trans* equilibria in prolyl peptides because its potential has already been proven in various biological systems.

REFERENCES

1. Voet, D. and Voet, J. G., *"Biochemistry,"* John Wiley and Sons, New York, 1990.
2. Yaron, A. and Naider, F., *Crit. Rev. in Biochem. and Mol. Biol.*, 28:31-81, 1993.
3. Blout, E. R., *Acc. Chem. Res.*, 9:106-113, 1976.
4. Ramachandran, G. N. and Mitra, A. K., *J. Mol. Biol.*, 107:85-92, 1976.
5. Momany, F. A., McGuire, R. F., Burgess, A. W., and Scheraga, H. A., *J. Phys. Chem.*, 79:2361-2381, 1975.
6. Piela, L., Nemethy, G., and Scheraga, H. A., *Biopolymers*, 26:1587-1600, 1987.
7. Zimmermann, S. S. and Scheraga, H. A., *Biopolymers*, 16:811-843, 1977.
8. Smith, J. A. and Pease, L. G., *"Reverse Turns in Peptides and Proteins (CRC Critical Reviews in Biochemistry 8),"* CRC Press, Boca Raton, 1980.
9. Ramachandran, G. N. and Sasisekharan, V., *Advan. Prot. Chem.*, 23:283-437, 1968.
10. Stewart, D. E., Sarkar, Atom, and Wampler, John E., *J. Mol. Biol.*, 214:253-260, 1990.
11. Heitman, J., Novva, N. Rao, and Hall, Michael N., *New Biol.*, 4:448-460, 1992.
12. MacKinnon, C., Veeraraghavan, S., Kreider, I., Allen, M. J., Liggins, J. R., and Nall, B. T., *"Chemical Aspects of Enzyme Biotechnology,"* edited by T. O. Baldwin *et al.*, Plenum Press, New York, 1990.
13. Schumacher, A. and Nordheim, A., *Clin. Investig.*, 70:773-779, 1992.
14. Schreiber, S. L., *Science*, 251:283-287, 1991.
15. Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T., and Schmid, F. X., *Nature*, 337:476-478, 1989.
16. Takahashi, N., Hayano, T., and Suzuki, M., *Nature*, 337:473-475, 1989.

17. Harding M. W., Galat, A., Uehling, D. E., and Schreiber, S. L., *Nature*, 341:758-760, 1989.
18. Siekierka, J. J., Hung, S. H. Y., Poe, M., Lin, C. S., and Sigal, N. H., *Nature*, 341:755-757, 1989.
19. Freedman, R. B., *Nature*, 341:692, 1989.
20. Koletsky, A. J., Harding, M. W., and Handschumacher, R. E., *J. Immunol.*, 137:1054-1059, 1986.
21. Hohman, R. J. and Hultsch, T., *New Biol.*, 2:663-672, 1990.
22. Friedman, J. and Weissman, I., *Cell*, 66:799-806, 1991.
23. Liu, J., Farmer, J. D., Jr., Lane, W. S., Friedman, J., Weissman, I., and Schreiber, S. L., *Cell*, 66:807-815, 1991b.
24. Emmel, E. A., Verweij, C. L., Durand, D. B., Higgins, K. M., Lacy, E., and Crabtree, G. R., *Science*, 246:1617-1620, 1989.
25. Rahfeld, J.-U., Schierhorn, A., Mann, K. and Fischer, G., *FEBS Lett.*, 343:65-69, 1994.
26. Kay, J. E., *Biochem. J.*, 314:361-385, 1996.
27. Stein, R. L., *Adv. Protein Chem.*, 44:1-24, 1993.
28. Kallen, J., Spitzfaden, C., Zurini, M. G. M., Wider, G., Widmer, H., Wüthrich, K., and Walkinshaw, M. D., *Nature*, 353:276-279, 1991.
29. High, K. P. and Handschumacher, R. E., *Infectious Agents and Disease*, 1:121-135, 1992.
30. Steinmann, B., Bruckner, P., and Superti-Furga, A., *J. Biol. Chem.*, 266:1299-1303, 1991.
31. Lodish, H. F. and Kong, N., *J. Biol. Chem.*, 266:14835-14838, 1991.
32. Fransson, C., Freskgård, P. O., Hebertsson, H., Johansson, Å., Jonasson, P., Mårtensson, L.-G., Svensson, M., Jonsson, B.-H., and Carlsson, U., *FEBS Lett.*,

296:90-94, 1992.

33. Lin, L. N. and Brandts, J. F., *Biochemistry*, 18:5037-5042, 1979.
34. Lin, L. N. and Brandts, J. F., *Biochemistry*, 18:43-47, 1979.
35. Fischer, G., Bang, H., Berger, E., and Schellenberger, A., *Biochim. Biophys. Acta*, 791:87-97, 1984.
36. Lin, L.-N. and Brandts, J. F., *Biochemistry*, 22:4480-4485, 1983.
37. Yaron, A., *Biopolymers*, 26:S215-S222, 1987.
38. Fischer, G. and Schmid, F. X., *Biochemistry*, 29:2205-2212, 1990.
39. Anfinsen, C. B., *Science*, 181:223-230, 1973.
40. Hartl, F.-U., Hlodan, R., and Langer, T., *TIBS*, 19:20-25, 1994.
41. Kelley, W. L. and Georgopoulos, C., *Curr. Opini. Cell Biol.*, 4:984-991, 1992.
42. Cleland, J. L., "*Protein Folding In Vivo and In Vitro*," edited by Cleland, J. L., ACS, Washington, DC, 1993.
43. Brandts, J. F., Halvorson, H. R., and Brennan, M., *Biochemistry*, 14:4953-4963, 1975.
44. Kiefhaber, T., Grunert, H.-P., Hahn, U., and Schmid, F. X., *Biochemistry*, 29:6475-6480, 1990.
45. Mayr, L. M., Kiefhaber, T., and Schmid, F. X., "*Protein Folding In Vivo and In Vitro*," edited by Cleland, J. L., ACS, Washington, DC, 1993.
46. Schreiber, G. and Fersht, A. R., *Biochemistry*, 32:11195-11203, 1993.
47. Boyd, D. B., "*Reviews in Computational Chemistry*," volume 1, edited by Lipkowitz, K. B. and Boyd, D. B., VCH Publishers, Inc., New York, 1990.
48. Foresman, J. B. and Frisch, A., "*Exploring Chemistry with Electronic Structure Methods: A guide to Using Gaussian*," Gaussian, Inc., Pittsburgh, PA, 1993.

49. Bowen, J. P. and Allinger, N. L., *"Reviews in Computational Chemistry,"* volume 2, edited by Lipkowitz, K. B. and Boyd, D. B., VCH Publishers, Inc., New York, 1991.
50. Clark, T., *"A Handbook of Computational Chemistry: A Practical Guide to Chemical Structure and Energy Calculations,"* John Wiley and Sons, Inc., New York, 1985.
51. *"HyperChem™ Computational Chemistry,"* Autodesk, Inc., 1992.
52. Kollman, P. A., *Ann. Rev. Phys. Chem.* 38:303-16, 1987.
53. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, Jr., S., and Weiner, P., *J. Am. Chem. Soc.*, 106:765-784, 1984.
54. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M., *J. of Comp. Chem.*, 4:187-217, 1983.
55. Allinger, N. L., *J. Am. Chem. Soc.*, 99:8127-8134, 1977.
56. Jorgensen, W. L. and Tirado-Rives, J., *J. Am. Chem. Soc.*, 110:1657-1666, 1988.
57. Levine, I. N., *"Quantum Chemistry,"* 4th ed., Prentice-Hall, Inc., New Jersey, 1991.
58. Hehre, W. J., Radom, L., Schleyer, P. v.R., Pople, J. A., *"Ab Initio Molecular Orbital Theory,"* John Wiley and Sons, Inc., New York, 1986.
59. Sordo, T. L., Barrientos, C., and Sordo, J. A., *"Computational Chemistry: Structures, Interactions and Reactivity,"* edited by S. Fraga in "Studies in Physical and Theoretical Chemistry," volume 77(A), Elsevier Science Publishers B. V., 1992.
60. Stewart, J. J. P., *J. of Comp.-Aided Mol. Des.*, 4:1-103, 1990.
61. Pople, J. A., Santry, D. P., and Segal, G. A., *J. of Chem. Phys.*, 43:s129-s135, 1965.
62. Pople, J. A., Beveridge, D. L., and Dobosh, P. A., *J. of Chem. Phys.*, 47:2026-2033, 1967.
63. Zerner, M. C., *"Reviews in Computational Chemistry,"* volume 2, edited by Lipkowitz, K. B. and Boyd, D. B., VCH Publishers, Inc., New York, 1991.
64. Stewart, J. J. P., *"Reviews in Computational Chemistry,"* volume 1, edited by

- Lipkowitz, K. B. and Boyd, D. B., VCH Publishers, Inc., New York, 1990.
65. Dewar, M. J. S. and Thiel, W., *J. Am. Chem. Soc.*, 99:4899- 4917, 1977.
66. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P., *J. Am. Chem. Soc.*, 107:3902-3909, 1985.
67. Stewart, J. J. P., *J. Comput. Chem.*, 10:221-264, 1989.
68. Grathwohl, C. and Wüthrich, K., *Biopolymers*, 15:2025-2041, 1976.
69. Stimson, E. R., Montelione, G. T., Meinwald, Y. C., Rudolph, R. K. E., and Scheraga, H. A., *Biochemistry*, 21:5252-5262, 1982.
70. Blomberg, F., Rüterjans, H., Lintner, K., Toma, F., and Femandjian, S., *Org. Magn. Reson.*, 11:598-602, 1978.
71. Grathwohl, C. and Wüthrich, K., *Biopolymers*, 20:2623-2633, 1981.
72. Hunston, R. N., Gerothanassis, I. P., and Lauterwein, J., *J. Am. Chem. Soc.*, 107:2654-2661, 1985.
73. Brandts, J. F. and Lin, L.-N., "*Methods in Enzymology*," vol. 131, edited by Colowick, S. P. and Kaplan, N. O., Academic Press, Inc., Orlando, Fl, 1986.
74. Lin, L.-N. and Brandts, J. F., *Biochemistry*, 22:559-563, 1983.
75. Beveridge, D. L. and DiCapua, F. M., *Annu. Rev. Biophys. Biophys. Chem.*, 18:431-492, 1989.
76. Mezei, M. and Beveridge, D. L., "*Computer Simulation of Chemical and Biomolecular Systems*," edited by Beveridge, D. L. and Jorgensen, W. L., *Annals of the New York Academy of Science*, vol. 482, New York Academy of Science, NY, 1986.
77. McCammon, J. A., Gelin, B. R., and Karplus, M., *Nature*, 267:585-590, 1977.
78. Brooks III, C. L. and Case, D. A., *Chem. Rev.*, 93:2487-2502, 1993.
79. McCammon, J. A. and Harvey, S. C., "*Dynamics of Proteins and Nucleic Acids*," Cambridge University press, Cambridge, 1987.

80. Pearlman, D. A. and Kollman, P. A., "*Computer Simulation of Biomolecular Systems*," edited by van Gunsteren, W. F. and Weiner, P. K., ESCOM, Leiden, 1989.
81. Bennet, C. H., *J. Comp. Phys.*, 22:245-268, 1976.
82. Brooks III, C. L., "*Computer Simulation of Biomolecular Systems*," edited by van Gunsteren, W. F. and Weiner, P. K., ESCOM, Leiden, 1989.
83. Pearlman, D. A., *J. Chem. Phys.*, 98:8946-8957, 1993.
84. Pearlman, D. A. and Kollman, P. A., *J. Chem. Phys.*, 90:2460-2470, 1989.
85. Beveridge, D. L. and DiCapua, F. M., "*Computer Simulation of Biomolecular Systems*," edited by van Gunsteren, W. F. and Weiner, P. K., ESCOM, Leiden, 1989.
86. Kollman, P. A., *Chem. Rev.*, 93:2395-2417, 1993.
87. Berendsen, H. J. C., "*Proteins: Structure, Dynamics, and Design*," edited by Renugopalakrishnan, V., Carey, P. R., Smith, I. C. P., Huang, S.-G., and Storer, A. C., ESCOM, Leiden, 1991.
88. Bash, P. A., Singh, U. C., Brown, F. K., Langridge, R., Kollman, P. A., *Science*, 235:574-576, 1987.
89. Hirono, S. and Kollman, P. A., *J. Mol. Biol.*, 212:197-209, 1990.
90. Cummins, P. L. and Gready, J. E., *J. Comp.-Aided Mol. Des.*, 7:535-555, 1993.
91. Duffy, E. M., Severance, D. L., and Jorgensen, W. L., *J. Am. Chem. Soc.*, 114:7535-7542, 1992.
92. Gao, J., *J. Phys. Chem.*, 96:537-540, 1992.
93. Gao, J., *J. Am. Chem. Soc.*, 115:2930-2935, 1993.
94. Hodel, A., Simonson, T., Fox, R. O., and Brünger, A. T., *J. Phys. Chem.*, 97:3409-3417, 1993.
95. Elber, R. and Karplus, M., *Science*, 235:318-321, 1987.
96. Gilquin, B., Perahia, D., and Dive, V., *Int. J. Peptide Protein Res.*, 43:384-392, 1994.

97. Suarez, C., Tafazzoli, M., True, N. S., Gerrard, S., LeMaster, C. B., and LeMaster, C. L., *J. Phys. Chem.*, 99:8170-8176, 1995.
98. Wood, R. H. and Mühlbauer, W. C. F., *J. Phys. Chem.*, 95:6670-6675, 1991.
99. Mountain, R. D. and Thirumalai, D., *J. Phys. Chem.*, 93:6975-6979, 1989.
100. Straub, J. E. and Thirumalai, D., *PROTEINS: Struc. Func. and Gen.*, 15:360-373, 1993.
101. Straub, J. E., Rashkin, A. B., and Thirumalai, D., *J. Am. Chem. Soc.*, 116:2049-2063, 1994.
102. Clarage, J. B., Romo, T., Andrews, B. K., Pettitt, B. M., and Phillips, Jr., G. N., *Proc. Natl. Acad. Sci. USA*, 92:3288-3292, 1995.
103. Valleau, J. P. and Torrie, G. M., *A Guide to Monte Carlo for Statistical Mechanics: 2. Byways*, In *Statistical Mechanics, Part A*, Chapter 5, edited by Berne, B. J., Plenum Press, New York, 1977.
104. Beutler, T. C. and Van Gunsteren, W. F., *J. Chem. Phys.*, 100:1492-1497, 1994.
105. Saito, M., *J. Phys. Chem.*, 99:17043-17048, 1995.
106. Jorgensen, W. L., Chandrasekhar, J., and Madura, J. D., *J. Chem. Phys.*, 79:926-935, 1983.
107. Guenot, J. and Kollman, P. A., *J. Comp. chem.*, 14:295-311, 1993.
108. Allen, M. P. and Tildesley, D. J., *Computer Simulation of Liquids*, Oxford University Press, Oxford, 1987.
109. Cieplak, P. and Kollman, P. A., *J. Comp.-Aided Mol. Des.*, 7:291-304, 1993.
110. Schreiber, H. and Steinhauser, O., *Biochemistry*, 31:5856-5860, 1992.
111. Berendsen, H. J. C., Van Gunsteren, W. F., Zwinderman, H. R. J., and Geurtsen, R. G., *Ann. N.Y. Acad. Sci.*, 482:269-286, 1986.
112. Foley, C. K., Pedersen, L. G., Charifson, P. S., Darden, T. A., Wittinghofer, A., Pai, E. F., and Anderson, M. W., *Biochemistry*, 31:4951-4959, 1992.

113. Ewald, P., *Ann. Phys.*, 64:253-287, 1921.
114. Hockney, R. W. and Eastwood, J. W., *Computer Simulation using Particles*, IOP Publishing Ltd., Bristol, England, 1988.
115. Luty, B. A., Davis, M. E., Tironi, I. G., and Van Gunsteren, W. F., *Mol. Simul.*, 14:11-20, 1994.
116. Luty, B. A. and Van Gunsteren, W. F., *J. Phys. Chem.*, 100:2581-2587, 1996.
117. Pearlman, D. A. and Kollman, P. A., *J. Chem. Phys.*, 94:4532-4545, 1991.
118. Mark, A. E., Van Gunsteren, W. F., and Berendsen, H. J. C., *J. Chem. Phys.*, 94:3808-3816, 1991.
119. Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A., *J. Comp. Chem.*, 16:1357-1377, 1995.

PART II

SOLVATION FREE ENERGIES OF PHENYLALANYL PEPTIDES

LITERATURE REVIEW

I. Molecular Properties for Efficient Drug Delivery

Understanding the biophysical mechanisms governing the membrane transport of drugs is essential to discover and develop drugs that may be transported more efficiently to target sites. Because most drug absorption occurs in the intestinal lumen by passive diffusion, rather than by carrier- or receptor-mediated processes,^[1] it is very important to know what kinds of molecular properties and parameters affect the permeability of membranes to the molecules of interest.

Parameters that are routinely examined in terms of drug delivery include partition coefficient, permeability, solubility, molecular weight, molecular charge, pKa of the solute, and pH of the solvent. The 1-octanol/water partition coefficient ($\log []_{1\text{-oct}}/[]_{\text{H}_2\text{O}} = \log P$) is the parameter most frequently used to characterize the transport properties of drugs, with a higher transport rate generally being predicted with increasing partition coefficient.^[2]

Recently, a series of papers reported a poor correlation between partition coefficients and permeability coefficients in studies of the mechanism of membrane transport of phenylalanyl peptides into Caco-2 (human-derived colon carcinoma cell)^[1] monolayer membrane systems.^[3,4,5] In these studies, a zwitterionic (at physiological pH) peptide series, (D-Phe)_nGly, and a neutral series, Ace(D-Phe)_nNH₂ (where n=1, 2, or 3), were employed. As would be expected, terminal charges on zwitterionic peptides reduced membrane permeability. Less expected was the observation that incremental addition of D-Phe also

decreased membrane permeability, even though the partition coefficient increased as n increased, suggesting that partition coefficients may be poor predictors of membrane absorption of peptides in this system.^[3]

It was proposed that a major impediment to passive absorption of peptides is the energy required to break water-peptide hydrogen bonds in order for the solute to enter the cell membrane (desolvation of the solute).^[4] In fact, it was found that methylation of amide nitrogens dramatically improved the permeability of phenylalanyl peptides across Caco-2 cell membranes.^[5] The increased absorption was not attributed to increased partition coefficient because methylation produced only a very small increase in this parameter. The authors suggested that because the N-H group can act as a hydrogen bond donor, adding a methyl group to the amide nitrogen may reduce solute-solvent intermolecular hydrogen bonding, resulting in more favored desolvation.^[5] These results are summarized in **Table 1**.

From these studies, it was suspected that the desolvation/solvation free energy ($\Delta G_{\text{desolv}}/\Delta G_{\text{solv}}$) of a peptide, where ΔG_{desolv} is the negative of ΔG_{solv} , might correlate better with its permeability than does the partition coefficient. However, values of ΔG_{solv} are not typically examined by researchers studying drug delivery across membranes, probably because of the difficulty in experimental determination of these values. Thus, a computational method to calculate values of ΔG_{solv} will provide an additional tool to apply in assessing the transport properties of drugs.

Table 1 Partition Coefficient (Log P) and Permeability Values in the Literature for D-Phenylalanyl Peptides

Peptide	Partition Coefficient	Permeability (cm/s) x 10 ⁶	Ref. #
Ace(D-Phe)NH ₂	0.16	7.3	3
	0.05	7.99	4
Ace(D-Phe) ₂ NH ₂	1.29	2.3	3
	1.19	2.20	4
Ace(D-Phe) ₃ NH ₂	2.8	0.7	3
	2.30	0.66	4
Ace(D-Phe) ₃ NH ₂ +(Me) ₁	2.63	2.78	5
	+(Me) ₂	5.68	5
	+(Me) ₃	13.8	5
	+(Me) ₄	23.8	5

Note: 'Ace' and 'Me' stand for acetyl and methyl groups, respectively.

II. Molecular Modeling Tools for Solvation Free Energy Calculation

There are several solvation free energy simulation programs available for biological molecules, which include a semiempirical method (AMSOL), a simplified microscopic simulation method (POLARIS), and free energy perturbation (FEP) methods (AMBER).

The semiempirical method provides four approaches to the calculation of solvation free energy in the AMSOL program: AM1-SM1 (Solvation Model 1), AM1-SM1a, AM1-SM2, and PM3-SM3.^[6] In these models, the free energy of solvation (ΔG_{solv}) is partitioned as follows:

$$\Delta G_{\text{solv}} = E_{\text{EN}}(\text{aq}) + G_{\text{p}}(\text{aq}) + G_{\text{CDS}}(\text{aq}) - E_{\text{EN}}(\text{g}) + \Delta G_{\text{vib}} + \Delta G_{\text{elec}}$$

where $E_{\text{EN}}(\text{aq})$ and $E_{\text{EN}}(\text{g})$ are the sums of the solute electronic kinetic and electronic-nuclear coulombic energies in aqueous and gas phase respectively; $G_{\text{p}}(\text{aq})$ is the solution polarization free energy; $G_{\text{CDS}}(\text{aq})$ is the sum of the cavitation, dispersion, and (solvent) structural free energies; ΔG_{vib} and ΔG_{elec} are vibrational and electronic free energy changes upon dissolution, which are ignored upon parameterization due to their small quantities.^[6] The partition of solvation effects into the G_{p} and G_{CDS} terms is the critical physical approximation in all SM models and it is assumed that besides G_{p} , the dominant remaining effects are all approximately proportional to the first solvent-accessible surface area of the solute upon hydration.^[6] Parameters used in these models include atomic radii, solvent accessible surface tensions, and constants in the bond order equations.^[6]

POLARIS employs the Protein Dipoles Langevin Dipoles (PDL) model to compute ΔG_{solv} .^[7] This approach approximates the average energy of the protein and water system by

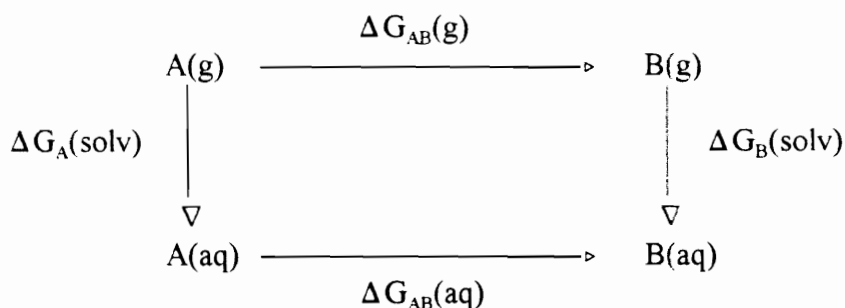
the electrostatic energy evaluated at the average structure.^[7] All of the key electrostatic contributions in molecules (e.g., permanent dipoles, induced dipoles, and surrounding water molecules) are represented, but a simplified set of Langevin dipoles is used to represent the solvent.^[8] The PDL method assumes that the solute is surrounded by a sphere, which contains a three-dimensional cubic grid. Each grid point within a specific van der Waals distance from a solute atom is deleted, and each of the remaining grid points represents a Langevin dipole.^[7] Three regions are defined in the PDL method: (1) the solute, (2) the sphere consisting of the grid surrounding the solute, and (3) the infinite surrounding beyond the finite grid system, which is represented as a continuum with the dielectric constant of water.^[7] The ΔG_{soln} of the solute is calculated as a sum of the following energies: (i) the total electrostatic interaction between the charges within region (1); (ii) the energy of interaction between the Langevin dipoles in region (2) and charges in region (1); and (iii) the electrostatic contribution from the bulk waters in region (3).^[7,8]

Different atomic charge calculation techniques may be utilized to calculate ΔG_{soln} values in POLARIS. These techniques include Mulliken population analysis, electrostatic potential analysis, the charge equilibration method, and the Gasteiger method. The oldest and most widely used is the Mulliken population analysis, which uses a density matrix in a Hartree-Fock wavefunction (see PART I).^[9,10,11] Net atomic charges (populations) are assigned from the diagonal terms, and overlap charges are obtained from the off-diagonal elements, where each charge is split between two atoms based on their orbital exponents.^[9,10,11] The electrostatic potential (ESP) analysis^[12] first calculates shells (molecular

surfaces) of points at a specified distance from the molecule, computes the electrostatic potential at these points based on quantum mechanical calculations, and finally fits the potential at those points to that calculated quantum mechanically.^[13] The charge equilibration (QEq) approach first considers how the energy of an isolated atom changes as a function of charge, which is expressed by the ionization potential, electron affinity, and electronegativity terms.^[14] Secondly, it adds the interatomic electrostatic energy term, which leads to a total electrostatic energy expression.^[14] Derivatives for all atoms in the molecule then are simultaneously solved to obtain atomic charges, assuming that at equilibrium all the atomic chemical potentials are equal.^[14] In the Gasteiger method, only the connectivities of the atoms (i.e., the topology of a molecule) are considered.^[15] First, the electronegativity of an atom is expressed as a function of the total charge in a quadratic equation and then the effect of charge transfer between directly neighboring atoms is represented as new atomic charges as a function of atomic electronegativities, from which the new total charge is calculated.^[15] This iteration continues until the atomic charge differences before and after the iteration are below 0.001.^[15]

Two of the programs mentioned above, AMSOL and POLARIS, compute the absolute free energy of solvation (ΔG_{solv} ; from the gas state to the aqueous state directly) for the molecule of interest. The FEP method also calculates ΔG_{solv} by mutating a complete molecule into dummy atoms so that the molecule eventually vanishes in a box of water molecules.^[16,17] This method, however, is more often used to calculate the relative free energy of solvation ($\Delta \Delta G_{\text{solv}}$) of two similar molecules by applying the following thermodynamic

cycle:



In this cycle the molecules, A and B, are either in the gas (g) or aqueous (aq) phase, ΔG_{AB} is the free energy difference for mutating A to B, and $\Delta G_A(\text{solv})$ and $\Delta G_B(\text{solv})$ are the free energies of solvation of A and B respectively.^[16] Since ΔG_{AB} can be calculated with the FEP method and the free energy is a state function, $\Delta \Delta G_{AB}(\text{solv}) = \Delta G_{AB}(\text{aq}) - \Delta G_{AB}(\text{g}) = \Delta G_B(\text{solv}) - \Delta G_A(\text{solv})$, the relative solvation free energy between A and B can be calculated.^[16]

The FEP method is useful not only in calculating the relative solvation free energy difference between two molecules, but also in estimating solvent effects on conformational equilibria, in determining partition coefficient of a molecule in two different solvents, in estimating the free energy of host-guest binding, and in studying chemical reactions in solution.^[18]

III. Solvation Free Energy in the Literature

Experimentally, the ΔG_{solv} value of a molecule can be obtained by measuring the concentration of the molecule in a dilute aqueous solution ($[]_{\text{aq}}$) and its concentration in the gas phase ($[]_{\text{gas}}$) that is in equilibrium with that solution. Activity coefficient values ($\gamma =$

$[]_{\text{aq}}/[]_{\text{gas}}$ and $\Delta G_{\text{solv}} = -RT \ln \gamma$) of many organic molecules have been reported.^[19] Similarly, ΔG_{solv} values of various hydrocarbons were reported based on their activity coefficient values.^[20] Also, the ΔG_{solv} values for the amino acid side chains were measured by determining their vapor pressure over dilute aqueous solutions.^[21] Similar techniques, where the vapor pressure and the water solubility are utilized, were employed to determine ΔG_{solv} of anisole, 1,2-dimethoxybenzene, and 1,2,3-trimethoxybenzene.^[22] In addition, the ΔG_{solv} for protonated valine and proline was measured by a high-pressure mass spectrometric method,^[23] however, ΔG_{solv} measurements of most of the amino acids or small peptides have not been reported.

On the other hand, several computer simulation studies of absolute ΔG_{solv} values have been published. POLARIS provided very good agreement between the calculated and the experimental ΔG_{solv} values for some small neutral and charged molecules.^[24] AM1-SM1 and DelPhi (where the electrostatic component of ΔG_{solv} is evaluated by solving the Poisson-Boltzman equation by a finite difference method) methods were compared by their performances in reproducing the experimental results for small charged and neutral molecules and it was concluded that the AM1-SM1 method outperforms DelPhi.^[25] Also with AMSOL, ΔG_{solv} for neutral and charged small molecules were calculated using SM1a, SM2, and SM3 methods; good agreement with experimental results was reported.^[6] It was indicated that overall both AM1-SM2 and PM3-SM3 offer improved performance over the previous models, and AM1-SM2 seems to be better than PM3-SM3 for some nitrogen-containing compounds.^[6] Finally, the FEP method has been used to compute the absolute ΔG_{solv} of H₂O,

NH₃, and MeNH₂ with good agreement with the experimental values.^[17]

Alternative computer simulations for solvation free energies of organic/biological molecules have employed the FEP methods, in which the relative solvation free energy ($\Delta\Delta G_{\text{solv}}$) of two molecules is measured. For example, $\Delta\Delta G_{\text{solv}}$ (methane $\rightarrow$ ethane), $\Delta\Delta G_{\text{solv}}$ (ethane $\rightarrow$ propane), and $\Delta\Delta G_{\text{solv}}$ (Ace-Ala-NMe $\rightarrow$ Ace-Val-NMe) were calculated and the first two free energy values were with good agreement with experiment, whereas no experimental value was available for the third simulation.^[26] Also, $\Delta\Delta G_{\text{solv}}$ values for amino acid side chains have been calculated and good agreement with experiment was obtained.^[16] Different atomic charges calculated by the ESP method with different basis sets were utilized to calculate $\Delta\Delta G_{\text{solv}}$ (anisole $\rightarrow$ benzene) and $\Delta\Delta G_{\text{solv}}$ (1,2,3-trimethoxybenzene $\rightarrow$ benzene) and the 6-31G* [which indicates (6 Gaussian functions for the core orbitals)-(3 Gaussian functions for the inner valence)(1 Gaussian function for the outer valence)(* = polarization effects added)] charge model was found to reproduce the experimental values better than STO-3G (Slater-type orbitals in terms of 3 Gaussian functions) and 4-31G basis sets.^[22]

Because the ΔG_{solv} of peptides have not often been examined experimentally and because it is difficult to measure these values, a computer simulation method to calculate ΔG_{solv} values will be very useful. However, in developing computer simulation methods, it is imperative that computed values are shown to agree with experimental values with a test set of compounds. In this project, the AMSOL and POLARIS methods are applied to small compounds for which experimental values of ΔG_{solv} are available as well as to the D-phenylalanyl peptides.

METHODS

Molecular models of the phenylalanyl peptides were built using QUANTA, version 3.3.1 (Molecular Simulations, Inc.). These peptides are denoted as in **Table 2**.

Following model building, conformational searches were performed for each peptide using the random sampling method in QUANTA. A total of 500 structures was generated for each peptide, with energy minimization of the structures being done using the Adopted-Basis Newton-Raphson method and an energy gradient tolerance of 0.001. For each of the 12 peptides, the five lowest energy structures from the conformational search and energy minimization were selected for further analysis.

To prepare structures for calculation of solvation free energy using POLARIS version 3.22 (Molecular Simulations, Inc.), atomic charges were calculated for each structure using MOPAC version 6.0 (Quantum Chemistry Program Exchange). The MNDO Hamiltonian was employed as recommended in the manual, with charges being calculated using both the ESP and Mulliken population analysis methods. POLARIS was then used to calculate solvation free energies for each structure with each of the sets of atomic charges.

For calculation of solvation free energies using AMSOL version 4.5 (Quantum Chemistry Program Exchange), MOPAC calculations were done using the AM1 Hamiltonian to calculate the heat of formation for each of the structures. These heats of formation were required for the AMSOL calculations, which were done using the AM1-SM2 method. This method was chosen because it is one of the two latest methods and the other method (PM3-

Table 2 D-Phenylalanyl Peptides Used for Solvation Free Energy Simulation

(D-Phe) ₁ Peptide	R1*	R2	Designation
Ace-N-Phe-CO-NH	H	H	F0
R1 R2	Me	H	F1
	Me	Me	F2

(D-Phe) ₂ Peptide	R1	R2	R3	Designation
Ace-N-Phe-CO-N-Phe-CO-NH	H	H	H	FF0
R1 R2 R3	H	Me	H	FF1
	Me	Me	H	FF2
	Me	Me	Me	FF3

(D-Phe) ₃ Peptide	R1	R2	R3	R4	Designation
Ace-N-Phe-CO-N-Phe-CO-N-Phe-CO-NH	H	H	H	H	FFF0
R1 R2 R3 R4	H	H	Me	H	FFF1
	H	Me	Me	H	FFF2
	Me	Me	Me	H	FFF3
	Me	Me	Me	Me	FFF4

* 'Me' indicates for a methyl group.

SM3) has been reported to have problems with nitrogen-containing molecules.^[6] The AMSOL calculation time with full optimization is very long: one AMSOL run for a mono-peptide takes approximately one to three weeks on a DEC Alpha Workstation. Because of the length of time required for geometry optimization during the AMSOL calculations, 'ISCF' calculations, where no geometry optimization is done, were performed using the geometries from AM1 optimization of the non-solvated system. With this simplified option, AMSOL runs much faster, requiring only 10-20 seconds to compute ΔG_{solv} for each of our peptides. In addition, the solvation free energies with and without 'ISCF' option had a difference of less than 3% (based on the comparisons for mono-phenylalanyl peptides and the molecules listed in **Table 3**; data not shown). Also, AMSOL performance with and without the 'ISCF' option using AM1-SM1^[25] and AM1-SM2^[27] methods has been reported and it was noted that the resulted solvation free energies were not significantly different between those options.

All the calculations were run on a Silicon Graphics Personal Iris Model 4D/35 except for AMSOL calculations which were done on a DEC Alpha Workstation in the Virginia Tech Computing Center. See APPENDIX C for step-by-step details of the methods.

RESULTS

Test Compounds

To validate methods for calculation of solvation free energy, we computed solvation free energy for some compounds for which experimental values are available (**Table 3**).

In general, ΔG_{solv} values calculated using POLARIS are in poor agreement with experimental values, especially for acetamide and NMAs. Because the ΔG_{solv} values are highly dependent upon atomic charges, several methods were applied to calculation of these charges. However, the calculated values were not improved by applying another Hamiltonian (AM1) in the semiempirical methods (data not shown) or by using the QEq and Gasteiger methods for atomic charge calculation (data not shown). Despite the poor agreement between the values of ΔG_{solv} calculated using POLARIS and the experimental values, the relative differences in the values among the compounds in **Table 3** are reasonably consistent, if not always in magnitude at least in sign.

Calculations of ΔG_{solv} using AMSOL also gave poor agreement with experimental values for small hydrophobic molecules. However, for the polar molecules, acetamide and NMA, which are often considered to be the smallest models for the peptide bond in proteins, agreement was excellent with AMSOL. To see if atomic charges in AMSOL were acceptable, we compared the calculated dipole moments in MOPAC (in gas phase) and in AMSOL (in aqueous phase) with experimental values (**Table 4**). Agreement was very good between MOPAC results and experimental values. Increased dipole moments in AMSOL

Table 3 Test Calculation Results of ΔG_{soln} by AMSOL and POLARIS

Mol. Name	AMSOL	POLARIS (MNDO)		Exp.
		ESP	Mulliken	
methane	1.0	2.5 [2.4]	2.5	2.0
benzene	-0.5 (-0.5)	-2.7 [-0.7]	-1.1	-0.9/-0.9*
toluene	-0.3 (-0.3)	-2.8 [-1.2]	-1.7	-0.9/-0.9*
acetamide	-11.4	-5.8	-4.4	-11.7
NMA(<i>trans</i>)	-9.2	-5.0 [-10.6]	-3.4	-10.2/ $\approx$ -10*
NMA(<i>cis</i>)	-10.2	-5.0	-4.4	

Note: Experimental and calculated [] values are from POLARIS User's Guides, where they had been obtained from Dr. Arieh Warshel's group (USC), and experimental values with asterisks and calculated () values are from Cramer and Truhlar.[6]
 'ESP' and 'Mulliken' stand for Electrostatic Potential and Mulliken population analysis methods to calculate atomic charges, respectively.
 'NMA' stands for N-methyl acetamide.

Table 4 Dipole Moments Comparison in MOPAC and AMSOL

Mol. Name	Dipole Moment (debyes)		
	MOPAC	AMSOL	Exp. (in gas)
methane	0.0	0.0	0.0
benzene	0.0	0.0	0.0
toluene	0.27	0.33	0.38
acetamide	3.76	4.90	3.76
NMA(<i>trans</i>)	3.51	4.50	3.73*
NMA(<i>cis</i>)	3.83	5.22	

Note: Experimental results are from *CRC Handbook of Chemistry and Physics*, 76th ed., CRC Press, 1995-1996 (* from 66th ed.)

indicates atomic charges used in AMSOL calculations were appropriate because charges are supposed to be further polarized upon solvation.^[6]

D-Phenylalanyl Peptides

Results from the calculations of solvation free energies of the D-phenylalanyl peptides using POLARIS are summarized in **Table 5**. A distinct decrease in ΔG_{solv} is observed, using either ESP or Mulliken charges, when comparing F0, FF0, and FFF0, which are the D-phenylalanyl peptides of one, two, and three residues, respectively, without additional methyl groups. Similar trends are observed for all the series of peptides with equivalent numbers of added methyl groups (e.g., F1, FF1, and FFF1). This decrease in ΔG_{solv} indicates that solvation becomes more favored as the peptide length increases. Thus, desolvation of the peptides becomes less favored with increasing length of peptide, which is consistent with the decreased permeability of the peptides as the chain length increases. As methyl groups are added to peptides of a given length, the ΔG_{solv} generally shows a small increase when ESP charges are used, except between FFF0 and FFF1. Thus, desolvation becomes more favorable as methyl groups are added, which is consistent with the permeability trend that was reported (**Table 1**). Free energy differences upon methylation, when Mulliken charges were used to calculate ΔG_{solv} , are not as obvious for all peptide lengths. However, **Table 3** suggests that the absolute ΔG_{solv} values calculated by POLARIS may not be reliable, in terms of reproducing experimental ΔG_{solv} values by simulation, even though ΔG_{solv} values for our peptides by POLARIS generally indicate good agreement with experiment.

Table 5 Solvation Free Energies Calculated by POLARIS with MNDO

Pep. Name	ΔG_{solv} (kcal/mol)			
	ESP	SD	Mulliken	SD
F0	-11.3	0.4	-8.6	0.3
F1	-10.5	0.7	-8.4	0.7
F2	-9.2	0.5	-7.8	0.4
FF0	-16.8	0.7	-13.2	0.7
FF1	-15.3	0.3	-12.2	0.4
FF2	-14.8	0.8	-12.0	0.5
FF3	-14.0	0.6	-11.9	0.5
FFF0	-19.8	1.1	-15.6	0.8
FFF1	-20.1	1.9	-15.7	1.4
FFF2	-19.3	0.7	-15.2	1.1
FFF3	-17.6	1.0	-14.3	0.9
FFF4	-17.4	1.5	-14.1	1.5

Note: The above values are the averages among five conformers.
For the individual calculation results, see Table 1 in APPENDIX D.
'ESP' and 'Mulliken' stand for Electrostatic Potential and Mulliken population analysis methods to calculate atomic charges, respectively.
'SD' is for standard deviation.

We also applied the AMSOL program to calculation of ΔG_{soln} of these peptides (**Table 6**). In general, the values for ΔG_{soln} obtained using AMSOL were similar to those obtained with POLARIS in that as the chain length increased, ΔG_{soln} value decreased. However, the effect of adding methyl groups was not clear, except between F1 and F2 and between FF2 and FF3, where a decrease in ΔG_{soln} of approximately 2 kcal/mol was seen.

Because the effect of methylation on ΔG_{soln} was not apparent, we performed some additional calculations to see if there were any effects of position of methylation on the solvation free energies of these peptides. The first test molecule, denoted test1, does not have any methyl groups on the backbone including the methyl in the N-terminal amide (HCO-NH-Phe-CO-NH₂). Test2, test3, and test4 are derived from test1 and have a methyl group added at the N-terminal formamide nitrogen (test2), the C-terminal amide nitrogen (test3), and the N-terminal formamide carbon and the C-terminal amide nitrogen (test4). Note that test4 also corresponds to F0 with an additional methyl group at the C-terminal amide nitrogen.

The AMSOL calculation results for these molecules are shown in **Table 7** and a comparison among the results in **Table 6** and those in **Table 7** is illustrated in **Figure 1**.

Data in **Figure 1** reveal that methylation at the C-terminal amide nitrogen consistently increased the solvation free energies by approximately 1.7 kcal/mol. Also, methylation at the N-terminal (formamide carbon), which makes the N-terminal an acetyl group, added up to 2.5 kcal/mol to the solvation free energy. On the other hand, methylation at the N-terminal formamide nitrogen had a lesser effect (+1.0 kcal/mol for test1 $\rightarrow$ test2) or no effect (-0.1 kcal/mol for F0 $\rightarrow$ F1 and 0.0 kcal/mol for test4 $\rightarrow$ F2) on the solvation free energies.

Table 6 Solvation Free Energies Calculated by AMSOL with 1SCF Option

Pep. Name	ΔG_{soln} (kcal/mol)	SD
F0	-12.9	0.0
F1	-13.0	0.0
F2	-11.3	0.0
FF0	-15.5	0.0
FF1	-15.6	0.1
FF2	-15.7	1.0
FF3	-13.7	0.3
FFF0	-17.2	1.5
FFF1	-16.7	2.6
FFF2	-17.1	1.2
FFF3	-16.0	2.1
FFF4	-16.7	1.4

Note: The above values are the averages among five conformers.
For the individual calculation results, see Table 2 in APPENDIX D.
'SD' is for standard deviation.

Table 7 AMSOL Test Calculation Results with 1SCF Option

Pep. Name	ΔG_{soln} (kcal/mol)	SD
test1 (HCO-NH-Phe-CO-NH ₂)	-15.4	0.0
test2 (HCO-NMe-Phe-CO-NH ₂)	-14.4	0.0
test3 (HCO-NH-Phe-CO-NHMe)	-13.7	0.0
test4 (Ace-NH-Phe-CO-NHMe)	-11.3	0.0

Note: The above values are the averages among five conformers.
For the individual calculation results, see Table 3 in APPENDIX D.
'SD' is for standard deviation.

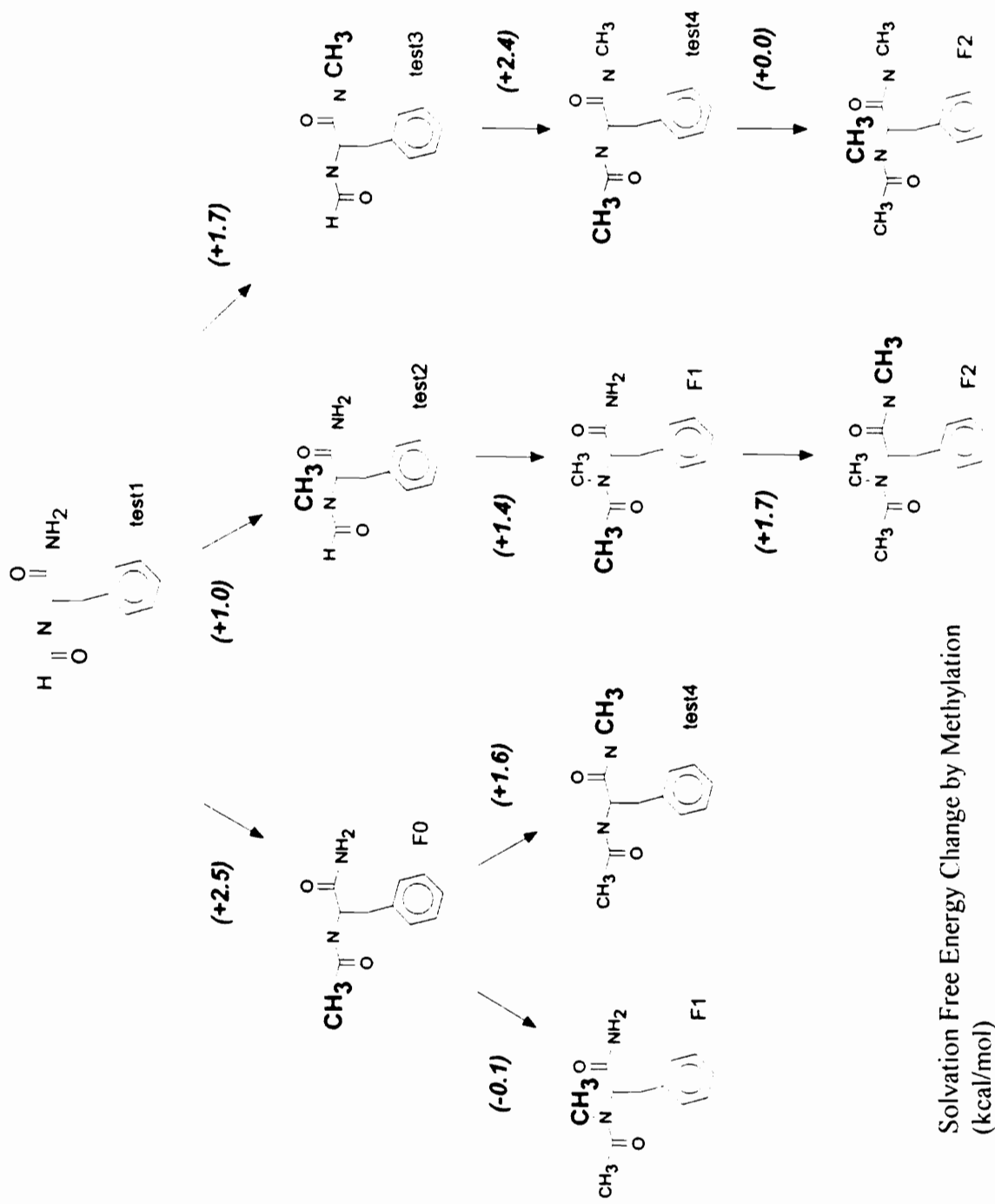


Figure 1 Solvation Free Energy Change by Methylation (kcal/mol)

Several properties of the peptides under study were compared in an attempt to understand some of the differences in solvation free energy. For example, from comparison of the structures using interactive molecular graphics, it was found that methylation at the N-terminal formamide carbons and at the C-terminal amide nitrogens did not affect the structures of the peptides but that alkylation at the N-terminal acetamide nitrogen did change the peptide structures in that the positions of the phenylalanyl side chains were moved, presumably to avoid the added methyl groups on the backbone. The backbone structures were slightly changed as well. However, structural changes that correlated with solvation free energies were not apparent.

Additional features examined from AMSOL runs were atomic charges of some of the peptides (**Table 8**), and the dipole moment and surface area (**Table 9**). In addition, the individual terms, the sum of which comprise the AMSOL ΔG_{solv} values, were tabulated and compared (**Table 10**).

Table 8 indicates that methylation at the N-terminal formamide carbon (test1 – F0, +2.5 kcal/mol) ΔG_{solv} only slightly shifted the amide nitrogen and carbon charges to that of the methyl carbon. For the others, methylation at a formamide nitrogen (test1 – test2, +1.0 kcal/mol) and at a C-terminal amide nitrogen (test1 – test3, +1.7 kcal/mol) did not show much changes in overall atomic charges, either.

The effect of methyl addition at different positions is more obvious in dipole moment values (**Table 9**) and in polarization (G-P) and cavity-dispersion-solvent structure free energies (G-CDS) (see **Table 10**). Methylation at N-terminal formamide carbons (test1 – F0

Table 8 AMSOL Charge Comparison among Mono-D-Phe Peptides

Atoms	test11	test21	test31	F01
<i>N-terminal</i>				
N	-0.4175	-0.4116	-0.4166	-0.3887
H	0.2655		0.2653	0.2536
C	0.2756	0.2869	0.2747	0.3190
O	-0.4491	-0.4461	-0.4505	-0.4432
H	0.2018	0.1993	0.2015	
<i>Cα-Side Chain</i>				
C α	0.0094	0.0223	0.0077	0.0110
H	0.1510	0.1496	0.1482	0.1515
C β	-0.1224	-0.1171	-0.1205	-0.1225
H	0.1110	0.1135	0.1120	0.1108
H	0.1123	0.1149	0.1120	0.1141
C γ	-0.0898	-0.0988	-0.0894	-0.0906
C δ	-0.1372	-0.1352	-0.1380	-0.1372
H	0.1415	0.1463	0.1407	0.1413
C δ	-0.1385	-0.1392	-0.1398	-0.1387
H	0.1437	0.1520	0.1443	0.1430
C ϵ	-0.1395	-0.1425	-0.1395	-0.1390
H	0.1493	0.1508	0.1492	0.1495
C ϵ	-0.1383	-0.1471	-0.1378	-0.1380
H	0.1511	0.1469	0.1514	0.1513
C ζ	-0.1399	-0.1396	-0.1402	-0.1391
H	0.1505	0.1504	0.1508	0.1509
<i>C-terminal</i>				
C	0.2839	0.2805	0.2952	0.2840
O	-0.4594	-0.4397	-0.4377	-0.4585
N	-0.4108	-0.4177	-0.4099	-0.4099
H	0.2375	0.2363	0.2650	0.2582
H	0.2584	0.2520		0.2380
<i>Methyl (on N-terminal)</i>				
		(at N)		(at C)
C		-0.1176		-0.2334
H		0.1129		0.1201
H		0.1274		0.1211
H		0.1101		0.1212
<i>Methyl (on C-terminal)</i>				
			(at N)	
C			-0.1359	
H			0.1068	
H			0.1068	
H			0.1244	

Note: test11 is one of the conformers of test1. The same is for test21, test31, and F01. (See **Table 7** for the test molecule information.)

Table 9 AMSOL Results - Dipole Moment and Surface Area

Pep. Name	Dipole Moment (debyes)	Surface Area (Å ²)
test1	4.1	382.0
test2	3.5	399.8
test3	3.6	423.1
test4	4.0	394.1
F0	4.5	405.3
F1	4.9	412.0
F2	4.4	429.7
FF0	7.2	558.2
FF1	5.9	589.1
FF2	6.2	592.3
FF3	5.3	617.0
FFF0	6.1	701.5
FFF1	5.6	723.9
FFF2	5.0	726.5
FFF3	4.1	719.9
FFF4	4.2	739.5

Note: The above values are the averages among five conformers.
For the individual calculation results, see Table 4 in APPENDIX D.

Table 10 AMSOL Results - Components of Free Energy of Solvation

Pep. Name	Energies (kcal/mol)					ΔG_{solv}
	E-EN (g)	G-P (aq)	G-ENP(aq)	G-CDS(aq)	G-S (aq)	
test1	-60.8	-11.9	-70.5	-5.6	-76.1	-15.4
test2	-53.2	-13.0	-63.5	-4.1	-67.6	-14.4
test3	-57.3	-13.0	-67.9	-3.1	-71.0	-13.7
test4	-63.1	-10.9	-72.3	-2.2	-74.5	-11.3
F0	-66.6	-9.8	-74.8	-4.7	-79.5	-12.9
F1	-58.5	-11.7	-68.0	-3.5	-71.5	-13.0
F2	-55.0	-12.8	-65.4	-0.9	-66.3	-11.3
FF0	-80.6	-13.0	-91.4	-4.7	-96.1	-15.5
FF1	-69.2	-14.4	-81.5	-3.4	-84.8	-15.6
FF2	-61.5	-16.0	-74.8	-2.5	-77.2	-15.7
FF3	-44.7	-17.5	-57.8	-0.6	-58.4	-13.7
FFF0	-94.6	-15.5	-108.1	-3.7	-111.6	-17.2
FFF1	-82.2	-16.7	-96.5	-2.4	-99.0	-16.7
FFF2	-71.5	-17.9	-86.8	-1.9	-88.7	-17.1
FFF3	-65.2	-17.4	-80.1	-1.1	-81.2	-16.0
FFF4	-61.1	-19.9	-77.9	0.2	-77.8	-16.7

Note: The above values are the averages among five conformers.
 For the individual calculation results, see Table 5 in APPENDIX D.

The free energy of solvation is broken down as follows:

E-EN(g) = gas-phase electronic-nuclear energy (Heat of Formation)

G-P(aq) = polarization free energy of solvation

G-ENP(aq) = electronic-nuclear-polarization energy of system
 = E-EN(aq) + G-P(aq)

G-CDS(aq) = cavity-dispersion-solvent structure free energy

G-S (aq) = free energy of system = G-ENP(aq) + G-CDS(aq)

ΔG_{solv} = free energy of solvation = G-S(aq) - E-EN(g)

and test3 – test4, +2.5 kcal/mol) increased the dipole moments by 0.4 debyes and G-P term also by 2.1 kcal/mol, indicating unfavorable polarization free energies. On the other hand, methyl addition at a formamide nitrogen (test1 – test2, +1.0 kcal/mol) and at a C-terminal amide nitrogen (test1 – test3, F1 – F2, and F0 – test4, +1.7 kcal/mol) decreased the dipole moments by 0.5-0.6 debyes and G-P term by 1.1 kcal/mol, indicating favorable polarization effects upon methylation. In addition, alkylation at the N-terminal acetamide nitrogens (F0 – F1 and test4 – F2, ≈ 0.0 kcal/mol) increased the dipole moments by 0.4 debyes but decreased G-P term by 1.9 kcal/mol, which suggests that there are no clear correlation between the values of the dipole moments and G-P term. However, the G-CDS term was consistently increased for all cases but in different degrees: methylation at a formamide carbon increased the G-CDS term the least (+0.9 kcal/mol), at acetamide and formamide nitrogens that term was increased by 1.3 and 1.5 kcal/mol, respectively, and finally at C-terminal amide nitrogens it was elevated by 2.5 kcal/mol. Among di- and tri-phenylalanyl peptides, the above-mentioned positional methylation effects were less consistent probably due to their conformational variabilities. No clear correlations between positional methylation effects and surface areas were found (**Table 9**).

Therefore, it seems that solvation free energy differences among our phenylalanyl peptides with added methyl groups at various positions on their backbones were mainly influenced by subtle free energy balances between G-P and G-CDS terms.

Small Amides and Amines

Recently, disagreement between calculations by the FEP method (see PART I) and experiment for solvation free energy of small amides and amines was reported by two separate groups, Mongantini et al.^[17] and Ding et al.^[28] Both groups found that the computed free energy of solvation monotonically increases as a function of methyl addition for both small amides (acetamide (ACT), N-methylacetamide (NMA), and N,N-dimethylacetamide (DMA)) and amines (ammonia, methylamine, dimethylamine, and trimethylamine), whereas experimental results show "anomalous orders" in that the orders of the solvation free energies of the amides and amines are $\text{NMA} < \text{ACT} < \text{DMA}$ and $\text{NH}_2\text{CH}_3 < \text{NH}_3 \approx \text{NH}(\text{CH}_3)_2 < \text{N}(\text{CH}_3)_3$, respectively.^[17,28] Their results along with our AMSOL calculation results are summarized in **Table 11** through **13**. Additional AMSOL results for amines are shown in **Table 14**.

For the amide series in **Table 11**, our AMSOL values show that methyl addition monotonically increases the free energy of solvation, which is in good agreement with the data in the literature. However, all the calculated ΔG_{soln} values for acetamide were not in agreement with the experimental value (-9.7 kcal/mol), which was originally obtained by Wolfenden.^[21,29] Because the experimental value of -11.66 kcal/mol was also found for acetamide in the literature, which is comparable with the calculated values, the experimental solvation free energy for acetamide might need to be re-evaluated. On the other hand, the solvation free energy values for the amine series by AMSOL shows an excellent agreement with experiment in that it could reproduce the "anomalous order" for the solvation free energies of the amines. Values from FEP runs failed to show this unusual N-methylation

Table 11 Comparison of Solvation Free Energies (ΔG_{solv}) of Amides and Amines in the Literature

Mol. Name	Solvation Free Energy (kcal/mol)			
	Morgantini ^[17]	Ding ^[28]	AMSOL	Exp.
ACT		-12.63	-11.4	-9.7(-11.7*)
NMA(<i>trans</i>) $\Delta G_{\text{solv}}(\text{ACT})+2.09^{\#}$		-9.60	-9.2	-10.1
NMA(<i>cis</i>)		-10.01	-10.2	-10.1
DMA $\Delta G_{\text{solv}}(\text{NMA})+2.41$		-8.67	-8.3	-8.6
NH ₃	-4.16	-4.0 ^{&}	-4.3	-4.3
NH ₂ (CH ₃)	-3.57	-3.7	-6.1	-4.6
NH(CH ₃) ₂	-1.95	-1.2	-4.2	-4.3
N(CH ₃) ₃	0.39	-0.6	-2.5	-3.2

Note: Experimental values are from [28], except * from POLARIS 3.0 User's Guide.
Morgantini[17] values were obtained by FEP method in AMBER 4.0 with RESP charges from 6-31G* basis set level.
Ding[28] values were obtained by FEP method using AMBER force field in IMPACT program with ESP charges from 6-31G* level.
[#]NMA(*trans*) and DMA values from Morgantini[17] were obtained as relative free energies. ACT, NMA, and DMA stand for acetamide, N-methylacetamide, and N,N-dimethylacetamide, respectively.
[&]Ding[28] values for amines were obtained with added polarization effects in their calculations.

effect in amines.

Table 12 indicates that with AMSOL charge separation was much greater than that with the others for all methylamines. AMSOL showed that the methyl carbon of $\text{NH}_2(\text{CH}_3)$ had approximately 30% of the total negative charge and that fraction of negative charges on methyl carbons kept increasing up to 53% as more methyl groups were added to amines. On the other hand, with the other methods the negative charges originally from nitrogens were much weakly separated within $\text{NH}_2(\text{CH}_3)$ and $\text{NH}(\text{CH}_3)_2$. However, all the calculated dipole moments were in good agreement to each other and also comparable with the experimental values (see **Table 13**).

In **Table 14** the incremental negative charge separation upon methylation of amines by AMSOL was also indicated in the G-P (aq) values, which consistently decreased as more methyl groups are added. On the other hand, the G-CDS (aq) values increased with methylation probably due to increased cavity sizes and hence unfavorable solvent structure rearrangement energies. Note that the G-P-CDS (aq) term, which combines G-P (aq) and G-CDS (aq) terms, clearly parallels ΔG_{soln} values with the "anomalous order" of the amine series. This again implies that the polarization free energy and the cavity-dispersion-solvent structure free energy of solvation together could contribute significantly in qualitative agreement with experiment.

Table 12 Atomic Charge Comparison in the Literature

Molecule	Atom	Atomic Charge		
		Morgantini ^[17]	Ding ^[28]	AMSOL
NH ₃	N	-1.065	-1.103	-0.410
	H	0.355	0.368	0.137
NH ₂ (CH ₃)	N	-1.010	-1.002	-0.395
	H	0.375	0.373	0.153
	C	0.314	0.307	-0.178
	H	-0.018	-0.017	0.089*
NH(CH ₃) ₂	N	-0.821	-0.691	-0.378
	H	0.381	0.373	0.168
	C	0.229	-0.011	-0.159
	H	-0.003	0.057	0.088*
N(CH ₃) ₃	N	-0.408	-0.331	-0.361
	C	-0.035	-0.106	-0.138
	H	0.057	0.072	0.086*

Note: * indicates that the value is the average of atomic charges of three hydrogens/methyl group. Morgantini[17] and Ding[28] values were obtained from 6-31G* RESP and 6-31G* level ESP charge calculations, respectively. AMSOL charges were obtained by Mulliken methods.

Table 13 Comparison of Dipole Moments for Amines with Literature Values

Mol. Name	Dipole Moment (debyes)			
	Morgantini ^[17]	Ding ^[28]	AMSOL	Exp.(gas)
NH ₃	1.96	1.89	1.86	1.47
NH ₂ (CH ₃)	1.78	1.63	1.58	1.31
NH(CH ₃) ₂	1.45	1.22	1.34	1.01
N(CH ₃) ₃	0.98	0.84	1.14	0.61

Note: Morgantini[17] and Ding[28] values were obtained from 6-31G* RESP and 6-31G* ESP charges, respectively. AMSOL moments were obtained from Mulliken charges.

Table 14 AMSOL Results for Amines - Components of Free Energy of Solvation

Mol. Name	Energies (kcal/mol)					ΔG_{solv}
	E-EN (g)	G-ENP(aq)	G-P (aq)	G-CDS(aq)	G-P-CDS (aq)	
NH ₃	-7.3	-7.6	-0.3	-4.0	-4.3	-4.3
NH ₂ (CH ₃)	-7.4	-8.9	-1.8	-4.6	-6.4	-6.1
NH(CH ₃) ₂	-5.6	-8.1	-3.1	-1.7	-4.9	-4.2
N(CH ₃) ₃	-1.7	-4.9	-4.1	0.7	-3.4	-2.5

Note: The free energy of solvation is broken down as follows:

E-EN(g) = gas-phase electronic-nuclear energy (Heat of Formation)

G-ENP(aq) = electronic-nuclear-polarization energy of system

= E-EN(aq) + G-P(aq)

G-P(aq) = polarization free energy of solvation

G-CDS(aq) = cavity-dispersion-solvent structure free energy

G-P-CDS(aq) = G-P(aq) + G-CDS(aq)

ΔG_{solv} = free energy of solvation = G-ENP(aq) + G-CDS(aq) - E-EN(g)

DISCUSSION

It was speculated that the poor agreement between POLARIS and experimental values (**Table 2**) might be caused by terms that are not taken into account in the calculations, such as polarizability and entropy effects. POLARIS mainly considers permanent dipole and induced dipole interactions as key electrostatic contributions,^[8] but solute (region 1) polarizability is ignored.^[7] Polarization effects can be very large for many organic and bio-organic molecules.^[6] In fact, it has been emphasized that a model which considers only point charge-point dipole interactions cannot describe experimental results adequately without consideration of higher order moments and the polarizability effects.^[30] In addition, POLARIS overlooks other effects like cavitation and loss of the translational and rotational degrees of freedom of the solute, which reduce entropy and depends only on the solvent-accessible area and not on the local polar/apolar character of the surface.^[31]

On the other hand, in AMSOL those entropy effects are included in the G-CDS term, which is specifically associated with the first-hydration-shell (first solvent-accessible surface area) and provides effects such as cavitation, dispersion, and solvent-structural rearrangement.^[6] Since the concept of solvent-accessible surface area was first introduced to describe the relationship of proteins to solvent quantitatively,^[32] this term has been associated with energetics of various phenomena such as hydrocarbon solubility,^[33] hydrophobic bonding,^[34] solvation energy in protein folding and binding,^[35] and receptor-ligand binding free energies,^[27,36] and good correlation of solvent-accessible surface area with

the parameters mentioned above has been obtained. AMSOL incorporates this parameter in the G-CDS term by assigning unique surface tensions to the solvent-accessible surface area of various functional groups within the solute. On the other hand, the polarization effects are included in the G-P term.^[27] It has been shown that polarization increases the free energies of solvation by 9-29% and that electronic relaxation is much more important than geometric relaxation.^[27] This is why our AMSOL free energy calculation with 'ISCF' option, where geometry optimization is not performed, is very reasonable.^[27] Therefore, inclusion of the polarizability term could also make a significant difference in its performance compared to POLARIS. In fact, **Table 14** shows that the G-P-CDS (G-P + G-CDS) term values closely correlated with the "anomalous order" of solvation free energies of methyl amines. Therefore, the various N-methylation effects shown with the D-phenylalanyl peptides (see **Figure 1**) could be quite reliable.

From our simulation study, it was shown that as the peptide length increases, solvation free energy becomes more favorable and hence permeability decreases as experimental results imply. On the other hand, experimental partition coefficient (log P) term suggests that as the peptide length increases, D-phenylalanyl peptide becomes more hydrophobic (see **Table 1**). These contradictory results could probably originate from the fact that peptide bonds has both proton donors and acceptors to form hydrogen bonds with hydrogen-bonding solvents such as water and 1-octanol. Therefore, the log P value may not be a good hydrophobicity parameter for molecules like peptides which could be attracted by 1-octanol not only because of their intrinsic hydrophobicity but also of their hydrogen-bonding capabilities. However,

increased permeability (hydrophobicity) upon methylation for tri-phenylalanyl peptides, as shown in **Table 1**, was not clearly correlated with either solvation free energy values from our simulation study or experimental log P values.

After the series of Caco-2 cell permeability experiments with D-phenylalanyl peptides, the same research group reported evidence for a polarized efflux system (probably P-glycoprotein) for peptides in the apical membrane of Caco-2 cells.^[37] In their experiment it was shown that the transport of FFF2 across Caco-2 cells in the apical to basolateral (AP → BL) direction was a function of peptide concentration and that the permeability of FFF2 across the cell monolayers in the BL → AP direction was approximately five times greater than that in the opposite direction whereas that of F0 was not polarized.^[37] In their previous papers, however, they had assumed no D-peptide active transport system and no concentration-dependent kinetics; therefore only passive transport was considered.^[3] If a concentration dependent transport is the case for these peptides, their permeability measurements for the series of D-phenylalanyl peptides could have created some inconsistencies in their results since the peptide concentrations used were variable: trace concentration^[3] and 200 to 300 μM ^[4] for F0, FF0, and FFF0, and 0.4 to 10 μM for FFF1 through FFF4.^[5] Therefore, considering the possible existence of the apically polarized peptide transport system on Caco-2 cells, it can be suggested that permeability differences among series of D-phenylalanyl peptides with various lengths and different degrees of alkylation could have been influenced not only by the intrinsic hydrophobicities of the peptides but also membrane properties of the monolayer cells used.

In conclusion, our simulation results successfully demonstrated that calculated solvation free energies of D-phenylalanyl peptides could correlate well with their lengths and permeability values in the literature, but failed to predict the effects of backbone-nitrogen methylation on peptides' permeability by computing their free energies of solvation. In fact, amines are known to show irregular solvation effects upon successive substitution of hydrogens on the nitrogen atom by methyl groups. However, our AMSOL results for small amines were very encouraging because they could clearly indicate the "anomalous order" of their solvation free energies, which had not been successfully predicted by the FEP methods. Therefore, further investigation is necessary to determine if any useful correlation exists between solvation free energies and permeability values of not only peptides but also other molecules of pharmaceutical interest because possible conditional correlation between those parameters was shown when peptide lengths vary and the potential of AMSOL was strongly indicated by our simulation study.

REFERENCES

1. Hilgers A. R., Conradi, R. A., and Burton, P. S., *Pharm. Res.*, 7:902-910, 1990.
2. Stehle, R. G. and Higuchi, W. I., *J. Pharm. Sci.*, 61:1922-1930, 1972.
3. Ho, N. F. H., Day, J. S., Barsuhn, C. L., Burton, P. S., and Raub, T. J., *J. Controlled Release*, 11:3-24, 1990.
4. Conradi, R. A., Hilgers, A. R., Ho, N. F. H., and Burton, P. S., *Pharm. Res.*, 8:1453-1460, 1991.
5. Conradi, R. A., Hilgers, A. R., Ho, N. F. H., and Burgon, P. S., *Pharm. Res.*, 9: 435-439, 1992.
6. Cramer, C. J. and Truhlar, D. G., *J. Comp.-Aided Mol. Design*, 6:629-666, 1992.
7. Lee, F. S., Chu, Z. T., and Warshel, A., *J. Comp. Chem.*, 14:161-185, 1993.
8. Warshel, A. and Creighton, S., "*Computer Simulation of Biomolecular Systems*," edited by van Gunsteren, W. F. and Weiner, P. K., ESCOM, Leiden, 1989.
9. Mulliken, R. S., *J. Chem. Phys.*, 23:1833-1840, 1955.
10. Wiberg, K. B. and Rablen, P. R., *J. Comp. Chem.*, 14:1504-1518, 1993.
11. Mulliken, R. S., *J. Chem. Phys.*, 36:3428-3439, 1962.
12. Momany, F. A., *J. Phys. Chem.*, 82:592-601, 1978.
13. Singh, U. C. and Kollman, P. A., *J. Comp. Chem.*, 5:129-145, 1984.
14. Rappé, A. K. and Goddard III, W. A., *J. Phys. Chem.*, 95:3358-3363, 1991.
15. Gasteiger, J. and Marsili, M., *Tetrahedron*, 36:3219-3228, 1980.
16. Bash, P. A., Singh, U. C., Langridge, R., and Kollman, P. A., *Science*, 236:564-568, 1987.
17. Morgantini, P.-Y. and Kollman, P. A., *J. Am. Chem. Soc.*, 117:6057-6063, 1995.

18. Jorgensen, W. L., *Acc. Chem. Res.*, 22:184-189, 1989.
19. Hine, J. and Mookerjee, P. K., *J. Org. Chem.*, 10:292-298, 1975.
20. Ben-Naim, A. and Marcus, Y., *J. Chem. Phys.*, 81:2016-2027, 1984.
21. Wolfenden, R., Andersson, L., Cullis, P. M., and Southgate, C. C. B., *Biochemistry*, 20:849-855, 1981.
22. Kuyper, L. F., Hunter, R. N., Ashton, D., Merz, Jr., K. M., and Kollman, P. A., *J. Phys. Chem.*, 95:6661-6666, 1991.
23. Meot-Ner, M. and Field, F. H., *J. Am. Chem. Soc.*, 96:3168-3171, 1974.
24. Lee, F. S., Chu, Z. T., and Warshel, A., *J. Comp. Chem.*, 14:161-185, 1993.
25. Alkorta, I., Villar, H. O., and Perez, J. J., *J. Comp. Chem.*, 14:620-626, 1993.
26. Sun, Y., Spellmeyer, D., Pearlman, D. A., and Kollman, P., *J. Am. Chem. Soc.*, 114:6798-6801, 1992.
27. Cramer, C. J. and Truhlar, D. G., *J. Am. Chem. Soc.*, 115:8810-8817, 1993.
28. Ding, Y., Bernardo, D. N., Krogh-Jespersen, K., and Levy, R. M., *J. Phys. Chem.*, 99:11575-11583, 1995.
29. Wolfenden, R., *Biochemistry*, 17:201-204, 1978.
30. Fawcett, W. R. and Blum, L., *J. Electroanal. Chem.*, 355:253-263, 1993.
31. Vajda, S., Weng, Z., Rosenfeld, R., and DeLisi, C., *Biochemistry*, 33:13977-13988, 1994.
32. Lee, B. and Richards, F. M., *J. Mol. Biol.*, 55:379-400, 1971.
33. Hermann, R. B., *J. Phys. Chem.*, 76:2754-2759, 1972.
34. Chothia, C., *Nature*, 248:338-339, 1974.
35. Eisenberg, D. and McLachlan, A. D., *Nature*, 319:199-203, 1986.

36. Krystek, S., Stouch, T., and Novotny, J., *J. Mol. Biol.*, 234:661-679, 1993.
37. Burton, P. S., Conradi, R. A., Hilgers, A. R., and Ho, N. F. H., *Biochem. Biophys. Res. Commun.*, 190:760-766, 1993.

PERSPECTIVES

With the continued development of the fields of computational chemistry and hardware technology, research activity in the various kinds of atomistic and numerical molecular modeling has been growing and having a positive impact on the rest of chemistry as well as other sciences, such as biology, medicine, and material science. The upward trend in the use of molecular simulation is evident in the scientific literature, which shows a steady rise in the percentage of papers that incorporate modeling. Thus, molecular modeling is proving influential and useful.

However, there are a number of issues which still need to be resolved regarding free energy simulations as described in this paper. So, before the potential of the modeling methods can be fully realized, we must derive more effective, consistent, and reliable methodologies for performing free energy calculations. In the meanwhile, currently available methodologies need to be fully examined in order to benefit from an alternative method at hand; e.g., testing of the TI method for the modeling of *cis/trans* equilibria of prolyl peptides.

Free energy is probably the most important general concept in chemistry because free energies of molecular systems describe their tendencies to associate and react. Therefore, being able to predict this quantity using computer simulation would be an enormously important advance. Testing of available molecular modeling tools is thus very critical to help further progress in computational chemistry so that molecular simulation could be a more accurate and practical method to predict molecular phenomena in routine basis.

APPENDIX A

PART I - Detailed Methods

All the work was done on a Silicon Graphics Personal Iris Model 4D/35.

1. Building a peptide in QUANTA (version 3.3.1)

- a. Open 'Applications' window and select 'Sequence Builder'.
- b. Open the amino acid residue library by choosing 'AMINOH.RTF'.
- c. Click 'TYR' and 'PRO' to make Ace-Tyr-Pro-NMe.
- d. Open 'Edit' to choose 'Apply Patches to Residues'.
- e. Click 'NACT' for an acetyl blocking group, then click 'TYR' in the sequence.
- f. Click 'CMAM' for an N-methyl amide blocking group, then click 'PRO' in the sequence.
- g. Open 'SequenceBuilder' window, and choose 'Returns to Molecular Modeling'.
- h. Save the structure.

2. Setting the torsion to cis in QUANTA

- a. Open 'Edit' and select 'Molecular Editor'.
- b. Click 'Edit Bonds...'
- c. Choose 'Set Torsion Angle...' and click the three bonds which define the torsion of interest.
- d. Enter 0 for 'New dihedral angle'.
- e. Save the change and exit from 'Molecular Editor'.
- f. Open 'CHARMm' and choose 'Constraints Options'.
- g. Select 'Dihedral Constraints' and 'Create Dihedral Constraints'.
- h. Choose 'Enter Dihedral' and click four atoms which define the dihedral.
- i. Enter 0 for 'Optimum Dihedral Angle' and click 'Finish'.
- j. Choose 'Dihedral Constraints' again and select 'Dihedral Constraints ON'.
Save the constraint.

3. Conformational Search in QUANTA

- a. Open 'CHARMm' and 'Minimization Options'.
- b. Select 'Adopted-Basis Newton Raphson'.
- c. Enter 3000 and 0.001 for 'Number of Minimization Steps' and 'Energy Gradient Tolerance', respectively.
- d. Open 'Applications' window and select 'Conformational Search'.
- e. Click 'Torsions...' and 'Define All Torsions', then 'Select Torsions...'.
- f. Select 'All Torsions' then 'Finish'. Close 'Torsions'.

- g. Click 'Setup Search...' then 'Random Sampling'.
- h. Enter 1000 and 360 for 'Number of Samples' and 'Torsion Angle Window', respectively.
- i. Select 'Minimize Each Structure', 'Cumulative Random Modifications', and 'Display Each Structure'.
- j. Click 'Do Search'. Define .csr file name.
- k. After the search, select 'ANALYSIS'.
- l. Open the search (.csr) file.
- m. Select 'Sort By Property...' in the 'ANALYSIS' window.
- n. Click 'Use Sorted Order' then 'First Structure'.
- o. Choose 'File Options/Filters...' and write a .msf file.
- p. Click 'Next Structure' and repeat l. and m.
- q. Exit 'ANALYSIS' and 'Conformational Search'.

4. Energy-minimization and making .pdb files in QUANTA

- a. Open the .msf file for the first structure from the conformational search.
- b. Click 'CHARMm Minimization' in the 'Modeling' window.
- c. Save the minimized structure.
- d. Open 'Edit' then 'Molecular Editor'.
- e. Click 'Edit Bonds...' then 'Display Kekule Structure'.
- f. Save and exit 'Molecular Editor'.

- g. Open 'File' and choose 'Export'.
- h. Select 'Protein Data Bank' and save the structure in .pdb file.
- i. Open the .msf file for the second structure from the conformational search and repeat b. through i.
- j. Open 'File' and exit QUANTA.

5. Editing .pdb files for AMBER (4.1) use

- a. Open the .pdb file by *jot* in unix shell.
- b. Eliminate the fifth, tenth, and eleventh columns.
- c. Change the order of atoms according to the residue orders (i.g., Ace(1)-Tyr(2)-Pro(3)-NMe(4)) by examining the third column from the right (for the atoms types).
- d. Change the residue names for the acety- and N-methyl amide group atoms in the fourth column (i.e., TYR to ACE; PRO to NME).
- e. Change the residue numbers in the fifth column so that ACE=1, TYR=2, PRO=3, and NME=4.
- f. Change the atom types in the third column according to the amino acid database in AMBER (named all.in under /disk3/applications/amber41/dat directory).
- g. Save this file as 'YPcis.pdb'.

6. Creating Topology and Coordinate files for AMBER runs in XLEAP

- a. Type 'xleap', then an XLEAP shell appears on the screen.
- b. In the XLEAP shell, type in the following (note: 'YPcis' is a temporary name for the molecular system in the XLEAP session):

```
verbosity 1 [ENTER]
```

```
YPcis = loadpdb YPcis.pdb [ENTER]
```

```
SolvateBox YPcis WATBOX216 5 [ENTER]
```

```
saveamberparm YPcis YPsolSANDER.top YPsol.crd [ENTER]
```

```
saveamberparmpert YPcis YPsolGIBBS.top YPsol.crd [ENTER]
```

- c. Exit XLEAP by typing 'quit' or clicking 'Quit' button under the 'FILE' menu.

7. Making SANDER minimization input file (water molecules only; minimization 1)

- a. Open a *jot* window in unix.
- b. Type the following:

*** Top of File ***

Title	!	title for the run
&cntrl	!	input options start here
timlim = 10000000.,	!	time limit for the job (in seconds)
imin = 1	,	! number of minimization run
ntpr = 100	,	! every NTPR steps energy info printed
ntwx = -1	,	! no printing for coord. in separate file
ntwv = -1	,	! no printing for veloc. in separate file
ntwe = 100	,	! every NTWE steps energy + temp are printed
ntwxm = 0	,	! after NTWXM steps NTWX will be inactive
ntwvm = 0	,	! after NTWVM steps NTWV will be inactive
ntwem = 0	,	! after NTWEM steps NTWE will be inactive

```

ioutfm = 0           ,      ! velocity and coord. sets will be formatted

ntb = 2              ,      ! constant pressure option
idiel = 1            ,      ! constant dielectric function
dielc = 1.0          ,      ! dielectric constant = 1.0
scee = 1.2           ,      ! scale factor for 1-4 electrostatic int.

ibelly = 1           ,      ! belly option - only selected atoms move

maxcyc = 500         ,      ! maximum number of cycles of minimization
ncyc = 50             ,      ! after NCYC cycles change the method
ntmin = 1            ,      ! steepest descent + conjugate gradient used
dx0 = 0.01           ,      ! the initial step length
dxm = 0.5            ,      ! the maximum step length
drms = 0.0001        ,      ! convergence criterion for RMS

ntp = 1              ,      ! constant pressure dynamics option

jfastw = 1           ,      ! fast water algorithm option
watnam = 'IP3'       ,      ! water molecule name
&end                 ! end of options
Title                ! title for the belly option
ATOM 48 1901         ! only use molecules 48 to 1901 (waters)
END
END
*** End of File ***

```

c. Save the file as 'YPsol_min1.in'.

d. Make a script file as follows:

```

*** Top of File ***
rm YPsol_min1.cor
rm YPsol_min1.vel
rm YPsol_min1.en
rm YPsol_min1.inf
rm YPsol_min1.out
rm YPsol_min1.rst

```

```

/disk3/applications/amber41/exe/sander -i YPsol_min1.in \
-o YPsol_min1.out \

```

```
-p YpsolSANDER.top -c YPsol.crd \  
-r YPsol_min1.rst -x YPsol_min1.crd\  
-v YPsol_min1.vel -e YPsol_min1.en\  
-inf YPsol_min1.inf\  
*** End of File ***
```

- e. Save it as 'YPrunmin' and make it executable.
- f. Run it by typing 'nohup YPrunmin &'.

8. Making SANDER minimization input file (for the whole system; minimization 2)

- a. Open 'YPsol_min1.in' by *jot* in unix.
- b. Change the following:
 - 1) remove 'ibelly = 1';
 - 2) erase the belly option description at the bottom of file.
- c. Save the file as 'YPsol_min2.in'.
- d. Change 'YPrunmin': all 'min1' become 'min2' and 'YPsol.crd' becomes 'YPsol_min1.rst'.
- e. Save the script file.
- f. Run it by typing 'nohup YPrunmin &'.

9. Making SANDER dynamics input file (equilibration before slow-warming; dynamics 1)

- a. Open a *jot* window in unix.
- b. Type the following:

*** Top of File ***

```

Title                                     ! title for the run
&cntrl                                  ! input options start here
    timlim = 10000000.,                 ! time limit for the job (in seconds)

    ntp = 100                          ,   ! every NTPR steps energy info printed
    ntwx = -1                          ,   ! no printing for coord. in separate file
    ntwv = -1                          ,   ! no printing for veloc. in separate file
    ntwe = 100                         ,   ! every NTWE steps energy + temp are printed
    ntwxm = 0                         ,   ! after NTWXM steps NTWX will be inactive
    ntwvm = 0                         ,   ! after NTWVM steps NTWV will be inactive
    ntwem = 0                         ,   ! after NTWEM steps NTWE will be inactive
    ioutfm = 0                        ,   ! velocity and coord. sets will be formatted

    ntf = 3                           ,   ! all the bond interactions are omitted
    ntb = 2                           ,   ! constant pressure option
    idiel = 1                         ,   ! constant dielectric function is used
    dielc = 1.0                      ,   ! dielectric constant for electrostatic int.
    ntnb = 1                         ,   ! non-bonded pair list will be generated
    nsnb = 10                        ,   ! after NSNB steps non-bonded list updated
    scee = 1.2                       ,   ! scale factor for 1-4 electrostatic int.

    nstlim = 5000                    ,   ! number of dynamics steps
    dt = 0.002                      ,   ! time step (in psec)

    temp0 = 10.0                    ,   ! reference temperature
    tempi = 10.0                    ,   ! initial temperature (in kelvin)
    ntt = 5                         ,   ! constant temperature option

    ntp = 1                          ,   ! const pressure dynamics option

    ntc = 3                          ,   ! SHAKE is on - all bonds are constrained

    jfastw = 1                      ,   ! fast water algorithm option
    watnam = 'IP3'                  ,   ! water molecule name
&end                                ! terminating input options
*** End of File ***

```

c. Save the file as 'YPsol_min3.in'.

d. Change 'YPrunmin': all 'min2' become 'min3' and 'YPsol_min1.rst' becomes

'YPsol_min2.rst'.

- e. Save the script file.
- f. Run it by typing 'nohup YPrunmin &'.

10. Making SANDER dynamics input file (slow warming)

- a. Open 'YPsol_min3.in' by *jot* in unix.
- b. Change the following:
 - 1) add 'nmrmax = 1' (slow-warming flag)
 - 2) add 'ntx = 7' (X, V, and box sizes are read formatted)
 - 3) add 'nrun = 1' (dynamics flag)
 - 4) set 'nstlim = 10000'
 - 5) add 'init = 4' (starting velocity option)
 - 6) set 'temp0 = 300.0' (reference temperature)
- c. Add the following at the bottom of the file:

*** Top of File ***

&cntrl ! main input section

...

&end

&wt type = 'TEMP0', ! options for the gradual warming see p.91 in the manual
 istep1 = 1 ,
 istep2 = 5000 ,
 value1 = 10.0 ,
 value2 = 300.0 ,
 iinc = 5 ,

&end

```
&wt type = 'END'
&end
*** End of File ***
```

d. Save the file as 'YPsol_sander.in'.

e. Make a script file as follows:

```
*** Top of File ***
```

```
rm YPsol_sander.crd
rm YPsol_sander.en
rm YPsol_sander.inf
rm YPsol_sander.out
rm YPsol_sander.rst
rm YPsol_sander.vel
```

```
/disk3/applications/amber41/exe/sander -i YPsol_sander.in \
-o YPsol_sander.out -p YPsolSANDER.top \
-c YPsol_min3.rst -r YPsol_sander.rst \
-x YPsol_sander.crd -v YPsol_sander.vel \
-e YPsol_sander.en -inf YPsol_sander.inf
```

```
*** End of File ***
```

f. Save it as 'YPrunsander' and make it executable.

g. Run it by typing 'nohup YPrunsander &'.

11. Making SANDER dynamics input file (pressure equilibration; dynamics 2)

a. Open 'YPsol_sander.in' by *jot* in unix.

b. Change the following:

1) remove 'nmrmax = 1'

2) set 'nstlim = 50000'

3) set 'tempi = 300.0'

- c. Save the file as 'YPsol_sander2.in'.
- d. Change 'YPrunsander': all 'sander' become 'sander1' and 'YPsol_min3.rst' becomes 'YPsol_sander.rst'.
- f. Save the script file.
- g. Run it by typing 'nohup YPrunsander &'.

12. Making GIBBS input file (forward run; GIBBS 1)

- a. Open a *jot* window in unix.
- b. Type the following:

*** Top of File ***

Title	!	title for the run
&cntrl	!	input options start here
timlim = 10000000.,	!	time limit for the job (in seconds)
ntx = 7	,	! coord, V, box info read in AMBER format
ntb = 2	,	! constant pressure option
nrun = 1001	,	! number of dynamics run
ntt = 5	,	! const temperature option
temp0 = 300.0	,	! reference temperature
ntp = 1	,	! const pressure dynamics option
nstlim = 1000	,	! number of dynamics steps
init = 4	,	! dynamics starting procedure option
dt = 0.002	,	! time step (in psec)
ntc = 3	,	! SHAKE is on - all bonds are constrained
ncorc = 1	,	! constraints included free e. calculations
ishkfl = 6	,	! option for SHAKE failure

```

ntf = 3           ,      ! all the bond interactions are omitted
ntnb = 1          ,      ! non-bonded pair list will be generated
nsnb = 10         ,      ! after NSNB steps non-bonded list updated
idiel = 1         ,      ! constant dielectric function is used
intprt = 5        ,      ! energy contribution option

scee = 1.2        ,      ! scale factor for 1-4 electrostatic int.
dielc = 1.0       ,      ! dielectric constant for electrostatic int.

ntpr = 1000       ,      ! every NTPR steps energy info printed
ntwx = 100000     ,      ! no printing for coord. in separate file
ntwv = -1         ,      ! no printing for veloc. in separate file
ntwe = -1         ,      ! every NTWE steps energy + temp are printed
ntwxm = 0         ,      ! after NTWXM steps NTWX will be inactive
ntwvm = 0         ,      ! after NTWVM steps NTWV will be inactive
ntwem = 0         ,      ! after NTWEM steps NTWE will be inactive
ioutfm = 0        ,      ! velocity and coord. sets will be formatted
isande = 1        ,      ! enthalpies and entropies printed

ntr = 0           ,      ! MD without restraint
intr = 1          ,      ! additional constraints will be read

almlda = 0.0      ,      ! starting value of lambda
almdel = 0.001    ,      ! d_lambda value
isldyn = 3        ,      ! window calculation
nstmeq = 500      ,      ! number of steps for equilibration
nstmul = 500      ,      ! " " " for data collection
ndmpmc = 1        ,      ! stat info is dumped every ndmpmc windows
idwide = 0        ,      ! double-wide sampling is performed

jfastw = 1        ,      ! fast water algorithm option
watnam = 'IP3'    ,      ! water molecule name
&end              ,      ! terminating input options
  9  26  28  38  0  0  2  0.00000  1.00000
  0.00000  0.00000  0.00000  180.00000  0  0      ! constraint options; see p.134
*** End of File ***                                ! in manual

```

c. Save the file as YPsol_gibbs.in file.

d. Make a script file as follows:

```
*** Top of File ***  
PIN = YPsol_gibbs.in  
PPARM = YPsolGIBBS.top  
PINCRD = YPsol_sander1.rst
```

```
POUT = YPsol_gibbs.out  
PREST = YPsol_gibbs.rst  
PINFO = YPsol_gibbs.inf  
PVEL = YPsol_gibbs.vel  
PEN = YPsol_gibbs.en  
MICSTAT = YPsol_gibbs.sta  
*** End of File ***
```

- e. Save it as 'YPsol_gibfile'.
- f. Make another script file as follows:

```
*** Top of File ***  
/disk3/applications/amber41/exe/gibbs YPsol_gibfile  
*** End of File ***
```

- g. Save it as 'YPrungibbs' and make it executable.
- h. Run it by typing 'nohup YPrungibbs &'.

13. Making GIBBS input file (equilibration run; GIBBS 2)

- a. Open 'YPsol_gibbs.in' by *jot*.
- b. Change the following:
 - 1) set 'nrun = 1'
 - 2) set 'ntwx = 1000'
 - 3) set 'almda = 1.0'
 - 4) set 'almdel = 1.0'

5) set 'nstmeq = 1000'

6) set 'nstmul = 0'

7) change the bottom section as follows:

*** Top of File ***

Title

&cntrl ! main input section

....

&end

```
9 26 28 38 0 0 2 1.00000 0.00000
0.00000 180.00000 0.00000 180.00000 0 0
```

*** End of File ***

c. Save the file as YPsol_equil.in file.

d. Change 'YPsol_gibfile': all 'gibbs' become 'equil' and 'YPsol_sander2.rst' becomes 'YPsol_gibbs.rst'.

e. Save the script file.

f. Run it by typing 'nohup YPrungibbs &'.

14. Making GIBBS input file (reverse run; GIBBS 3)

a. Open 'YPsol_gibbs.in' by *jot*.

b. Change the bottom section as follows:

*** Top of File ***

Title

&cntrl ! main input section

....

&end

```
9 26 28 38 0 0 2 0.00000 1.00000
0.00000 180.00000 0.00000 0.00000 0 0
```

*** End of File ***

- c. Save the file as 'YPsol_back.in'.
- d. Change 'YPsol_gibfile': all 'equil' become 'back' and 'YPsol_gibbs.rst' becomes 'YPsol_equil.rst'.
- e. Save the script.
- f. Run it by typing 'nohup YPrungibbs &'.

APPENDIX B

PART I - Additional Calculation Results

Table 1 Free Energy Change for Ace-Tyr-Pro-NMe *in vacuo* by AMBER 4.0 in 100 Windows

Torsion (degs)	Free Energy (kcal/mol)							
	500/500*		500/5000		5000/500		5000/5000	
	Forwd	Bckwd	Forwd	Bckwd	Forwd	Bckwd	Forwd	Bckwd
0	0.27	0.00	0.29	0.00	0.36	0.00	0.27	0.00
9	1.78	1.66	1.44	0.83	1.72	1.12	1.45	0.58
18	3.55	3.28	2.78	1.43	3.07	1.70	2.81	1.10
27	5.60	5.02	4.27	2.06	4.56	2.34	4.30	1.68
36	7.87	6.84	5.95	2.83	6.26	3.11	5.96	2.42
45	10.35	8.71	7.98	3.91	8.27	4.16	7.97	3.46
54	12.97	10.61	10.28	5.25	10.62	5.54	10.37	4.88
63	15.25	12.31	12.66	6.71	12.99	6.99	12.74	6.32
72	17.26	13.55	14.62	7.73	14.95	8.01	14.68	7.30
81	18.95	14.36	16.28	8.45	16.62	8.74	16.33	7.99
90	20.38	14.93	17.74	8.97	18.08	9.26	17.78	8.48
99	21.54	15.20	18.94	9.21	19.28	9.49	18.96	8.69
108	22.24	14.94	19.71	8.93	20.04	9.18	19.72	8.38
117	22.32	13.93	19.75	7.75	20.07	7.97	19.72	7.18
126	21.55	11.97	18.99	5.69	19.30	5.89	18.95	5.09
135	20.15	9.46	17.66	3.08	17.93	3.23	17.61	2.45
144	18.04	6.28	15.62	0.00	16.00	0.20	15.68	-0.59
153	14.75	2.69	13.40	-2.71	13.77	-2.53	13.48	-3.30
162	11.68	-1.00	11.39	-5.56	11.91	-5.34	11.31	-6.15
171	10.29	-3.68	11.27	-7.76	11.81	-7.53	10.63	-8.83
180	10.20	-5.20	12.16	-8.70	12.68	-8.46	11.00	-10.38

Note: * for (the number of steps in equilibration/data collection)

Table 2 Free Energy Change for Ace-Tyr-Pro-NMe *in vacuo* by AMBER 4.0 in 500 Windows

Torsion (degs)	Free Energy (kcal/mol)							
	500/500*		500/5000		5000/500		5000/5000	
	Forwd	Bckwd	Forwd	Bckwd	Forwd	Bckwd	Forwd	Bckwd
0	0.05	0.00	0.06	0.00	0.07	0.00	0.05	0.00
9	0.34	0.20	1.00	0.81	1.08	0.87	0.93	0.72
18	0.66	0.36	2.02	1.64	2.09	1.69	1.93	1.53
27	1.31	0.83	3.13	2.56	3.20	2.61	3.03	2.45
36	2.38	1.69	4.41	3.66	4.50	3.72	4.29	3.52
45	4.02	3.13	6.02	5.07	6.11	5.13	5.92	4.94
54	6.04	4.95	7.86	6.71	7.95	6.78	7.85	6.68
63	8.17	6.90	9.92	8.59	10.01	8.65	9.92	8.56
72	9.90	8.45	11.56	10.03	11.64	10.09	11.55	10.01
81	11.23	9.59	12.88	11.17	12.96	11.22	12.87	11.14
90	12.19	10.36	14.00	12.09	14.08	12.14	13.99	12.06
99	12.78	10.77	14.87	12.77	14.95	12.82	14.86	12.73
108	12.95	10.75	15.31	12.99	15.38	13.04	15.28	12.94
117	12.57	10.19	14.93	12.37	15.01	12.42	14.86	12.27
126	11.52	8.93	13.70	10.88	13.79	10.94	13.59	10.74
135	9.83	6.66	11.87	8.79	11.95	8.84	11.71	8.61
144	7.39	4.24	9.48	6.17	9.58	6.24	9.36	6.05
153	2.89	-0.69	7.19	3.78	7.29	3.85	7.10	3.68
162	-1.27	-5.30	4.87	1.35	4.97	1.41	4.90	1.27
171	-4.41	-8.81	3.51	-0.42	3.63	-0.33	4.24	0.16
180	-6.05	-10.78	3.39	-0.97	3.52	-0.87	4.81	0.33

Note: * for (the number of steps in equilibration/data collection)

Table 3 Free Energy Change for Ace-Tyr-Pro-NMe *in vacuo* by AMBER 4.0 in 1000 Windows

Torsion (degs)	Free Energy (kcal/mol)							
	500/500*		500/5000		5000/500		5000/5000	
	Forwd	Bckwd	Forwd	Bckwd	Forwd	Bckwd	Forwd	Bckwd
0	0.03	0.00	0.03	0.00	0.03	0.00	0.03	0.00
9	1.04	0.97	0.87	0.77	0.88	0.76	0.87	0.76
18	2.02	1.86	1.82	1.62	1.82	1.61	1.85	1.64
27	3.07	2.82	2.87	2.58	2.87	2.57	2.92	2.62
36	4.30	3.96	4.09	3.69	4.08	3.69	4.15	3.76
45	5.86	5.42	5.66	5.17	5.66	5.16	5.71	5.23
54	7.65	7.11	7.55	6.96	7.54	6.95	7.59	7.00
63	9.62	8.98	9.57	8.89	9.57	8.88	9.61	8.93
72	11.21	10.48	11.17	10.39	11.16	10.38	11.20	10.42
81	12.49	11.67	12.45	11.58	12.45	11.57	12.48	11.61
90	13.57	12.65	13.52	12.55	13.52	12.54	13.55	12.58
99	14.40	13.38	14.34	13.28	14.34	13.27	14.37	13.30
108	14.79	13.67	14.72	13.55	14.72	13.53	14.73	13.55
117	14.40	13.16	14.24	12.95	14.25	12.94	14.23	12.93
126	13.18	11.81	12.91	11.49	12.92	11.48	12.86	11.43
135	11.37	9.87	10.97	9.41	10.97	9.40	10.87	9.32
144	8.94	7.31	8.04	6.37	8.04	6.36	7.92	6.26
153	6.53	4.85	3.58	1.75	3.59	1.74	3.86	2.05
162	4.17	2.44	-0.20	-2.20	-0.20	-2.20	-0.12	-2.08
171	2.73	0.82	-2.39	-4.49	-2.37	-4.49	-1.87	-3.95
180	2.85	0.74	-3.54	-5.76	-3.53	-5.75	-2.47	-4.70

Note: * for (the number of steps in equilibration/data collection)

Table 4 Free Energy Change for Ace-Tyr-Pro-NMe *in vacuo* by AMBER 4.0 in 10000 Windows

Torsion (degrees)	Free Energy (kcal/mol)	
	500/500*	
	Forward	Backward
0	0.00	0.00
9	0.83	0.82
18	1.77	1.75
27	2.80	2.77
36	4.00	3.96
45	5.51	5.46
54	7.42	7.36
63	9.41	9.34
72	10.96	10.88
81	12.19	12.11
90	13.22	13.12
99	14.00	13.90
108	14.32	14.21
117	13.77	13.64
126	12.35	12.21
135	10.32	10.16
144	7.60	7.44
153	4.51	4.34
162	1.42	1.25
171	-0.34	-0.53
180	-0.62	-0.83

Note: * for (the number of steps in equilibration/data collection)

Table 5 Gibbs Free Energy Change for Ace-Tyr-Pro-NMe in water by AMBER 4.1 (1000 windows; 500 equilibration and 500 data collection steps)

Torsion (degrees)	Gibbs Free Energy (kcal/mol)			
	<i>cis</i> → <i>trans</i>		<i>trans</i> → <i>cis</i>	
	Forward	Backward	Forward	Backward
0	0.01	0.00	1.02	0.85
9	0.48	0.47	1.51	1.36
18	0.97	0.95	2.11	1.95
27	1.55	1.53	2.67	2.51
36	2.16	2.14	3.40	3.25
45	2.84	2.83	4.23	4.08
54	3.64	3.61	4.90	4.75
63	4.47	4.44	5.79	5.66
72	5.24	5.21	6.63	6.50
81	6.06	6.04	7.39	7.25
90	6.56	6.52	8.08	7.93
99	6.68	6.63	8.28	8.13
108	6.72	6.67	8.16	8.00
117	5.40	5.35	6.85	6.71
126	3.43	3.39	4.88	4.76
135	1.64	1.60	3.14	2.99
144	0.34	0.28	1.46	1.31
153	-1.02	-1.09	-0.33	-0.48
162	-2.22	-2.28	-1.81	-1.94
171	-2.84	-2.93	-2.91	-3.04
180	-3.26	-3.37	-3.26	-3.37

Table 6 Gibbs Free Energy Change for Ace-Asn-Pro-NMe in water by AMBER 4.1 (1000 windows; 500 equilibration and 500 data collection steps)

Torsion (degrees)	Gibbs Free Energy (kcal/mol)			
	<i>cis</i> → <i>trans</i>		<i>trans</i> → <i>cis</i>	
	Forward	Backward	Forward	Backward
0	0.01	0.00	0.27	0.15
9	0.29	0.25	0.70	0.59
18	0.84	0.77	1.20	1.09
27	1.60	1.47	1.79	1.65
36	2.32	2.22	2.38	2.28
45	3.14	3.02	3.05	2.98
54	4.02	3.88	3.63	3.60
63	4.67	4.52	4.40	4.40
72	5.22	5.14	5.21	5.22
81	5.79	5.78	5.97	6.01
90	6.34	6.34	6.56	6.53
99	6.66	6.68	6.79	6.79
108	5.68	5.67	6.35	6.28
117	4.44	4.44	5.31	5.20
126	2.86	2.81	4.31	4.18
135	1.51	1.40	2.67	2.54
144	-0.06	-0.14	0.95	0.84
153	-1.34	-1.44	-0.72	-0.78
162	-2.68	-2.73	-2.28	-2.35
171	-3.46	-3.54	-3.30	-3.33
180	-3.74	-3.76	-3.74	-3.76

APPENDIX C

PART II - Detailed Methods

All the work was done on a Silicon Graphics Personal Iris Model 4D/35 except that AMSOL runs performed on an IBM RS6000 Workstation in the Computing Center.

Outline of the calculation methods:

- 1) Build each peptide in QUANTA (12 peptides in total).
- 2) Perform the conformational search generating 500 structures for each peptide.
Take the five lowest-energy structures each (60 peptides in total).
- 3) Calculate atomic charges by two different methods (ESP and Mulliken analysis) for each conformer in MOPAC.
- 4) Copy the calculated atomic charges into .bgf files, creating two new .bgf files for each conformer (120 files in total).
- 5) Calculate the ΔG_{soln} for each peptide conformer in POLARIS.
- 6) Also calculate ΔG_{soln} for some of the peptides in AMSOL.

Detailed calculation methods:

1. Building a peptide in QUANTA (version 3.3.1)
 - a. Open 'Applications' window and select 'Sequence Builder'.

- b. Open the amino acid residue library by choosing 'AMINOH.RTF'.
- c. Click 'PHE' to make Ace-Phe-NH₂.
- d. Open 'Edit' to choose 'Apply Patches to Residues'
- e. Click 'NACT' for an acetyl blocking group, then click 'PHE' in the sequence.
- f. Click 'CAMD' for an amide blocking group, then click 'PHE' in the sequence.
- g. Open 'SequenceBuilder' window, and choose 'Return to Molecular Modeling'.
- h. Save the structure.

2. Adding methyl groups to the peptide in QUANTA

- a. Open 'Edit' and 'Molecular Editor'.
- b. Select 'Add Atoms...'
- c. Click 'Convert to Methyl'.
- d. Click the backbone-hydrogen atom to which a methyl group is attached.
- e. Click 'Save and Exit' to save the new molecule.
- f. Repeat a. through e. to add more methyl groups one by one to the peptide.

3. Conformational Search in QUANTA

- a. Open 'CHARMm' and 'Minimization Options'.
- b. Select 'Adopted-Basis Newton Raphson'.
- c. Enter 3000 and 0.001 for 'Number of Minimization Steps' and 'Energy Gradient Tolerance', respectively.

- d. Open 'Applications' window and select 'Conformational Search'.
- e. Click 'Torsions...' and 'Define All Torsions', then 'Select Torsions...'.
- f. Select 'All Torsions' then 'Finish'. Close 'Torsions'.
- g. Click 'Setup Search...' then 'Random Sampling'.
- h. Enter 500 and 360 for 'Number of Samples' and 'Torsion Angle Window', respectively.
- i. Select 'Minimize Each Structure', 'Cumulative Random Modifications', and 'Display Each Structure'.
- j. Click 'Do Search'. Define .csr file name.
- k. After the search, select 'ANALYSIS'.
- l. Open the search (.csr) file.
- m. Select 'Sort By Property...' in the 'ANALYSIS' window.
- n. Click 'Use Sorted Order' then 'First Structure'.
- o. Choose 'File Options/Filters...' and write an .msf file.
- p. Click 'Next Structure' and repeat l. and m.
- q. Exit 'ANALYSIS' and 'Conformational Search'.

4. Energy-minimization and making .bgf files in QUANTA

- a. Open the .msf file for the first structure from the conformational search.
- b. Click 'CHARMm Minimization' in the 'Modeling' window.
- c. Save the minimized structure.

- d. Open 'Edit' then 'Molecular Editor'.
 - e. Click 'Edit Bonds...' then 'Display Kekule Structure'.
 - f. Save and exit 'Molecular Editor'.
 - g. Open 'File' and choose 'Export'.
 - h. Select 'BioDesign BGF' to save the structure in .bgf format.
 - i. Open the .msf file for the other structures from the conformational search and repeat b. through h. for each of the five lowest-energy conformers.
 - j. Repeat 1-a. through 3-i. for all peptides.
 - 1) For di-Phe peptides, click 'PHE' twice at 1-c.
 - 2) For tri-Phe peptides, click 'PHE' three times at 1-c.
- This procedure creates 60 Phe-peptides.
- k. Copy the .bgf files to the BIOGRAF directory.

5. Making MOPAC/AMSOL input files in BIOGRAF (version 3.21)

- a. Click 'in-out' in the 'Top menu'.
- b. Select 'read', 'BioDesign', then a .bgf file name.
- c. Click 'write', 'mopac', 'yes', then enter the MOPAC file name. This writes a file in cartesian coordinates.
- d. Click 'return', 'utilities', then 'program reset'.
- e. Click 'OK'.
- f. Repeat a. through e. for the rest of .bgf files.

- g. Click 'exit' then 'OK' to exit BIOGRAF.

6. Editing MOPAC input files

- a. Copy the MOPAC input (.mpc) files to the MOPAC directory,
'/usr/local/455_sgrw/mopac6.0/bin'.
- b. Change the directory to the above MOPAC directory.
- c. Open a .mpc file using the *jot* editor.
- d. Type in the following keywords:

```
*** Top of File ***
XYZ MNDO !SCF GEO-OK MMOK NOXYZ NOINTER      ! MOPAC keywords
information                                     ! run info line
                                                ! blank line
N  0.3629 1  0.9600 1  -6.8706 1              ! coordinate info
.....                                         ! (Cartesian coord)
*** End of File ***
```

- e. Rename the file to a .dat file.
- f. Run MOPAC point energy calculation by typing:

nohup RunMopac filename > & log &

where 'filename' does not have an extension part (= .dat).
- g. Using the output file resulting from the previous MOPAC run, create a similar input file using its internal coordinates which appear toward the top of the output file.
- h. At the top of the new input file, enter the new keywords as follows:

```
*** Top of File ***
```

```

MNDO ESP GNORM=0.0 EF GEO-OK MMOK NOXYZ NOINTER  ! keywords
information                                         ! run info line
                                                    ! blank line
N                                                    ! internal coordinate starts from the fourth line
H  0.99712  1                                     1
C  1.45764  1 119.29365  1                         1  2
H  1.09212  1 107.96183  1 140.23335  1  3  1  2
...
*** End of File ***

```

- i. Repeat c. through h. for all .mpc files.

7. Running MOPAC (version 6.0) and editing .bgf file charges

- a. Execute MOPAC calculations by typing:

nohup RunMopac filename > & log &

where 'filename' does not have an extension part (= .dat).

- b. Repeat a. for all .dat files. Up to four to five simultaneous runs are possible according to the molecular size on Personal IRIS 4D/35.
- c. After completion of each run, open the output file (.out) using the *jot* editor.
- d. Find the section 'NET ATOMIC CHARGES AND DIPOLE CONTRIBUTIONS'.
- e. Copy the atomic charges, which appear under 'CHARGE', to the corresponding original .bgf file created in QUANTA by replacing the old charge values. Save the new .bgf file.
- g. Find the section 'ELECTROSTATIC POTENTIAL CHARGES'.
- h. Copy the electrostatic potential (ESP) charges, which appear under

'CHARGE', to the corresponding original .bgf file created in QUANTA by replacing the old charge values. Save the second new .bgf file.

- i. Repeat a. through h. for all 60 .bgf files, making 120 new .bgf files.

8. Calculating ΔG_{solv} in POLARIS (version 3.0.1)

- a. Copy all the new .bgf files to the POLARIS directory.
- b. In 'Top menu:', click 'in-out', 'read', 'BioDesign', then select a .bgf file name. Click 'OK' and 'return'.
- c. Select 'build', 'organic' or 'peptide', depending on the molecule of interest, then the file name in 'Files:'.
- d. Click 'edit id', 'auto type', then 'auto'. Keep clicking 'return' five times to go to 'Top menu:'.
- e. Select 'simulate', 'electrostatics', 'general', and 'output type' to make the output type 'verbose'. Click 'return', 'energy', 'modify run' three times to have 'multiple run', then 'return' twice.
- f. Choose 'region I', the file name, 'return', then 'current charges'.
- g. Click 'dG solvation'. In the unix window read the ΔG_{solv} value from toward the bottom of the output under 'PDL MICROSCOPIC ESTIMATE' section where two 'dG' values appear, and the one on the right corresponds the ΔG_{solv} .
- h. Click anywhere in the black window to stop the molecular rotation, then

'return'.

- i. Select 'utilities', 'program reset', then 'OK'.
- j. Repeat b. through i. for all .bgf files.
- k. To exit from POLARIS, click 'exit' in 'Top menu:', then 'OK'.

9. Editing MOPAC files (.mpc) for AMSOL (version 4.5)

- a. In the MOPAC directory, edit the .dat file by changing 'MNDO' to 'AM1' in the MOPAC keyword section (in the first line). Rename the .dat file.

- b. Run MOPAC calculation by typing:

```
nohup RunMopac filename >> & log &
```

where 'filename' does not have an extension part (= .dat).

- c. Using the output file resulting from the previous MOPAC run, take the 'FINAL HEAT OF FORMATION' value which appears after the energy-minimization steps.
- d. Open a .mpc file using the *jot* editor.
- e. Type in the following keywords for the AMSOL run:

*** Top of File ***

XYZ AM1 SM2 DERINU TRUES T=2000000 PRECISE GEO-OK MMOK NOXYZ

NOINTER ! MOPAC keywords in the first line

information ! run info in the second line

! blank in the third line

-66.62926 ! heat of formation from MOPAC run

N 0.3629 1 0.9600 1 -6.8706 1 ! Cartesian coord info from the fifth

line

....

*** End of File ***

- f. Rename the file to a new .dat file.
- g. Repeat c. through h. for the selected .mpc files.

10. Running AMSOL calculation

- a. Run AMSOL by typing the following:

nice amsoli filename > filename.out &

- b. After the completion of the run, take the ΔG_{soln} value which appears toward the bottom of the output file.

11. AMSOL 1SCF calculation

- a. Repeat 9. a. through d., except that the .dat file header looks like the following:

*** Top of File ***

AM1 ESP GNORM=0.0 EF GEO-OK MMOK NOINTER ! MOPAC keywords
information ! run info line
! blank line

N ! internal coordinate starts from the fourth line

H 0.99712 1 1

C 1.45764 1 119.29365 1 2

H 1.09212 1 107.96183 1 140.23335 1 3 1 2

...

*** End of File ***

- b. Make a file which contains the following:

*** Top of File ***

XYZ 1SCF AM1 SM2 DERINU TRUES PRECISE GEO-OK MMOK NOXYZ

NOINTER
information
*** End of File ***

!MOPAC keywords in the first line
! information in the second line

- c. Leave the third column as a blank and add the heat of formation value from the previous MOPAC run on the fourth line.
- d. Append the optimized Cartesian coordinates from the previous MOPAC .out file. Rename the file to .dat. This file looks like the following:

*** Top of File ***

XYZ 1SCF AM1 SM2 DERINU TRUES PRECISE GEO-OK MMOK NOXYZ
NOINTER
information

-60.76583

N 0.0000 1 0.0000 1 0.0000 1

H 0.9927 1 0.0000 1 0.0000 1

C -0.6474 1 1.2882 1 0.0000 1

....

*** End of File ***

- e. Repeat a. through d. for the selected .mpc files.
- f. Run AMSOL as in 10.

APPENDIX D

PART II - Additional Calculation Results

Table 1 POLARIS ΔG_{solv} Calculation with MNDO and Atomic Charges by ESP and Mulliken Methods

Pep. Name	ΔG_{solv} (kcal/mol)			
	ESP	SD*	Mulliken	SD
F01	-11.314	0.376	-8.739	0.305
02	-10.741		-8.071	
03	-11.375		-8.614	
04	-11.794		-8.845	
05	-11.369		-8.711	
F11	-10.988	0.713	-9.042	0.654
12	-9.341		-7.433	
13	-11.132		-8.938	
14	-10.273		-8.159	
15	-10.578		-8.531	
F21	-8.400	0.509	-7.071	0.414
22	-9.672		-8.159	
23	-9.506		-7.975	
24	-9.352		-7.815	
25	-8.958		-7.797	
FF01	-16.227	0.671	-12.459	0.741
02	-17.137		-13.288	
03	-17.446		-14.182	
04	-16.026		-12.467	
05	-17.395		-13.569	

Table 1 - continuation

Pep. Name	ΔG_{soln} (kcal/mol)			
	ESP	SD*	Mulliken	SD
FF11	-15.780	0.298	-12.823	0.430
12	-15.179		-12.014	
13	-15.224		-12.347	
14	-15.224		-11.975	
15	-14.981		-11.703	
FF21	-15.838	0.812	-12.507	0.536
22	-14.289		-11.770	
23	-14.820		-12.263	
24	-13.828		-11.164	
25	-15.398		-12.269	
FF31	-13.639	0.578	-12.212	0.464
32	-13.731		-11.296	
33	-14.975		-12.304	
34	-13.632		-11.419	
35	-14.233		-12.030	
FFF01	-20.612	1.124	-15.454	0.834
02	-20.986		-16.565	
03	-20.047		-16.098	
04	-18.652		-15.541	
05	-18.528		-14.343	
FFF11	-22.237	1.905	-17.256	1.448
12	-20.506		-16.088	
13	-17.073		-13.310	
14	-20.790		-16.018	
15	-19.715		-15.637	
FFF21	-18.733	0.722	-14.853	1.059
22	-18.900		-14.256	
23	-20.537		-16.983	
24	-19.192		-15.115	
25	-19.051		-14.659	

Table 1 - continuation

Pep. Name	ΔG_{soln} (kcal/mol)			
	ESP	SD*	Mulliken	SD
FFF31	-17.751	0.984	-14.643	0.872
32	-16.408		-13.199	
33	-16.829		-13.673	
34	-17.978		-14.407	
35	-18.903		-15.446	
FFF41	-16.188	1.495	-13.197	1.456
44	-15.565		-11.983	
43	-17.610		-14.624	
44	-19.045		-15.211	
45	-18.562		-15.394	

* Standard Deviation of values from five conformers

Table 2 AMSOL Calculation Results with 1SCF

Pep. Name	ΔG_{soln} (kcal/mol)	SD (kcal/mol)*
F01	-12.896	0.001
F02	-12.897	
F03	-12.897	
F04	-12.895	
F05	-12.899	
F11	-12.979	0.002
F12	-12.983	
F13	-12.978	
F14	-12.979	
F15	-12.980	
F21	-11.341	0.001
F22	-11.340	
F23	-11.340	
F24	-11.339	
F25	-11.339	
FF01	-15.491	0.001
FF02	-15.492	
FF03	-15.490	
FF04	-15.492	
FF05	-15.490	
FF11	-15.648	0.073
FF12	-15.490	
FF13	-15.640	
FF14	-15.649	
FF15	-15.671	
FF21	-17.514	1.023
FF22	-15.191	
FF23	-15.189	
FF24	-15.192	
FF25	-15.353	

Table 2 - continuation

Pep. Name	ΔG_{solv} (kcal/mol)	SD (kcal/mol)*
FF31	-14.036	0.288
FF32	-14.037	
FF33	-13.500	
FF34	-13.521	
FF35	-13.513	
FFF01	-18.333	1.478
FFF02	-16.205	
FFF03	-19.127	
FFF04	-16.647	
FFF05	-15.636	
FFF11	-18.443	2.578
FFF12	-17.898	
FFF13	-12.167	
FFF14	-17.248	
FFF15	-17.824	
FFF21	-17.756	1.197
FFF22	-15.071	
FFF23	-17.815	
FFF24	-17.059	
FFF25	-17.924	
FFF31	-13.887	2.146
FFF32	-14.381	
FFF33	-17.657	
FFF34	-15.352	
FFF35	-18.859	
FFF41	-16.792	1.424
FFF42	-14.275	
FFF43	-17.412	
FFF44	-17.313	
FFF45	-17.858	

* Standard Deviation of values from five conformers

Table 3 AMSOL Test Calculations with 1SCF

Pep. Name	ΔG_{soln} (kcal/mol)	SD (kcal/mol)*
test11	-15.347	0.003
test12	-15.350	
test13	-15.354	
test14	-15.352	
test15	-15.347	
test21	-14.430	0.000
test22	-14.431	
test23	-14.431	
test24	-14.432	
test25	-14.431	
test31	-13.721	0.002
test32	-13.726	
test33	-13.726	
test34	-13.725	
test35	-13.722	
test41	-11.311	0.001
test42	-11.310	
test43	-11.312	
test44	-11.312	
test45	-11.310	

* Standard Deviation of values from five conformers

Table 4 AMSOL Results - Dipole Moment and Surface Area

Pep. Name	Dipole Moment (debyes)				Surface	
	in vacuum	SD*	in water	SD	Area (Å ²)	SD
test11	3.056	0.002	4.134	0.002	381.98	0.01
test12	3.054		4.132		381.98	
test13	3.052		4.129		382.00	
test14	3.053		4.131		381.99	
test15	3.055		4.133		381.97	
test21	2.787	0.002	3.466	0.001	399.74	0.01
test22	2.784		3.466		399.76	
test23	2.783		3.466		399.76	
test24	2.784		3.468		399.75	
test25	2.786		3.467		399.74	
test31	3.240	0.002	3.640	0.002	423.15	0.03
test32	3.244		3.637		423.10	
test33	3.242		3.636		423.13	
test34	3.238		3.636		423.18	
test35	3.243		3.639		423.11	
test41	2.399	0.001	4.024	0.003	394.15	0.00
test42	2.399		4.030		394.14	
test43	2.399		4.027		394.15	
test44	2.401		4.022		394.15	
test45	2.399		4.029		394.15	
F01	3.503	0.002	4.508	0.003	405.32	0.03
F02	3.499		4.502		405.38	
F03	3.502		4.506		405.33	
F04	3.504		4.508		405.31	
F05	3.503		4.508		405.32	
F11	3.680	0.000	4.856	0.000	412.01	0.01
F12	3.680		4.856		412.03	
F13	3.680		4.856		412.00	
F14	3.681		4.856		412.02	
F15	3.680		4.856		412.02	

Table 4 - continuation

Pep. Name	Dipole Moment (debyes)				Surface	
	in vacuum	SD*	in water	SD	Area (Å ²)	SD
F21	3.435	0.000	4.376	0.000	429.69	0.01
F22	3.436		4.376		429.70	
F23	3.435		4.376		429.70	
F24	3.435		4.375		429.69	
F25	3.435		4.376		429.68	
FF01	5.523	0.001	7.216	0.001	558.22	0.01
FF02	5.521		7.217		558.20	
FF03	5.522		7.216		558.21	
FF04	5.520		7.214		558.21	
FF05	5.520		7.214		558.20	
FF11	4.081	1.113	5.279	1.431	593.82	9.96
FF12	4.083		5.303		592.58	
FF13	4.107		5.314		593.90	
FF14	4.080		5.280		593.81	
FF15	6.576		8.493		571.30	
FF21	5.281	0.427	7.466	0.736	576.41	14.53
FF22	4.317		5.798		602.90	
FF23	4.317		5.797		602.92	
FF24	4.318		5.798		602.88	
FF25	4.357		5.896		576.32	
FF31	4.084	0.296	5.294	0.011	616.79	0.15
FF32	4.085		5.296		616.80	
FF33	3.542		5.270		617.11	
FF34	3.548		5.281		617.02	
FF35	3.545		5.275		617.07	
FFF01	2.756	2.474	3.764	2.997	691.73	19.43
FFF02	9.031		11.174		701.52	
FFF03	4.631		6.146		674.31	
FFF04	3.295		4.024		719.65	
FFF05	4.471		5.449		720.11	

Table 4 - continuation

Pep. Name	Dipole Moment (debyes)				Surface	
	in vacuum	SD*	in water	SD	Area (Å ²)	SD
FFF11	7.525	2.115	9.290	2.456	757.13	21.05
FFF12	4.601		5.992		715.70	
FFF13	1.759		2.559		700.03	
FFF14	3.342		4.556		719.53	
FFF15	4.138		5.460		727.34	
FFF21	3.434	0.511	4.419	0.719	718.46	15.67
FFF22	3.501		4.676		743.42	
FFF23	4.238		5.687		739.70	
FFF24	3.103		4.223		705.13	
FFF25	4.236		5.749		725.57	
FFF31	2.672	0.516	3.512	0.743	726.98	9.06
FFF32	2.881		3.781		710.87	
FFF33	3.886		5.321		731.23	
FFF34	2.613		3.582		711.68	
FFF35	2.897		4.182		718.96	
FFF41	3.803	1.433	5.037	1.768	737.68	8.52
FFF42	1.808		2.625		752.80	
FFF43	4.611		6.106		729.06	
FFF44	1.331		1.982		739.19	
FFF45	3.902		5.084		738.62	

* Standard Deviation of values from five conformers

Table 5 AMSOL Results - Components of Solvation Free Energy

Pep. Name	Energies (kcal/mol)					
	E-EN(g)	E-EN(aq)	G-P(aq)	G-ENP(aq)	G-CDS(aq)	G-S(aq)
test11	-60.766	-58.553	-11.925	-70.478	-5.636	-76.113
test12	-60.766	-58.552	-11.925	-70.478	-5.638	-76.116
test13	-60.766	-58.552	-11.927	-70.479	-5.641	-76.120
test14	-60.766	-58.552	-11.925	-70.478	-5.640	-76.118
test15	-60.766	-58.553	-11.925	-70.477	-5.635	-76.113
test21	-53.202	-50.561	-12.971	-63.533	-4.099	-67.632
test22	-53.202	-50.561	-12.973	-63.534	-4.099	-67.633
test23	-53.202	-50.561	-12.973	-63.534	-4.099	-67.633
test24	-53.202	-50.560	-12.975	-63.535	-4.099	-67.633
test25	-53.202	-50.561	-12.973	-63.534	-4.099	-67.633
test31	-57.281	-54.888	-13.020	-67.908	-3.094	-71.003
test32	-57.281	-54.887	-13.021	-67.908	-3.099	-71.008
test33	-57.281	-54.887	-13.022	-67.909	-3.099	-71.008
test34	-57.281	-54.888	-13.020	-67.908	-3.099	-71.006
test35	-57.281	-54.887	-13.020	-67.908	-3.095	-71.003
test41	-63.147	-61.358	-10.948	-72.307	-2.151	-74.458
test42	-63.147	-61.358	-10.950	-72.309	-2.149	-74.458
test43	-63.147	-61.358	-10.951	-72.309	-2.150	-74.459
test44	-63.147	-61.358	-10.949	-72.307	-2.152	-74.459
test45	-63.147	-61.358	-10.950	-72.309	-2.149	-74.457
F01	-66.629	-65.079	-9.756	-74.835	-4.691	-79.526
F02	-66.629	-65.079	-9.754	-74.833	-4.693	-79.526
F03	-66.629	-65.079	-9.757	-74.836	-4.691	-79.526
F04	-66.629	-65.080	-9.754	-74.834	-4.690	-79.524
F05	-66.629	-65.079	-9.758	-74.837	-4.691	-79.528
F11	-58.476	-56.289	-11.687	-67.975	-3.479	-71.455
F12	-58.476	-56.288	-11.691	-67.979	-3.480	-71.459
F13	-58.475	-56.289	-11.685	-67.914	-3.479	-71.454
F14	-58.476	-56.289	-11.686	-67.975	-3.479	-71.455
F15	-58.476	-56.289	-11.688	-67.976	-3.479	-71.456

Table 5 - continuation

Pep. Name	Energies (kcal/mol)					
	E-EN(g)	E-EN(aq)	G-P(aq)	G-ENP(aq)	G-CDS(aq)	G-S(aq)
F21	-54.967	-52.592	-12.772	-65.364	-0.944	-66.308
F22	-54.967	-52.592	-12.771	-65.364	-0.944	-66.307
F23	-54.967	-52.592	-12.772	-65.364	-0.944	-66.308
F24	-54.967	-52.593	-12.770	-65.363	-0.944	-66.306
F25	-54.967	-52.592	-12.770	-65.363	-0.944	-66.306
FF01	-80.577	-78.405	-12.975	-91.379	-4.689	-96.068
FF02	-80.577	-78.404	-12.975	-91.380	-4.689	-96.068
FF03	-80.577	-78.405	-12.973	-91.379	-4.688	-96.067
FF04	-80.577	-78.406	-12.974	-91.380	-4.689	-96.069
FF05	-80.577	-78.406	-12.972	-91.378	-4.689	-96.068
FF11	-69.618	-67.556	-14.121	-81.677	-3.589	-85.266
FF12	-69.601	-67.541	-14.019	-81.560	-3.532	-85.091
FF13	-69.617	-67.545	-14.125	-81.670	-3.587	-85.257
FF14	-69.618	-67.556	-14.123	-81.679	-3.589	-85.268
FF15	-67.694	-65.142	-15.635	-80.777	-2.588	-83.365
FF21	-57.964	-54.288	-18.218	-72.505	-2.972	-75.477
FF22	-62.448	-59.942	-15.422	-75.365	-2.274	-77.639
FF23	-62.448	-59.942	-15.421	-75.364	-2.273	-77.637
FF24	-62.448	-59.942	-15.424	-75.366	-2.274	-77.640
FF25	-62.319	-59.777	-15.430	-75.207	-2.466	-77.673
FF31	-58.940	-56.211	-16.586	-72.797	-0.179	-72.976
FF32	-58.940	-56.211	-16.587	-72.798	-0.179	-72.977
FF33	-35.129	-29.724	-18.048	-47.772	-0.856	-48.629
FF34	-35.130	-29.725	-18.069	-47.794	-0.857	-48.651
FF35	-35.129	-29.728	-18.058	-47.786	-0.856	-48.643
FFF01	-95.350	-93.259	-16.041	-109.300	-4.384	-113.084
FFF02	-94.749	-92.747	-15.005	-107.752	-3.202	-110.954
FFF03	-92.128	-89.253	-18.061	-107.314	-3.942	-111.255
FFF04	-95.273	-93.698	-14.526	-108.223	-3.697	-111.921
FFF05	-95.364	-93.655	-14.101	-107.756	-3.244	-111.000

Table 5 - continuation

Pep. Name	Energies (kcal/mol)					
	E-EN(g)	E-EN(aq)	G-P(aq)	G-ENP(aq)	G-CDS(aq)	G-S(aq)
FFF11	-84.037	-81.051	-19.071	-100.122	-2.358	-102.480
FFF12	-83.937	-81.225	-16.974	-98.199	-3.636	-101.834
FFF13	-82.767	-81.365	-13.115	-94.480	-0.454	-94.933
FFF14	-78.143	-75.572	-17.560	-93.132	-2.260	-95.391
FFF15	-82.364	-79.882	-16.842	-96.724	-3.464	-100.188
FFF21	-75.068	-72.719	-17.024	-89.743	-3.080	-92.823
FFF22	-73.928	-71.886	-16.108	-87.993	-1.006	-89.000
FFF23	-68.836	-65.471	-20.180	-85.651	-1.001	-86.652
FFF24	-70.030	-67.635	-16.381	-84.016	-3.073	-87.089
FFF25	-69.805	-66.513	-19.962	-86.475	-1.255	-87.729
FFF31	-65.016	-63.122	-15.486	-78.608	-0.295	-78.903
FFF32	-65.839	-63.922	-15.600	-79.522	-0.698	-80.220
FFF33	-68.166	-65.320	-18.935	-84.255	-1.568	-85.823
FFF34	-64.500	-62.254	-17.089	-79.343	-0.509	-79.852
FFF35	-62.413	-59.288	-19.647	-78.935	-2.337	-81.272
FFF41	-64.543	-61.452	-20.158	-81.610	0.275	-81.335
FFF42	-59.655	-57.130	-18.477	-75.607	1.677	-73.930
FFF43	-63.135	-60.135	-19.585	-79.719	-0.828	-80.548
FFF44	-60.407	-57.516	-19.377	-76.893	-0.827	-77.720
FFF45	-57.594	-53.991	-21.920	-75.911	0.458	-75.452

VITA

Tamaki Kurusu was born April 19, 1958, in Funabashi City, Chiba Prefecture, Japan to Tamao and Yoshie Kurusu. After graduating with B.A. in English from Kanagawa University in Yokohama City, Japan, she came to the United States to attend Virginia Polytechnic Institute and State University in Blacksburg, Virginia in 1985. She received her B.S. in Mathematics in May, 1988. Ms. Kurusu then entered the graduate program in the Department of Biochemistry at Virginia Polytechnic Institute and State University in 1990. Ms. Kurusu completed the requirements for her M.S. degree in August of 1996.

Tamaki Kurusu