Conformational Analysis of Protein Sidechains: Empirical Energy Parameters for Proline and Development of a Backbone-Dependent Rotamer Library

A thesis presented

by

Roland Leslie Dunbrack, Jr.

to

The Committee on Higher Degrees in Biophysics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biophysics

Harvard University

Cambridge, Massachusetts

May 1993

Conformational Analysis of Protein Sidechains: Empirical Energy Parameters for Proline and Development of a Backbone-Dependent Rotamer Library

Roland Leslie Dunbrack, Jr.
Research advisors: Martin Karplus and Jack L. Strominger
May 1993

Abstract

Empirical potential energy parameters for proline are developed for use with the molecular mechanics program CHARMM. *Ab initio* calculations on model compounds demonstrate dependence of the proline ring structure on the backbone dihedrals ϕ and ψ. This dependence is incorporated into a new potential energy parameterization for proline, which is able to reproduce the x-ray structures of several proline-containing peptides as well as the dynamics of four proline ring systems in antamanide.

Catalysis of the isomerization of the peptide bond preceding proline has been studied using *ab initio* calculations. NH_4^+ hydrogen bonded to the nitrogen of amides is able to reduce the calculated isomerization barrier isomerization from 13-18 kcal/mol to 0-9 kcal/mol depending on the orientation of the nitrogen substituents relative to the amide carbonyl. Arg66 of dihydrofolate reductase is known to catalyze the folding of DHFR by lowering the barrier to isomerization of a nearby proline. The calculations indicate that catalysis may proceed via hydrogen bonding to the ring nitrogen, and is of sufficient magnitude to play a significant role in the folding mechanisms of many proteins.

A backbone-dependent rotamer library is developed from protein crystal structures for use in predicting proteub sidechain conformation. The library in combination with energy minimization is successful in predicting 78% of χ_1 values (within 40°) of six proteins from their backbone coordinates alone.

A simple enumeration of steric interactions similar to the interactions of the terminal atoms of butane and pentane in the gauche and g+,g- conformations respectively

is used to analyze the experimental rotamer libraries. The enumeration is successful in explaining the general features of both the backbone-independent and dependent rotamer libraries. CHARMM is found to be successful in predicting alkane sidechain rotamer preferences, but is less successful in the cases of polar and aromatic sidechains.

The backbone rotamer library is used to aid homology modeling of six human Class I HLA proteins. The results are used to rationalize the previously determined sequence motifs of peptides known to bind to HLA-B35, HLA-B53, HLA-Cw4, HLA-Cw6, and HLA-Cw7.

Table of Contents

Acknowledgments

I would first like to thank my advisors, Prof. Martin Karplus and Prof. Jack Strominger. Prof. Karplus has been an inspiration to me since I walked into Chemistry 10 the first day of classes my freshman year at Harvard College. Over the course of one semester, he went from describing "the particle in a box" to molecular dynamics simulations of proteins and showed how the rules of physics could be used to explain everything from the structure of the hydrogen atom to oxygen binding in hemoglobin. During my junior and senior years under his supervision and encouragement, I worked on calculating the electronic structure of polyenes. That exposure to theoretical and computational chemistry led me to return Harvard and his research group to pursue a Ph.D. in biophysics. As my doctoral advisor, he has continued to inspire me to look at research problems broadly, deeply and carefully. He has always placed a strong emphasis on good writing, and this thesis is considerably clearer than it would have been without his careful reading.

Prof. Jack Strominger deserves a great deal of credit for putting up with a graduate student who couldn't perform an experiment to save his life. Jack has taught me a great deal about experimental biology and immunology, and I have tried to return the favor by explaining a little biophysics to him. Our conversations about science have meant a great deal to me. By providing me a desk and a bench in his lab even when I had long given up on experiments, he gave me access to post-doctoral fellows in his laboratory from whom I learned a great deal about the other side of the fence. These include Carolyn Doyle who supervised my rotation in the lab my first year of graduate school, and who encouraged me from the start. Also I would like to thank Jaekyoon Shin, Robert Ulrich, Roman Chicz and especially April Baker from the Strominger lab.

I must thank both of my advisors for their gentle encouragement at certain moments in my graduate career when my interests in science were flagging. I will always remember one winter when they seemed to have jointly decided to give me a kick

in the pants. They scheduled a meeting where I presented some meager results. Then they scheduled another meeting two weeks later where I presented a great deal more. Then they scheduled another meeting two weeks after that, where I presented even more. Then they tried to schedule another until I blurted out that I would get more done if they would leave me alone for awhile, the hint very much having been taken.

From my undergraduate years at Harvard, I must also express warm thanks to Prof. Veronica Vaida and Dr. Russ Hemley. Veronica was very kind to a sophomore who wanted to try out research for a summer. I will always remember her saying that with her I would not be asked to sort bolts. "People should sort their own bolts," she said. Even after I broke an expensive piece of equipment my first week in the lab, she said it happens to everyone and that I should go back to work. That one summer working in her lab was enough to convince me that I wanted to do a Ph. D. in science. I hope I can give the same opportunity some day to undergraduates considering a career in science in return for the kindness Veronica showed to me.

Russ Hemley was a post-doctoral fellow for both Prof. Karplus and Prof. Vaida. He was the first to suggest I try doing some calculations on the molecules I was studying in the lab. He introduced me to Martin, and after Veronica left Harvard to go to the University of Colorado during my junior year, I joined the Karplus group and continued to work with Russ. He was always very encouraging and just plain fun to work with.

After college, I spent two years at Cambridge University that changed my life in many ways beyond the merely scientific. I would like to thank Herchel Smith for creating the fellowship that sent me from Harvard to England for two years. I would also like to thank my advisor in the Department of Theoretical Chemistry, Prof. A. D. Buckingham, for encouraging me and showing great patience during what was a very difficult time in my life.

I would like to thank members of the Karplus group, past and present, who have made my graduate career much more productive and enjoyable than it otherwise would

have been, especially Bruce Tidor, Jeremy Smith, Martin Field, Stefan Fischer, Stephen Michnick, Hong Guo, Hsiang-Ai Yu, Louis Kuchnir, Krzysztof Kuczera, Andrej Š ali, Aaron Dinner, Roland Stote, Jeff Evanseck, Tom Ngo, Georgios Archontis, Herman van Vlijmen, Diane Joseph-McCarthy, Marci Karplus, and Alexander MacKerell Jr. I must especially acknowledge Alex's unique methods of inspiring me to complete the proline parameterization project. More seriously, I must thank him for many useful conversations and suggestions and for carefully reading Chapter 1 of this thesis in several draft stages. Diane performed the minimizations and simulations on the crystal compounds of Chapter 1 for earlier versions of the parameters and also proofread Chapter 1. Stefan and Stephen provided many insights on the calculations presented in Chapter 2. Andrej made many useful suggestions on Chapter 4. I would also like to thank Meredith Wynne, our simply amazing Biophysics administrator, who made negotiating the bureaucracy of completing a Ph. D. a breeze.

I must express my debt to my many friends at Mather House for the wonderful experience of having lived among them for the last four years. Jeff and Nancy Williamson hired me to be a resident tutor in biochemistry in the spring of 1989, and I am most thankful for the experiences that have come from that decision. I can not list all of the undergraduates I have gotten to know at Mather, but I will remember many of them always. I have been a member of a marvelous group of resident and non-resident tutors, whose dedication to Harvard students and to the House has been an inspiration and a blessing. In addition to my closest friends at Mather who will be mentioned later, I would like to thank Dr. Pat Hoy whose gentle and wise leadership as Mather's senior tutor will always serve as a model for me in my future endeavors. I have been the beneficiary of much good advice and counsel from him, and am honored to have worked with him these four years. And I certainly must mention Anne Aubrey, with whom I have spent many hours talking about nearly everything and everyone.

Finally, I would like to thank my family and friends to whom I am most indebted. My father provided me with a Harvard College education as well as instilled a love for math and science in me at a very young age. My mother has always enthusiastically cheered me on in my educational achievements, and I believe I would never have gone so far without her faith and support and interest in my education. To my parents and my sister Lynne, I express my love and gratitude always.

My closest friends over the last six years mean the world to me. I owe them much more than I can possibly express. They are Julia Dogra-Brazell, Lorna Brazell, Becky Gould, Larry O'Toole, Ian Tidswell, Alan Boudreau, John Patterson, Almaz Zelleke, Otto Coontz, Javier Torres, Jeff Dusenberry, and Brian Ballard. I dedicate this thesis to them and to my parents.

Introduction

The purpose of conformational analysis in organic and biological chemistry is to explain the physical and chemical properties of molecules in terms of their chemical structure, their conformation(s), and their dynamics. In a series of pioneering papers in the 1950's, D. H. R. Barton was able to correlate steric and electrostatic interactions of cyclohexane substituents with their chemical reactivity. Such interactions can be enumerated, assigned relative magnitudes from experimental data, and totaled to predict the likelihood that a molecule will exist in particular conformations. The physical and chemical properties of the molecule can then usually be inferred from the conformations the molecule is predicted to take on.

This is feasible in small systems for which the number of interactions is also small. The estimates of the interaction energies often do not have to be very accurate, since frequently one or two components of the energy dominate in determining the properties of interest. In much larger systems (e.g. proteins), it is not possible to do a simple back-of-the-envelope calculation to evaluate the energies of alternative conformations or their dynamical properties. To calculate these properties, empirical potential energy functions have been developed and implemented with complex computer programs.

This thesis presents several aspects of the conformational analysis of protein sidechains and its dependence on backbone conformation. It consists of two major components. The first is the development of a new potential energy parameterization of the amino acid proline for the molecular mechanics program CHARMM. Proline presents a major challenge because of the interdependence of the backbone conformation and the conformation and dynamics of the proline ring. The second major project of this thesis is an analysis of the dependence of protein sidechain conformation on the conformation of the protein backbone using data from the Brookhaven Protein Databank. While they were conceived of independently, they rest on the same principle that the

backbone and the sidechains can not be treated independently in any analysis of the structure and energetics of proteins.

In Chapter 1, an extensive series of *ab initio* calculations on N-acetyl proline amide and N-acetylproline N´-methyl amide is presented. These include a series of calculations in which the backbone dihedral ψ is fixed at different values, and the two minimum energy structures of the proline ring are found/optimized. The dependence of the structures and relative energies of these minimum energy conformations is striking, and provided an early impetus to examine the phenomenon of backbone/sidechain conformation interdependence more generally.

One element of proline structure and dynamics that is particularly important is the isomerization of the peptide bond preceding proline. The process of cis-trans isomerization is a rate-limiting step in the folding of many proteins. There are two families of enzymes that have been found to catalyze the cis-trans isomerization of proline residues as well as one example of a protein which catalyzes its own proline isomerization via an arginine sidechain. In Chapter 2, *ab initio* calculations on the interaction of an ammonium ion with the nitrogen atom of formamide and N,N-dimethyl acetamide are presented. This hydrogen bond is found to lower the barrier to cis-trans isomerization from 13-18 kcal/mol to 0-9 kcal/mol, depending on the basis set and the orientation of the nitrogen substituents in the transition state. A large number of proteins have been found in the Protein Databank that have hydrogen bonding groups interacting with the nitrogen of proline. These proteins are candidates for analysis of the dependence on the folding rate on the interaction of proline with a nearby positively charged sidechain.

In Chapter 3, the phenomenon of backbone/sidechain conformation dependence is extended to all of the amino acid sidechains in proteins. The context for this work was the development of a homology modeling scheme for the sidechains of HLA antigens (Chapter 5). For some sidechains there is no information in the known protein (the

template) for building the sidechain in the unknown structure (the target), and so a backbone-dependent rotamer library was developed from the Brookhaven Protein Databank. Backbone conformations are divided into 20° by 20° regions of the Ramachandran ϕ,ψ space. For each 20° by 20° block, the rotamer preferences for all sidechain types are given individually. Although the library was developed to be used in conjunction with information from a homologous protein, it is tested on the more difficult problem of predicting the conformations of all of the sidechains of six proteins from their backbone coordinates. It is found to be quite successful, and is applied to a homology modeling problem of penicillopepsin to rhizopuspepsin.

Since the backbone-dependent rotamer library showed fairly striking patterns of rotamer preferences as a function of the backbone dihedrals ϕ and ψ, I decided to investigate the reasons for this dependence. The key to solving the problem was found in a chemistry department lecture by Prof. Stuart Schreiber. He was discussing the backbone-independent rotamer preferences of sidechains and used simple Newman projections to show the steric interactions of heavy atoms separated by three and four chemical bonds (1-4 (gauche) and 1-5 (syn-pentane) steric interactions). These interactions have been used for some time in conformational analysis to predict the relative energies and populations of various conformations of organic molecules.

An analysis of the rotamer preferences of protein sidechains both in a backbone-independent and a backbone-dependent context is given in Chapter 4. This analysis consists of three parts – the enumeration of gauche and syn-pentane interactions in each likely rotamer conformation, the calculation of the energies and Boltzmann-weighted probabilities of these conformations with CHARMM, and the comparison of the enumeration and CHARMM results with the backbone-independent and backbone-dependent rotamer libraries developed in Chapter 3. The enumeration of steric interactions was quite successful in explaining the general features of the two rotamer libraries. The CHARMM calculations were quite successful in predicting the rotamer

preferences of alkane sidechains, but exhibited some limitations in predicting the conformations of polar and aromatic sidechains. These limitations are at least partly due to performing calculations without considering the effects of solvent and the rest of the protein, but may also be due to overestimation of electrostatic interactions between the sidechain and the backbone by the CHARMM potential. These effects will require further investigation.

In Chapter 5, a homology modeling method is described that combines information from the template protein and the backbone-dependent rotamer library in a consistent and easily implemented manner. The method is applied to the prediction of six HLA protein structures from the known coordinates of HLA-B27. The structure of one of these, HLA-Aw68, has been determined experimentally and is used to evaluate the new method. For the remaining five proteins, the sequences of peptides which bind to a cleft in the molecule are known, and the predicted structures are used to interpret the bound peptide sequence motifs. These include HLA-B53 which is known to provide resistance to malaria while the closely related allele HLA-B35 does not. We calculate the structures of both of these proteins and attempt to rationalize the peptides which bind to them. We have also calculated the structures of three proteins from the HLA-C gene locus. To date, there is no experimental structure of a protein from this locus, and there has been debate as to their function. It has been shown recently that these proteins do bind peptides for presentation to the immune system, and the sequence motifs have been determined. We use the predicted structures of the peptide binding sites to rationalize the peptide sequence motifs.

The first duty in life is to assume a pose. What the second is, no one has yet discovered.

– Oscar Wilde

Chapter 1

Derivation of Empirical Energy Parameters for Proline from *Ab initio* Calculations on N-acetylproline Amide and N-acetylproline N´-methylamide

Abstract

Molecular mechanics parameters for proline are developed from a series of *ab initio* calculations on the compounds N-formyl pyrrolidine, N-acetyl pyrrolidine, N-acetylproline amide, N-acetylproline N´-methyl amide, and proline amide. The proline backbone parameters were fit to the 6-31g*//3-21g energies of N-acetylproline N´-methylamide optimized at five values of the backbone dihedral angle ψ. Since the position of the succeeding peptide group is found to have an effect on the structures of the ring, the proline ring parameters are derived mainly from 6-31g* calculations on the Cγ-endo, Cγ-exo, and N-exo (or N-endo) conformations of N-acetylproline amide, with the backbone angle ψ fixed at 150° and -50°, values commonly found in proteins. The relative energies of the Cγ-endo and Cγ-exo minima vary with ψ, as does the energy of the saddle point between them. The saddle point structure is usually found to be N-exo with the ring in a non-planar conformation and the N bent down away from the proline carbonyl.

The parameters have been tested in energy minimizations and molecular dynamics simulations of five crystal structures of proline containing peptides as well as in a simulation of the cyclic decapeptide antamanide, which contains four prolines. The force field was found to reproduce well the structures and dynamics of these crystal structures and the solution dynamics of the four proline rings in antamanide. Adiabatic potential energy surfaces calculated with the new potential also correlate well with an extensive survey of proline rings contained in the Cambridge Crystallographic Database.

I. Introduction

Proline occupies a special place in both experimental and theoretical studies of protein structure and folding. The cis-trans isomerization of the peptide bond preceding a proline residue can affect both the rate and mechanism of protein folding. Many proteins fold with a rate-limiting barrier of approximately 20 kcal/mol that has been attributed to cis-trans isomerization of one or more proline residues [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. Two classes of enzymes – the cyclophilins and the FK506-binding proteins – have been found to catalyze cis-trans proline isomerization, although their role in cellular protein folding is not known [12, 13]. Since proline lacks the amino hydrogen, it has the ability to break [14] or kink helices [15], including transmembrane helices [16, 17]. In addition, prolines are commonly found in turns [18, 19, 20]. There are a number of proline-rich proteins, such as collagen, with important structural and mechanical properties that rely on the properties of prolyl residues [21]. These properties make proline unique among the amino acids and make it an important target for individual study in deriving molecular mechanics potentials for proteins.

The puckering of the proline ring, the cis-trans isomerism, and the planarity of the nitrogen center have been studied extensively with NMR experiments [22, 23, 24, 25, 26, 27] and x-ray crystallography [28, 29, 30, 31, 32, 33, 34, 35, 36, 37]. It is clear from crystallographic and NMR results that the conformation and dynamics of the ring depend on both the intra- and intermolecular environment of the proline residue. The proline ring has two clusters of conformations generally found in crystal surveys [28, 31]: Cγ-endo/Cβ-exo twist and Cγ-exo/Cβ-endo twist, where endo and exo indicate a position for the Cγ or Cβ atoms above or below the plane determined by Cδ, N, and Cα (above being the same side as the proline carbonyl; see Figures A1 and A3 in the Appendix) respectively [31]. Unlike the nitrogen center in pyrrolidine which is tetrahedral, the nitrogen center in proline crystal structures is almost, but not quite, planar [29].

In this paper, we present the derivation of an empirical potential energy function for proline that accurately reflects the energies and structures for the two ring pucker conformations, cis and trans peptide bond conformations, and the transitions between them. In addition to the intrinsic interest of the proline residue, this project was motivated by the implementation of an all atom potential function for proteins [38]. We began with the recently developed all-atom CHARMM parameters for alkanes [39], the peptide backbone [40], and amides [41]. To provide information which is not available experimentally, we have done extensive *ab initio* calculations on a number of conformations of N-formylpyrrolidine, N-acetylpyrrolidine, N-acetylproline amide, N-acetylproline N′-methylamide, and proline amide, representing internal, C–terminal and N-terminal prolines. After a brief description of the methods used, we present the results of the *ab initio* calculations, the fitting of these results to a potential energy function, and then a comparison of the *ab initio* and CHARMM potential results with x-ray crystal structure surveys, individual crystal structures, and NMR data on ring puckering and cis/trans isomerization. Several previous theoretical descriptions of proline exist [30, 42, 43, 44, 45, 46, 47, 48], and we compare these earlier results with those described here. Several empirical potential energy parameterizations for proteins have been developed, including parameters for proline, but they have not focused on unique properties of proline [49, 50, 51] and are not discussed here.

II. Methods

A. *Ab initio* calculations

Ab initio calculations were performed with the Gaussian 88 [52] and Gaussian 90 [53] programs at the 3-21g, 6-31g, and 6-31g* Hartree-Fock levels. The model compounds chosen are shown in Figure 1. Initial optimizations were performed at the 3-21g and 6-31g levels. Some structures were then fully optimized with the 6-31g* basis. The 6-31g* results were used as the basis for parameter development. In all of the *ab*

initio calculations reported, the structures were fully minimized with the Berny minimizer to the default tolerances specified in Gaussian 88 and Gaussian 90.

A normal mode analysis was performed on the unconstrained minimum energy structure of trans N-acetylproline amide ($AcProNH_2$) at the 6-31g* level. The potential energy distribution corresponding to the *ab initio* normal modes was calculated with the program MOLVIB [54], and used to adjust the CHARMM force constants. The normal mode coordinates have been expressed as a sum of bond stretching, angle deformation, rocking, scissor, wagging, twisting, and dihedral torsional components according to the recommendations of Pulay et al. [55]. Deformations of the five membered ring have been divided into two ring deformations (symmetric "ring def" and asymmetric "ring def'") and two ring torsions (symmetric "ring tor" and asymmetric "ring tor'") [55].

Positively charged proline amide was used to model N-terminal prolines. Both the Cγ-endo/Cβ-exo and N-endo structures were calculated at 6-31g*, as well as a complex of proline amide and a water molecule to determine appropriate charges for the ring nitrogen, the two hydrogen atoms bonded to nitrogen, Cα, and Cδ atoms. The complex was calculated by fixing proline amide in the configuration determined for proline amide alone (Cγ-endo/Cβ-exo) and configuring the water molecule as a CHARMM modified version [56] of the TIP3P water model [57], with the proline nitrogen, one water hydrogen, and the water oxygen fixed along a straight line. The minimization was performed as a function of three variables: the distance of the water hydrogen to the nitrogen, the angle of the hydrogen bond to the N-Cα bond, and the dihedral of the second water hydrogen with the water oxygen, ring nitrogen, and Cα.

B. CHARMM potential function fitting

The potential function used in version 22 of the program CHARMM [49] is of the form:

$$E = \sum_{\text{bonds}} \frac{1}{2} K_b(b - b_o)^2 \; +$$

$$\sum_{\text{bond angles}} \frac{1}{2} K_\theta(\theta - \theta_o)^2 \; + \; \sum_{\text{1-3 pairs}} \frac{1}{2} K_{ub}(s - s_o)^2$$

$$+ \sum_{\text{dihedral angles}} K_\phi [1 + \cos(n\phi - \delta)] + \sum_{\substack{\text{improper} \\ \text{dihedral angles}}} K_\omega(\omega - \omega_o)^2 \; +$$

$$\sum_{\text{non-bonded pairs}} \left\{ \varepsilon_{ij} \left[\left(\frac{Rmin_{ij}}{r} \right)^{12} - \left(\frac{Rmin_{ij}}{r} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon r} \right\}$$

No scaling of the non-bonded terms was performed for atoms separated by three bonds (1-4 terms). Non-bonded terms are included only for atoms separated by three or more bonds. No explicit hydrogen-bond terms were included. Calculations were performed with the dielectric constant equal to 1.0 (vacuum) and with no distance cutoff on electrostatic and Lennard-Jones terms. The parameters were fit to structures minimized with 1000 steps of the CHARMM conjugate gradient minimizer to a final root mean square gradient of less than 10^{-6}.

Initial parameters were derived from the alkane parameters of MacKerell et al. [39], protein backbone parameters of Kuchnir et al. [40], and amide parameters of MacKerell et al. [41]. Charges for N-acetylproline amide were set to agree with the CHARMM all-atom backbone charges as well as charges used in the alkane and amide parameterizations. The atom names and atom types used to describe the proline model compounds in the CHARMM potential are shown in Figure 1 and 2 respectively. For the ring carbon atoms, it was necessary to use three different atom types: CP1 for Cα,

CP2 for Cβ and Cγ, and CP3 for Cδ to allow for different bond lengths between the ring carbons. The other atom types for the model compounds correspond to standard CHARMM 22 atom types. Non-bonded parameters (ε_{ij} and $Rmin_{ij}$) were borrowed from alkane, amide, and backbone parameters used in the CHARMM all-atom parameter set [38]. For proline amide, the *ab initio* minimization of proline amide plus water was used to determine the appropriate charges for the positively charged N-terminus. The resulting interaction energy was used to determine a set of charges for atoms adjacent to and including the positively charged nitrogen. Although many combinations of charges satisfy the water/proline amide interaction, we have chosen a set that is reasonable compared to other CHARMM atom charges. The charges for the model compounds are listed in the Appendix.

The bond, bond angle, dihedral, and improper dihedral parameters of proline were adjusted to fit the 6-31g* *ab initio* structures of $AcProNH_2$ and N-acetylproline N´-methylamide ($AcPRONHCH_3$). Bond equilibrium distances (b_o's) were adjusted to reproduce the *ab initio* structures. These are less dependent on other parameters than are the angles and dihedrals. The heavy atom bond and angle force constants (K_b's and K_θ's) around the ring were then adjusted to give approximately the same normal mode frequencies for the minimized CHARMM structure as found in the *ab initio* calculation. The dihedrals angle parameters about the Cα-C bond (ψ) were adjusted to give a potential approximating the 6-31g*//3-21g energies of $AcProNHCH_3$ at ψ values of -150°, -75°, 0°, 75°, and 150°. The ring angle equilibrium values (θ_o's) and dihedral force constants (K_ϕ's) were then adjusted repeatedly until the Cγ-endo/Cβ-exo, Cγ-exo/Cβ-endo, and the endo/exo transition structures and relative energies of $AcProNH_2$ at 6-31g* with ψ fixed at 150° and -50° corresponded as closely as possible to the *ab initio* results. The peptide bond dihedral force constants were also adjusted to fit the experimental transition state energy of cis-trans isomerization of AcProOMe in low dielectric, non-hydrogen bonding solvents measured by Eberhardt et al. [58]. They found that in dioxane, benzene, and

toluene (ε=2.21, 2.27, and 2.38 respectively) $\Delta G^{\ddagger}$ of cis to trans isomerization was approximately 18 kcal/mol, while $\Delta G^{\ddagger}$ of trans to cis isomerization was approximately 19 kcal/mol. The free energy of activation was considerably higher in hydrogen bonding solvents. Also, the barrier was found to be almost entirely enthalpic. The height of the barrier in CHARMM was determined with the routine TRAVEL [59] by locating the saddle point on the cis-trans isomerization potential energy surface. Since $\Delta H^{\ddagger} \sim \Delta G^{\ddagger}$ in the experimental data of Eberhardt et al. [58], the peptide bond dihedrals parameters were adjusted until the cis-trans and trans-cis barriers agreed with their results.

Changes in the calculated normal mode frequencies and bond lengths were monitored and adjustments to the parameters made if necessary, as the parameters were adjusted to fit the *ab initio* geometries and relative energies. An iterative procedure of adjusting parameters to fit energies, geometries, and normal mode frequencies was used until convergence of the parameters was achieved.

C. Other calculations

In principle the conformation of the ring is determined by 9 degrees of freedom (5 ring atoms x 3 coordinates - 6), for instance 5 bond lengths and 4 ring atom dihedrals. In practice, the bond lengths are more or less constant in energetically accessible regions of the potential energy surface, so that only the 4 dihedrals are necessary to describe the ring pucker (the 5th is redundant since the 5 dihedrals must sum to 0) . Many authors choose to describe the ring with only two parameters – the pseudorotation angle and amplitude, which are complex functions of the ring dihedrals [60, 61]. We use the ring dihedrals themselves to describe the ring conformation. The dihedrals used to describe proline conformations in this paper are depicted for $AcProNH_2$ in Figure 2. In calculating a series of adiabatic potential energy surfaces with the new force field, dihedrals were constrained with force constants of 10000 kcal/mol at a series of values on a one- or two-dimensional grid (between -45° and 45° at 2.5° intervals for χ_1 and χ_2; between -180° and 180° at 5°

intervals for ψ; between -120° and -20° at 4° intervals for ϕ) and minimized for 1000 conjugate gradient steps with CHARMM [49].

Minimizations and molecular dynamics simulations of several crystals of proline containing peptides were performed with CHARMM. The x-ray structures used were acetyl-L-proline-N-methylamide (APNMA or AcProNHCH$_3$) [32], cyclo-L-prolyl-glycyl (CPG) [33], cyclo-L-prolyl-L-leucyl (CPL) [34], L-leucyl-L-prolylglycine (LPG) [35, 36], and L-proline (LPRO) [37]. The crystal of each of these compounds was minimized with 200 steps of ABNR minimization with the lattice held fixed and then 500 steps with the lattice parameters allowed to vary to a final gradient of less than 10^{-6}. The minimizations of each crystal structure were done at four image non-bonded term cutoffs (*cutim*) of 16, 18, 20, and 23 Å to ensure that the minimizations had converged. The other non-bond cutoffs, *cutnb, ctofnb,* and *ctonnb* were set to *cutim* minus 1, 2, and 4 Å respectively.

Molecular dynamics simulations of the crystals were performed for 100,000 steps (100 ps) of constant volume, constant temperature dynamics at the 23 Å cutoffs. Simulations were started from the minimizations after the 200 steps of lattice-fixed ABNR minimization. The lattice was held fixed throughout the simulations.

Four molecular dynamics simulations of the cyclic decapeptide antamanide were performed starting from the following conformations: the x-ray structure of antamanide [62] which has positive ϕ dihedrals for phenylalanines 5 and 10; a solution structure proposed by Kessler et al. [63] (ϕ_5 and ϕ_{10} both negative); and two other conformations proposed by Brunne et al. [64] from a GROMOS simulation with $\phi_5 > 0°$, $\phi_{10} < 0°$ in one case and $\phi_5 < 0°$, $\phi_{10} > 0°$ in the other. The simulations were performed for 100,000 steps (0.001 ps per step) of temperature equilibration, followed by 1,000,000 (1 ns) steps of dynamics. No cutoffs were used on non-bonded terms, and a dielectric constant of 1.0 was used. Structures were saved every 50 steps. Average residence times in the Cγ-endo and Cγ-exo conformations were calculated by analyzing the trajectories with

CHARMM: when $\chi_2 < 0$, the conformation was taken to be Cγ-endo; when $\chi_2 > 0$, the conformation was taken to be Cγ-exo. If the conformation of a proline ring changed from endo to exo and back to endo or vice versa for a very short period of time (less than 1 ps), the conformation was deemed not to have changed. This allows for brief ring flips that are not stable, and not likely to contribute to the experimental lifetimes. Once the residence times for the two conformations were calculated for each proline in each of the four trajectories, the exchange time constants (τ_{ex}) was calculated from the equation [25]:

$$\tau_{ex}^{-1} = \tau_{endo}^{-1} + \tau_{exo}^{-1}$$

III. Results

A. *Ab initio* calculations

A full list of the *ab initio* calculations performed and their relative energies is presented in Table 1, and a selected group of heavy atom dihedrals is listed in Table 2. The molecules with their CHARMM atom names are depicted in Figure 1.

Ab initio optimizations were performed on a series of small molecules of increasing size to find the most suitable model compounds for the derivation of parameters for proline. Pyrrolidine (cyclic $(CH_2)_4N$) is clearly insufficient for modeling proline since the nitrogen is sp^3 hybridized [65]. Optimization of N-formylpyrrolidine at the 3-21g level resulted in a Cγ-endo/Cβ-exo conformation (where *endo* is defined as positioned below and *exo* as above the plane of the ring determined by the other 4 ring atoms, where "above" is defined as the view in which the α,β,γ,δ carbons are in counterclockwise order and the α-carbon is *trans* to the aldehyde hydrogen). From crystal studies of proline containing compounds it is clear that there is another minimum with a Cγ-exo/Cβ-endo ring structure. The transition state between the Cγ-endo/Cβ-exo and Cγ-exo/Cβ-endo conformations of the pyrrolidine ring was located by fixing the Cα-Cβ-Cγ-Cδ dihedral (χ_2) to 0° and minimizing. The transition state was found to be N-exo , and 2.37 kcal/mol

above the Cγ-endo structure at the 3-21g level. The dihedral χ_1 (N-Cα-Cβ-Cγ) was found to be 18.8° in the transition state, and presumably a nearly equivalent N-endo transition state configuration with χ_1 equal to approximately -19° could also be found. To check that the N-exo structure was the saddle point on the pseudorotation potential energy surface, a second derivative matrix of the energy was calculated and diagonalized. The existence of a single negative eigenvalue for this matrix confirmed the location of the saddle point.

Rather than model the aldehyde with the CHARMM potential, we performed a series of *ab initio* optimizations on N-acetylpyrrolidine, which more closely resembles a peptide. The 3-21g minimized Cγ-endo/Cβ-exo structure was used as a starting structure for a 6-31g minimization, which in turn was used to initiate the 6-31g* minimization. The N-formylpyrrolidine N-exo ring conformation was used as the starting conformation for the ring pucker transition state structure of N-acetylpyrrolidine at 6-31g. The optimized structure was found to be 2.99 kcal/mol above the Cγ-endo/Cβ-exo minimum at 3-21g, 2.59 kcal/mol higher at 6-31g and 2.73 kcal/mol higher at 6-31g*, which are all slightly higher than the value for N-formylpyrrolidine (2.37 kcal/mol at 3-21g). An optimized structure with a "perpendicular" (or "perp" in Table 1) peptide bond was calculated to model the transition state between cis and trans proline by fixing the dihedral O_{i-1}-C_{i-1}-N_i-$C\alpha_i$ (where i denotes proline and i-1 denotes the previous amino acid) at -90°. (This is close to but not quite the top of the cis-trans barrier). This structure was 13.29 kcal/mol above the Cγ-endo/Cβ-exo minimum at 6-31g*. The experimental barrier between cis and trans proline is considerably higher at approximately 18-20 kcal/mol [1, 58].

It is interesting to follow the values of the improper dihedral about nitrogen (C_{i-1}-Cα-N-Cδ) in the N-acetylpyrrolidine structures: the Cγ-endo/Cβ-exo structures are nearly planar about the nitrogen atom, whereas in the ring puckering transition structure the peptide bond is bent upwards from the Cδ-N-Cα plane by 7° (N-improper=173°) at

6-31g and 18° (N-improper=162°) at 6-31g*. Apparently the addition of d-orbitals significantly effects the structure at the nitrogen. The 6-31g* perpendicular peptide bond structure has a nearly tetrahedral arrangement around the nitrogen with the improper equal to -125.3° (180° is a planar arrangement and ±120° is tetrahedral).

Preliminary CHARMM parameter optimization was performed using the N-acetylpyrrolidine *ab initio* calculations. However, upon optimizing the ring conformations and relative energies of the Cγ–endo/Cβ-exo and N-exo conformations of N-acetylproline amide ($AcProNH_2$) at 3-21g, we found that the presence of the proline ring carbonyl changes the relative energies of the ring conformations as well as their structures. Further *ab initio* minimizations were therefore performed on N-acetylproline amide.

A number of conformations of $AcProNH_2$ were optimized without restrictions on the value of the backbone dihedral ψ (N_i-$C\alpha_i$-C_i-N_{i+1}). In addition to the trans Cγ-endo/Cβ-exo conformation, the trans Cγ-exo/Cβ-endo, N-endo, and N-exo conformations as well as the cis Cγ-endo/Cβ-exo and perp Cγ-endo/Cβ-exo conformations were calculated at the 3-21g and 6-31g* levels. The minimized trans Cγ-endo/Cβ-exo, N-exo, and Cγ-exo/Cβ-endo conformations (structures *at1*, *at2*, and *at3* in Table 2; Figures A1, A2, and A3 in the Appendix respectively) with 6-31g* relative energies of 0.0, 2.70, and 1.90 kcal/mol were found to have a hydrogen bond between one amide NH_2 hydrogen and the carbonyl oxygen of the acetyl group. These structures are in the C_7 peptide conformation (so called because of the closing of the ring formed by 7 atoms – acetyl O_{i-1}, C_{i-1}, proline N_i, $C\alpha_i$, C_i, and amide N_{i+1}, H_{i+1}) with ψ having values of 76.5°, 84.0°, and 86.4° for the three structures respectively at 6-31g*. The trans N-*endo* calculation with χ_2 fixed at 0° was started with χ_1=-16°, χ_3=+16°, χ_4=-28°, χ_5=+28° (i.e. each ring dihedral opposite in sign of the N-exo structure), but the structure minimized back to the N-exo configuration with χ_1 =+16.4°. The reason seems to be that to maintain the C_7 hydrogen bond, the ring structure must be N-exo. The cis

Cγ-endo/Cβ-exo conformation was optimized and found to be 3.28 kcal/mol above trans Cγ-endo/Cβ-exo at 6-31g*. The cis structure (Figure A4 in the Appendix) had a value of ψ of -6.6°, placing an amide hydrogen pointing toward the ring nitrogen allowing for favorable electrostatic interactions with the π orbitals of the peptide bond and/or the partial lone pair of the nitrogen atom. Some of the energy difference is likely to be due to the presence of the C_7 hydrogen bond in the trans structure and its absence in the cis structure. Finally, the perpendicular peptide bond conformation (O_{i-1}-C_{i-1}-N_i-$C\alpha_i$ fixed at -90°) was calculated at 3-21g and 6-31g*, yielding energies 23.59 and 19.99 kcal/mol above the trans Cγ-endo/Cβ-exo structure respectively. The 6-31g* structure is shown in Figure A5 in the Appendix.

To investigate the potential energy surface of proline as a function of the backbone dihedral ψ, we optimized the proline dipeptide $AcProNHCH_3$ (trans and cis) with values of ψ at -150°, -75°, 0°, +75°, and 150°. The optimizations were performed at 3-21g, and the single-point energies of these structures at 6-31g* were then calculated. For the cis molecule the ψ=0° structure was optimized with the dihedral ψ unconstrained since the minimum energy structure has a value of ψ near 0°. The same procedure was followed for the ψ=75° structure for trans $AcProNHCH_3$, since the minimum energy structure has a value of ψ near 70°. The optimizations at other values of ψ were performed with the value of ψ held constant. Subsequently, optimizations were performed starting from the trans/ψ=150° structure and the cis/ψ=150° with ψ unconstrained. Starting from the trans structure, a local minimum was found with ψ=148° in trans with an energy 2.46 kcal/mol (6-31g*//3-21g) above the global minimum at ψ=70°. For the cis structure, a local minimum was found with ψ=174° and 2.69 kcal/mol at 6-31g*//3-21g above the cis minimum at ψ=-2°. The two local energy minima and the energies of the other fixed ψ conformations were used to parameterize the ψ potential energy surface of proline.

The orientation of the peptide group following proline (determined by the value of ψ) has an effect on the structure of the ring nitrogen in relation to its substituents. In Table 2, the structural parameters of these conformations are listed (structures *mt1* and *mc1* with ψ constrained to several different values). When ψ is near 180° (often referred to as trans´) in the ψ=150° or -150° structures, the improper dihedral on nitrogen (C_{i-1}-Cα-N-Cδ) is approximately 170-175°. But when ψ is in the range -75° to +75°, the orientation of the slight umbrella shape of the imino group is flipped to an improper dihedral between -173 and -164°. The effect was found in both cis and trans structures and has not been noted before to our knowledge. To check that this was not just an artifact of the Hartree-Fock calculations, we performed a database search on the structures of small peptides containing proline from the Cambridge Crystallographic Database (Figure 9), confirming the results of the *ab initio* calculations (described further below).

Since there was a striking effect of the position of ψ on the planarity of the imide group in the 3-21g proline dipeptide structures, we calculated the three ring conformations (Cγ-endo/Cβ-exo, N-exo, and Cγ-exo/Cβ-endo) of trans N-acetylproline amide with ψ constrained to 150° and -50° at 6-31g* to provide *ab initio* data for the parameterization of the proline ring. The values of ψ were chosen to correspond to the two conformations commonly found for proline residues in proteins. In trans $AcProNH_2$, the improper was found to be 171°, 158°, and 172° in Cγ-endo/Cβ-exo, N-exo, and Cγ-exo/Cβ-endo conformations when ψ was fixed at 150°. When ψ was fixed at -50°, the puckering transition state minimized to N-endo instead of N-exo. The N-impropers were -165, -151, and -167° for Cγ-endo/Cβ-exo, N-endo, and Cγ-exo/Cβ-endo respectively. The barrier between Cγ-endo/Cβ-exo and N-exo with ψ = 150° was found to be 0.5 kcal/mol less than the barrier in the unconstrained molecule (2.2 vs. 2.7 kcal/mol), and the difference in energy between the Cγ-endo/Cβ-exo and Cγ-exo/Cβ-endo structures was found to be 1.0 kcal/mol in the constrained ψ=150°

molecule versus 1.9 in the unconstrained molecule. With ψ fixed at -50° the Cγ-endo/Cβ-exo and Cγ-exo/Cβ-endo energies are reversed in order, with Cγ-exo/Cβ-endo being 0.3 kcal/mol *lower* in energy than Cγ-endo/Cβ-exo. The pseudorotation transition N-endo structure was 2.2 kcal/mol higher than the Cγ-endo structure. In all three ring conformations of $AcProNH_2$ with ψ unconstrained, the C_7 hydrogen bond is present, affecting the ring structures and energies. The CHARMM parameters were adjusted primarily to maximize their ability to match the minimized structures and relative energies of the structures with ψ fixed at 150° and -50°, since the C_7 peptide conformations for proline are quite uncommon in proteins, and therefore are of less relevance for a potential to be used for macromolecular systems.

In addition to the cis N-acetylproline amide structure with ψ unconstrained, a structure with ψ constrained at 150° was minimized. In both cases, the cis Cγ-endo/Cβ-exo structures were 3.3 kcal/mol higher in energy than the relevant trans Cγ-endo/Cβ-exo structures. The reasons for the energy differences in both cases are probably primarily electrostatic, arising from interactions of the C-terminal amide group with the carbonyl of the acetyl group and the proline ring nitrogen.

To model an N-terminal, positively charged proline, proline amide was minimized at 3-21g and 6-31g*. The minimum energy structure was found to be Cγ-endo/Cβ-exo. A transition structure between endo and exo was calculated by fixing χ_2 at 0° and was found to be N-endo (χ_1 = -20.6°) and 2.2 kcal/mol higher in energy than Cγ-endo/Cβ-exo. The N-exo structure was not calculated by *ab initio* methods for proline amide, but the CHARMM potential indicates that it exists and has approximately the same energy as the N-endo structure. Similarly, the Cγ-exo/Cβ-endo structure was not calculated for proline amide, but the CHARMM results show that its energy is very similar to the Cγ-endo/Cβ-exo energy. The Cγ-endo and the N-endo structures of $ProNH_2$ are shown in Figures A6 and A7 in the Appendix respectively.

Normal modes were calculated at the 6-31g* level on the Cγ-endo/Cβ-exo structure of trans N-acetylproline amide (ψ unconstrained) for the fitting of force constants. These are listed in Table 6 and described further below in comparison to the CHARMM normal modes of $AcProNH_2$.

B. CHARMM potential fitting

The procedure for fitting the CHARMM potential parameters to the experimental and *ab initio* data is described in detail in the Methods section above. In sum, empirical parameters for the CHARMM potential were adjusted to fit the structures and relative energies of the three ring conformations of trans $AcProNH_2$ with $\psi = 150°$ and the three ring conformations with $\psi = -50°$. The potential as a function of ψ was fitted to the 6-31g*//3-21g results on $AcProNHCH_3$. The normal modes of trans $AcProNH_2$ with ψ unconstrained were used to adjust the empirical force constants. When comparing crystal structures with the *ab initio* results, it was found that the 6-31g* results give a peptide bond that is slightly too long. Consequently, we shortened the bond length to correspond to experiment [32]. This bond length, the experimental cis/trans barrier, and information used to determine the alkane, amide, and peptide backbone parameters along with the *ab initio* results described above comprised all of the information used to determine the proline parameters.

The final parameters for proline and the model compounds are listed in the Appendix. The 6-31g* and CHARMM minimized energies and conformations of trans $AcProNH_2$ and cis $AcProNH_2$ are compared in Tables 3 and 4. Also the differences between the 6-31g* and CHARMM bond lengths, bond angles, and dihedral angles are listed in Table 5.

The relative energies of the three ring conformations with $\psi = 150°$ and $\psi = -50°$ are well-reproduced by the CHARMM potential. The difference between the Cγ-exo/Cβ-endo and Cγ-endo/Cβ-exo energies for the unconstrained ψ structures is not

well matched. The CHARMM potential underestimates the energy of the unconstrained ψ Cγ-exo structure by 0.8 kcal/mol. The CHARMM energies for the cis $AcProNH_2$ structures are also listed in Table 3. In the three conformations, the Cγ-exo/Cγ-endo differences are -0.07, 1.11, and 0.32 kcal/mol for the ψ=-50°, ψ unconstrained (~0°), and ψ=150° compared to the trans differences of -0.24, 1.09, and 0.92 kcal/mol respectively. The cis N-exo/Cγ-endo differences are 2.34, 3.11, and 2.76 kcal/mol (ψ=-50°,unconstrained,150°) and the trans differences are 2.31, 2.88, and 2.42 kcal/mol. We have not calculated the respective 6-31g* cis structures for comparison to these CHARMM results.

The trends in the N-improper values for the ψ = 150°, ψ = -50°, and the unconstrained ψ were reproduced reasonably well, except for the trans Cγ-exo/ψ=150° structure. For the ψ = 150° structures, the three N impropers are 171.3°, 157.9°, and 172.2° in 6-31g* and 173.3°, 163.8°, and -169.7° in CHARMM. The ψ = -50° structures, as noted earlier, have improper dihedrals of the opposite sign, i.e., -164.5°, -151.4, and -167.1°. These are also well matched by the CHARMM potential with values of -170.6°, -160.3, and -169.7°. The ψ free/Cγ-endo *ab initio* improper is -168.4°, while the CHARMM improper is flat at 180.0°. The CHARMM ψ free/N-exo structure has an improper of 165.9° compared to the 6-31g* value of 171.3°. The ψ free/ Cγ-exo structure calculated with CHARMM has an improper of -163.6°, while the 6-31g* value is -167.6°. In the two cis $AcProNH_2$ conformations (ψ free and ψ=150°) , the impropers were -178.1° and 174.9 in the CHARMM minimizations and -162.4° and 175.5° in the *ab initio* optimizations.

In Table 3, the energies of $AcProNHCH_3$ calculated at 6-31g* from the 3-21g optimized structures at various values of ψ are compared with the CHARMM results. To parameterize the potential as a function of ψ, it was necessary to add dihedral terms on the Cα-C bond to reproduce the *ab initio* energies. The resulting potential has two local minima in both trans and cis at values of ψ fairly close to the *ab initio* results. The

trans minima are at ψ = 70° and 149° in *ab initio* and 70° and 165° in CHARMM. The cis minima are at -2° and 174° in *ab initio* and -4° and 174° in CHARMM. The CHARMM relative energies of the two local minima for both cis and trans are in good agreement with the an initio results. The trans difference is 3.02 kcal/mol in CHARMM and 2.46 kcal/mol in *ab initio*; the cis difference is 2.81 kcal/mol in CHARMM and 2.69 kcal/mol in *ab initio*. The energies at other values of ψ are also well matched. At ψ=-75° the trans $AcProNHCH_3$ CHARMM and *ab initio* energies are 11.32 and 10.00 respectively, the cis/ψ=-75° energies are 5.13 and 6.70 kcal/mol, and the cis/ψ=+75° energies are 6.75 and 6.15 kcal/mol.

The energies of proline amide at four different ring conformations are also listed in Table 3. The 6-31g* transition state between Cγ-endo and Cγ-exo is an N-endo structure 2.2 kcal/mol higher than the Cγ-endo conformation. The CHARMM energy of this structure is 2.3 kcal/mol. We have used CHARMM to calculate the Cγ-exo and N-exo energies, which are 0.2 and 2.4 kcal/mol above the Cγ-endo conformation respectively. The values of ψ in the CHARMM minimizations are quite close to those in the *ab initio* optimizations of Cγ-endo and N-endo.

The calculated *ab initio* and CHARMM frequencies and potential energy distribution analysis of the normal modes of $AcProNH_2$ in the trans Cγ-endo/Cβ-exo (ψ unconstrained) conformation are listed in Table 6. The *ab initio* values have been scaled by a factor of 0.89, since 6-31g* *ab initio* frequencies are known to be too large [66]. The form of the normal modes and their frequencies correspond fairly well. The low frequency modes are particularly important for macromolecular modeling since these motions dominate large scale motions of proteins. While these frequencies depend on a large number of parameters, there is good agreement between the *ab initio* results and the CHARMM results, with most frequencies within 10% of the *ab initio* values and very similar potential energy distributions. In particular, the ring torsion modes (*ab initio* frequencies 99.3 and 187.2 cm^{-1} and CHARMM frequencies 110.5 and 200.7 cm^{-1}) are

important for the dynamics of the ring system, and are modeled accurately by the potential. Also, the peptide bond torsion and the ψ torsion at 45.9 and 88.6 cm^{-1} in *ab initio* are well reproduced by the CHARMM potential at 57.6 and 84.7 cm^{-1}.

C. Comparison with experiment

(i) *Proline crystal surveys and CHARMM potential energy surfaces*

Surveys of proline structures [28, 29, 30, 31] in peptide and protein crystals demonstrate a wide variety of conformations of the peptide backbone, the pyrrolidine ring, and positions of the ring nitrogen relative to its substituents. DeTar and Luthra [28] have completed the most thorough survey of proline crystals and their structural parameters, listing the bond lengths, bond angles, and dihedrals of 40 molecules containing proline.

Balaji et al. [29] have studied the structure of the nitrogen center in the ring in 107 compounds in the Cambridge Crystallographic Database, and found that the angle between the C_{i-1}-N bond and the Cα-N-Cδ plane (δ_1) varies between -13.4 and 18.2° for trans proline and between -6.0 and 12.8° for cis proline (their $\delta_1 > 0$ refers to the carbonyl carbon preceding proline bent upward toward the proline C=O group; $\delta_1=0$ would be a flat structure). Also, they found that the projection of the C-N bond onto the Cα-N-Cδ plane makes an angle (δ_2) with the Cα-N-Cδ angle bisector of between -0.6° and -8.1° in trans proline, and between 1.6° and 3.7° in cis proline (their $\delta_2 > 0$ refers to the C preceding proline bent towards Cδ; i.e., Cδ-N-C < Cα-N-C). Cung et al. [27] have noted that for ϕ in the range -95° to -75°, only Cγ-endo structures are found in their crystal structure survey. For ϕ above -55° only Cγ-exo (all cis) structures are observed. In the -75° to -55° range, there are examples of both structures.

We have performed a search of the Cambridge Crystallographic Database, and have found 161 prolines (84 trans and 77 cis) with resolution better than 0.08 Å. These structures can be described by a combination of the dihedrals ϕ, ψ, χ_1, χ_2, and the

improper dihedral about nitrogen (C_{i-1}-Cα-N-Cδ). We have compiled correlations among these dihedrals for the 161 prolines, and calculated adiabatic potential energy surfaces of $AcProNHCH_3$ for comparison with the survey data.

In Figure 3, the conformations of trans (filled circles) and cis (open circles) are shown as a function of their ϕ and χ_2 dihedrals. The survey shows that for all ϕ > -50° (up to -20° in the survey), all prolines in the survey are both cis and Cγ-exo ($\chi_2 > 0°$). In the range -50° > ϕ > -75°, there is a mixture of both cis and trans prolines with Cγ-endo ($\chi_2 < 0°$) and Cγ-exo ($\chi_2 > 0°$) rings . From ϕ = -75° to ϕ = -95°, there are both cis and trans prolines, but the rings are almost all Cγ-endo (with one cis structure with $\chi_2 = 0°$). Below -95°, the structures are all cis and Cγ-endo. These are essentially the same correlations found by Cung et al., who interpret these results to indicate that prolines with ϕ in the range -75° to -50° exist in a kinetic equilibrium between Cγ-endo and Cγ-exo, while in the other ranges, prolines are generally fixed in one ring conformation. In Figures 4a-d, we show potential energy curves for $AcProNHCH_3$ as a function of χ_2. Figures 4a and b are for trans $AcProNHCH_3$ and Figures 4c and d are cis. In Figures 4a and 4c the dihedral ψ is constrained at 150°, and in Figures 4b and 4d it is fixed at -50°. In each figure there are six curves, each with ϕ constrained to a different value, from -20° to -120°. From the curves in all the plots, it is clear that when ϕ is greater than -40°, there is a relatively deep Cγ-exo potential energy well and only a very shallow Cγ-endo well. The opposite effect is seen when ϕ is equal to or more negative than -80°, where there is a deep Cγ-endo well and a small or absent Cγ-exo well. In between (ϕ=-60°), the wells are each greater than a kcal/mol in depth. The potentials correspond well to the survey results, where ϕ < -75° produces only Cγ-endo structures, and ϕ > -50° produces only Cγ-exo structures. The physical explanation for the effect is simple: when ϕ is very negative, the proline carbonyl carbon is rotated *away* from the previous carbonyl group in the peptide. This rotates Cβ of the ring *down* below the plane of the ring. In this conformation, the Cγ is forced upwards above the ring plane to minimize ring strain and

fixed there, resulting in a Cγ-endo/Cβ-exo ring conformation. The opposite situation holds when $\phi > -50°$: with Cβ pulled upwards from the ring plane, Cγ is pushed downwards producing a Cγ-exo/Cβ-endo structure. In between ($-75° < \phi < -50°$), the Cβ is placed approximately level with the ring plane, and can be pushed slightly upwards or downwards to form either the Cγ-endo/Cβ-exo or Cγ-exo/Cβ-endo conformations. These results are confirmed in experimental [67, 68] and simulation results on the peptide antamanide which are described below. The question remains as to whether the value of ϕ is primarily determined by the conformation of the previous or subsequent amino acids to proline or both, or other intra- and intermolecular interactions. There is no clear explanation from the curves why cis prolines exhibit a much greater range in ϕ than trans prolines.

In Figures 5a (trans) and 5b (cis) we show the dependence of the nitrogen improper dihedral angle (N-impr) on ϕ for prolines in the CCDB. For both trans and cis, when ϕ is below -80°, the improper is nearly always above 180° (to have a continuous variable about the planar configuration with N-impr=180°, "negative" impropers will be referred to as impropers above 180°). Above $\phi = -80°$, there is a mix of both positive and negative impropers. The N-impr of the minimized structures represented in the energy curves in Figures 4a-d are shown in Figures 6a-d. It is clear from the plots that when ϕ is very negative (below $\phi = -80°$), the impropers are all above 180°. When ϕ is above -80°, the value of the N-impr ranges from 150° to 200°, depending on the value of ϕ and χ_2. When ϕ is -20°, -40°, and -60°, the nitrogen improper sometimes rises near $\chi_2 = 0°$ compared to the minimum energy structures. For the other curves in each figure, the motion of the nitrogen is inverted, moving from high impropers (>180°) at the Cγ-endo minimum ($\chi_2 \sim -35°$) to lower values near the transition state ($\chi_2 \sim 0°$) (occasionally below 180° entailing an inversion) back to higher ones at the Cγ-exo minimum ($\chi_2 \sim +35°$). The improper for Cγ-exo structures are almost always higher than for the corresponding Cγ-endo values. In some cases, when the Cγ-endo minimum has an improper below 180°,

the Cγ-exo improper may be above 180°. In the survey, this appears to be true for trans prolines (Figure 5a), but not for cis prolines (Figure 5b).

In Figure 7, the distribution of prolines as a function of ψ and χ_2 is shown. The figure shows that in the more heavily populated regions, (ψ < 0°; ψ > 100°), the value of ψ does not affect the structure of the ring. It is also evident that cis prolines exhibit values of ψ that are rarely seen in trans prolines (ψ= 20-100°). In this region there is a split in cis prolines between Cγ-endo structures when 20° < ψ < 65° and Cγ-exo structures when 65° < ψ < 100°. Figures 8a (trans) and 8b (cis) are the corresponding energy plots as a function of χ_2 with ψ constrained to different values for each curve. The dihedral ϕ is left unconstrained. The curves show that the value of ψ has only a small effect on the pseudorotation energy surface, with the Cγ-endo energy lower than Cγ-exo at ψ=150° and -150°, while at ψ=-30° and +30° the energies are approximately equal. The effect of ψ on cis prolines is not seen in the curves.

The relation between the nitrogen improper and ψ in the crystallographic survey is depicted in Figures 9a (trans) and 9b (cis). In the region where ψ < 0°, for both trans and cis prolines, the improper is usually above 180°. Above ψ=100°, trans Cγ-endo structures have impropers mostly below 180°, and trans Cγ-exo prolines impropers above 180°. Cis prolines exhibit impropers above 180° when ψ < 0°, but in the other regions there is less correlation between the improper and ψ. The cis prolines between ψ=65° and ψ=100° noticeably have impropers all near or above 180°. The Cγ-exo structures between ψ= 20° and 65° have mostly N-improper below 180°.

In Figures 10a (trans) and 10b (cis), the N-improper for the energy curves in Figures 8a and 8b are shown. In Figures 11a (trans) and 11b (cis), the values of ϕ for the same minimized structures are shown. The curves in Figure 10 have the same shape as those in Figures 6a-d, when ϕ < -60°. Again, Cγ-exo structures have larger values for the improper in both cis and trans. The Cγ-endo structures (χ_2~-35°) have impropers below 180°, and the Cγ-exo structures (χ_2~+35°) have impropers well above 180°. This

corresponds to the survey results for trans prolines (Figure 9a), but not for cis prolines (Figure 9b). For both cis and trans, the curves with ψ = -150° and 150° have the lowest values for the improper, corresponding to the *ab initio* results on AcProNHCH3 at 3-21g and AcProNH2 at 6-31g*, and the survey. The CHARMM results for the trans Cγ-exo impropers reproduce the experimental tendency toward impropers above 180°, while the *ab initio* improper is +172°.

The surveys suggest that the distribution of proline nitrogen improper dihedrals is correlated with the value of ψ. The cause of this effect is likely to be the influence of the proline carbonyl and succeeding amide dipoles on the position of the partial lone pair of the ring nitrogen. When ψ is near 180° the oxygen is in close proximity to the nitrogen, repelling the partial lone pair of the nitrogen to a position below the ring. This bends the substituents up toward the ring carbonyl. Conversely, positioning the amide hydrogen near the ring nitrogen may pull electrons from the π orbitals to a more sp^3-like arrangement with the lone pair above the ring, and so the nitrogen substituents point down. The precise mechanism may be more complex, but we feel this provides a reasonable if elementary explanation.

Figures 12a, 12b, 12c, and 12d (trans-endo, trans-exo, cis-endo, and cis-exo respectively, where "endo" is defined as $\chi_2 < 0$ and "exo" is defined as $\chi_2 > 0$) show the distribution of prolines on the ϕ-ψ map, divided on each figure into those with impropers below 180° ("up") and those with impropers above 180° ("down"). The ϕ-ψ surveys summarize the information described above – the strong effect of ϕ on χ_2 and the N-improper, especially in cis prolines where endo-down conformations are most likely when $\phi < -70°$ and exo-up conformations are most likely when $\phi > -70°$. For trans prolines, ϕ has a strong effect on χ_2, and ψ seems to skew Cγ-endo-prolines where $\psi >$ 120° toward an "up" conformation.

In Figures 13a and 13b, the energy as a function of ϕ and ψ is shown for trans and cis $AcProNHCH_3$ with χ_2 constrained to be -38° (a Cγ-endo conformation) with a weak

force constant of 100 kcal/mol. The constraint on χ_2 was used to prevent distortions on the surface caused by a ring flip from endo to exo. The electrostatic interactions between the amide carbonyl group and the acetyl carbonyl group in AcProNHCH$_3$ dominate the forms of each map, causing the global minimum for cis AcProNHCH$_3$ to be near ϕ,ψ = {-80°, 0°} and for trans proline near {-80°, 70°}. There are also local minima near ϕ,ψ = {-77°, 165°} in trans and ϕ,ψ = {-76°, 170°} in cis. In cis AcProNHCH$_3$, there is a tendency toward higher (less negative) value of ϕ when ψ is in the range -80° < ψ < -40°, and toward lower (more negative) value of ϕ when -30° < ψ < +50°. This effect is not seen in the peptide crystal survey where most cis prolines with 20° < ψ < 60° have values of ϕ above -60°.

Potential energy curves for both cis and trans AcProNHCH$_3$ as a function of ψ with ϕ unconstrained are shown in Figure 14a, and the values of ϕ and the N-improper for the same structures are shown in Figures 14b and 14c respectively. Trans has a deep minimum in the C_7 conformation at ψ=+70°, and a shallow minimum at ψ=165°. Trans has a large barrier to rotation about ψ (at ψ=-110°) of almost 12 kcal/mol above the minimum. Cis has a deep global minimum well approximately 7 kcal/mol deep at ψ = -5°, and a local minimum well 4-5 kcal/mol at ψ=170°.

The ϕ-curves in Figure 14b have similar shapes for cis and trans, except between ψ of 0° and 90° where the trans curve does not dip as far as the cis curve. The cis and trans N-improper curves are similar in shape, although there is less variation in cis than in trans. The N-improper is well above 180° in the central region and below 180° when ψ < -100° and ψ > 100°.

In Figure 15a-f, adiabatic energy surfaces for trans (Fig. 15a-c) and cis (Fig. 15d-f) AcProNHCH$_3$ as a function of χ_1 and χ_2 are shown with ψ unconstrained, ψ = 150°, and ψ = -50°. The lack of an N-endo (χ_1 < 0°) structure is apparent in the transition region between Cγ-endo and Cγ-exo (where χ_2 = 0°) in all of the plots except for trans/ψ=-50° where a fairly flat saddle point area allows for N-endo (χ_1<0°)

conformations as well as N-exo ($\chi_1 < 0°$) conformations of the ring. The cis/ψ=-50° plot also does not have a sharp rise for N-endo structures near the top of the barrier. In each case, this is due to electrostatic interactions between the amide and acetyl carbonyl groups, which are most pronounced in trans/ψ=unconstrained and trans/ψ=150°.

(ii) *Crystal structure minimizations and dynamics*

To judge the accuracy and transferability of the parameter set, crystal structures of several small peptides were studied. Energy minimizations and dynamics simulations of the structures in the crystal environment were performed with CHARMM. The calculated structures were then compared with the experimental crystal structures. The structures selected were acetyl-L-proline-N-methylamide (APNMA) [32], cyclo-L-prolyl-glycine (CPG) [33], cyclo-L-prolyl-L-leucine (CPL) [34], L-leucyl-L-prolylglycine (LPG) [35, 36], and L-proline (LPRO) [37]. These compounds were selected because they have exhibit a range of conformations of the ring (both Cγ-endo and Cγ-exo). The values of the χ_2 ring dihedrals are 36.0° (exo) for CPL, 35.4° (exo) for CPG, -36.2° (endo) for APNMA, and -41.0° (endo) for LPRO. LPG was found experimentally to exhibit both exo ("LPG2") and endo ("LPG") conformations of the ring. The molecules are depicted schematically in Figure 16, along with their atom names corresponding to the internal coordinate descriptions in Table 8, 9, and 10.

Minimizations for each of the crystals were performed with image non-bonded cutoffs of 16 Å, 18 Å, 20 Å, and 23 Å by minimizing for 200 steps with the lattice held fixed and 500 steps with the lattice dimensions allowed to change. Convergence was established at an image cutoff (*cutim* in CHARMM22) of 23 Å, which was also used for the simulations from the observation that the rms derivative of the potential energy function is close to zero at the end of each minimization, and the total, electrostatic, and van der Waals energies from the 16 Å, 18 Å, 20 Å, and 23 Å minimizations of each molecule are very similar. The energies and the unit cell dimensions and volumes for

minimizations of each of the crystals are listed in Table 7. An examination of the lattice parameters at the end of the 16 Å, 18 Å, 20 Å, and 23 Å minimizations of each molecule affirms the conclusion that the 23 Å cutoff is large enough. The differences in the final unit cell volume between the 16.0 and 23.0 Å minimizations are only 0.7 % for APNMA, 1.3% for CPG, 0.7 % for CPL, 0.6 % for LPG, 0.7 % for LPG2, and 1.4 % for LPRO. For APNMA and CPL, the unit cell volumes at the end of the 23.0 Å minimization are slightly lower than at the end of the 16.0 minimization, whereas for CPL, LPG, LPG2, and LPRO they are slightly higher. While the 23.0 Å cutoff appears to be large enough so as not to have a major effect on the results, the choice of the cutoff does have subtle effects on the minimized crystals. For each molecule, we will discuss in detail only the minimization done using the *cutim* value of 23.0 Å.

In addition to minimizations of the crystal structures, we performed molecular dynamics simulations of each of the crystals at constant volume and temperature (NVT). Simulations were run for 100,000 steps (0.001 ps per step, or 100 ps).

For each molecule the internal coordinate differences between the x-ray structure and the minimized and dynamics averaged structure were calculated, and the largest bond length, bond angle, and dihedral angle differences were examined. The internal coordinates for the CHARMM minimized structures, the x-ray structures, and the differences are listed in Table 8 according to their identity with reference to the proline residue in each peptide. Each proline atom is prefixed with "p", while each atom from the residues N-terminal and C-terminal to the proline residue is prefixed with "n" or "c" respectively (see Figure 16). Selected intermolecular distances between atoms are listed in Table 9, and the dynamics averaged internal coordinates, the differences in these coordinates from the x-ray structure, and fluctuations about the averages are listed in Table 10. We discuss each crystal structure individually below.

APNMA. The internal coordinate differences of APNMA in Tables 8 (minimized IC's – x-ray IC's) and 10 (simulation averaged IC's – x-ray IC's) show that the new parameter set describes the structure APNMA quite well. The ring in the APNMA x-ray structure is in a Cγ-endo/Cβ-exo conformation, and the dihedral ψ is -16°, indicating that the amide hydrogen is just above the ring quite near the ring nitrogen.

One of the largest heavy atom bond length difference is the peptide bond from proline to the methyl amide nitrogen of 0.03 Å in both the minimization and dynamics. The ring bond length CB-CG is too long by 0.03 Å. The ring CD-N bond length is too long by 0.04 Å in the simulation average. The largest heavy atom bond angle differences are the acetyl group angle {n-CA n-C n-O} of -3.2° and -3.8°, the backbone angles {p-CA p-C c-N} of 2.0° and 1.6° and {p-C c-N c-CA} of 1.2° and 1.5°, the ring angle {p-CG p-CD p-N} of 1.4° and 2.5°, and the angles around the ring nitrogen {n-C p-N p-CA} and {n-C p-N p-CD} of 1.5° and -2.0° in the minimized structure and 1.1° and -1.9° in the dynamics. The latter two mean that the C-N bond is angled away from the CA side of the ring by ~2° more than it should be. The heavy atom dihedrals show that the overall conformation of the minimized structure is correct with the largest backbone dihedral error of 7.4° and 4.7° in ψ. The errors in ϕ are -1.3° and 0.0°. The ring dihedrals are also quite close to the experimental structure with errors in the χ_1-χ_5 angles of -0.1°, 0.8°, -0.8°, 0.3°, and -0.4° in the minimized structure respectively. The dynamics averaged structure has larger ring dihedral differences, up to 3.9°. The nitrogen improper, which in the x-ray structure is -170.0°, deviates by -0.5° in the CHARMM minimization and -0.7° in the dynamics.

The fluctuations about the dynamics average show that the ring has approximately a ±10° motion as does the improper. Backbone Δϕ and Δψ are 7.5° and 10.0° respectively. The x-ray value of ψ is well within the fluctuation about the mean in the dynamics simulation. Similarly the heavy atom bond length fluctuations are all approximately 0.03 Å and the heavy atom angle fluctuations 3°, indicating that the both

the minimized structure and the x-ray structure are within the fluctuations about the dynamics averaged structure.

In Table 9, several intermolecular distances for APNMA are listed. The largest difference between the CHARMM minimizations and the x-ray structure is in an acetyl oxygen (n O)/amide hydrogen (c HN) distance of 0.13 Å. This can be attributed to the error in ψ which would move the amide relative to the acetyl group. The hydrogen bond between molecules in the crystal between the acetyl oxygen and the amide hydrogen is 2.05 Å in the x-ray structure and 1.92 Å in the minimized structure, indicating only a slight tightening of the intermolecular hydrogen bond. The remaining short intermolecular distances are all within 0.05 Å of the crystal structure.

CPG. Cyclic prolyl-glycine has two cis peptide bonds from glycine to proline and from proline to glycine, a Cγ-exo/Cβ-endo ring structure, and proline ψ = 38°. The bond length differences in CPG are all less than or equal to 0.015 Å in the minimization. Only the glycine C /proline N peptide bond is off by 0.04 Å in the simulation. The largest bond angle differences are the proline ring angles CB-CG-CD of 1.8° and 1.6° in the minimization and simulation respectively. The backbone dihedrals ϕ and ψ in glycine are in error by 2.3° and -5.1° in the minimized structure and 2.9° and -5.7° in the dynamics, while the proline backbone dihedrals are much closer to the x-ray structure at -2.5° and -1.3° in the minimization and -2.7° and -1.6° in the dynamics. The minimized and dynamics averaged rings are quite similar to the x-ray structure with the largest difference in χ_5 of 3.9° in the minimization and 4.7° in dynamics. The nitrogen improper which is -172.3° in the crystal structure is -168.0° in the minimization and -167.2° in the dynamics. The fluctuation about the improper is ±11.7°. The x-ray internal coordinates are well within the fluctuations indicating that the proline potential does a good job of representing the x-ray structure of CPG. Finally, the largest deviation in intermolecular interatomic distances occurs between proline oxygen of one molecule and proline nitrogen of another, which is 0.23 Å longer in the minimization than in the x-ray structure. The

intermolecular hydrogen bond between glycine nitrogen and oxygen atoms is only 0.03 Å shorter in the minimization than in the x-ray structure.

CPL. The x-ray structure of cyclic prolyl-leucine is quite similar to CPG described above. The CHARMM results are also similar to the results for CPG with the largest errors in the proline CB-CG bond length of 0.05 Å, the N-CA-C bond angle in leucine of 2.0/1.9° (minimization/dynamics), the proline ring angles CG-CD-N of 2.5/8.2° and CB-CG-CD of -3.9/-3.4°, and the glycine backbone dihedral angles ϕ and ψ of 4.0/4.1° and -8.6/-8.3° respectively. The leucine sidechain dihedrals χ_1 and χ_2 differ by 5.1/6.6° and -1.9/-1.1° respectively. The proline backbone dihedrals ϕ and ψ differ by -5.7/-5.3° and -0.5/-0.8°. The ring dihedrals for the minimized Cγ-exo/Cβ-endo ring are all less than 5.4° from the x-ray structure. In the dynamics simulation, however, there was significant fluctuation of the ring dihedrals ($\Delta\chi_2 = \pm 29°$), indicating that the ring flipped to the Cγ-endo structure at least once during the simulation. The predominant structure in the simulation was Cγ-endo, since the average χ_2 was -20.9°, in contrast to the x-ray structure with χ_2 of +36.0°. The only significant difference in interatomic distances was between the CB and oxygen of the leucine residue of -0.12 Å.

LPG. It is worth noting that the structure of LPG was originally solved in 1957 [35] based on visual estimates of the intensities of 1697 reflections and refined to a final R-factor of 12.9%. This structure was refined in 1979 by modern least-squares techniques to a final R-factor of 5.5% [36]. The earlier structure showed that Cγ of the proline ring was disordered, and located with approximately equal probability on either side of the plane of the ring. The refined structure showed that both Cβ and Cγ of the proline ring are disordered with two alternative conformations. The minor (41%) contributor (LPG2) shows Cγ as 0.70 Å out of the plane of the other four ring atoms and exo to the proline carbonyl, while the major (59%) contributor (LPG) shows Cγ 0.55 Å out of the plane of the ring and endo to the proline carbonyl. For the crystal minimizations, we used the highly refined coordinates [36] of both LPG (the

Cγ-endo/Cβ-exo ring) and LPG2 (the Cγ-exo/Cβ-endo ring). The hydrogen coordinates for Cβ, Cγ, and Cδ of proline for both LPG and LPG2 were generated using CHARMM. Also the hydrogen coordinates for the water molecule that are given in the paper were not used because they are incorrect (i.e., on a molecular graphics system, the water hydrogen atoms are not within bonding distance to the water oxygen) and instead were generated with CHARMM.

The only significant bond length difference was for the proline ring CA-CB bond of 0.04 and -0.09 Å for LPG and LPG2 respectively in both the minimization and the simulation. The largest angle differences were in the leucine backbone {n-N n-CA n-C} (4.3/4.1° and 4.4/4.2° in LPG and LPG2 respectively) and {n-CA n-C p-N} (3.5/3.9° and 4.4/4.0°), the proline backbone angles {p-CA p-C c-N} (3.7/3.6° and 4.4/3.7° in LPG and LPG2 respectively), {p-CB p-CA p-C} (-6.0/-5.4° and 5.3/7.4°), and the ring angles, all about 3-4° from the x-ray structure in the minimization, and somewhat further in the dynamics simulations. The minimized energy ring structure of LPG is closer to the x-ray structure than is LPG2, where some of the ring dihedrals are off by more than 9°. The minimized LPG ring dihedrals are all within 2° of the x-ray. However, the proline Cγ in the minimized LPG2 is still out of the plane of the ring and exo to the carbonyl as it is in the x-ray structure. The proline nitrogen impropers of both LPG and LPG2 (both 178.4°) have minimized to 179.9° and -170.2°. Since ψ is 161°, the CHARMM potential produces a flatter nitrogen center for the Cγ-endo conformation, but a non-planar structure with a negative improper for the Cγ-exo conformation. Finally, Table 9 reveals that the intermolecular distance between the N-terminal nitrogen of one peptide and the C-terminal oxygens of another is too short by 0.22/0.21 Å (LPG/LPG2) in one case and too close by 0.14/0.14 Å in another (different pairs of molecules) in the minimizations. The water molecule which is hydrogen bonded to the C-terminal oxygens has also moved substantially further away from the position of OT2 in the x-ray structure (by 0.30 Å in LPG).

In the simulation of LPG2, the proline ring has undergone significant fluctuation, which can be seen from the large values of the ring dihedral fluctuations about the simulation averages in Table 10. In the LPG simulation, the ring dihedral χ_2 never reached values above 0°, and in the LPG2 simulation the ring moved from exo to endo approximately 5 ps into the simulation, and remained there for the rest of the 100 ps simulation. This is in contrast to the experimental results, which indicate that the Cγ-exo conformation is populated approximately 40% of the time. The average value of ϕ (-72°) is in a region of the ϕ-χ_2 map that would indicate that equilibrium between the endo and exo conformations should be possible. It may be that the constraints of constant volume favor the endo conformation.

LPRO. L-proline is the zwitterion with a charged nitrogen in the proline ring. The minimization of this structure is a test of the N-terminal proline parameters.

The proline ring N-CA and N-CD bonds have the largest deviations from the x-ray structure in both the minimization and dynamics of LPRO at -0.03 Å and 0.02 Å respectively. The largest bond angle difference are the N-CA-C and CA-C-OT2 angles in the backbone of 4.0/3.3° and -2.9/-3.1° and the ring angle CA-CB-CG of 3.3/3.9°. The ring dihedrals are all quite close (<2.1° deviation in the minimization and <3.9° in the simulation) to the x-ray structure, and the only significant dihedral difference (-20.4° in the minimization and -6.4° in the simulation) is in the equivalent of the ψ dihedral N-CA-C-OT2, where OT2 is one of the carboxylate oxygens. The intermolecular distances between the terminal oxygen OT1 of one molecule and CA of another molecule is 0.20 Å smaller in the minimization than in the crystal structure because of the error in ψ. But the intermolecular salt-bridge length between the heavy atoms nitrogen and OT1 and OT2 are quite close to the x-ray structure distance (-0.02 Å and -0.06 Å deviations respectively).

The six crystal minimizations and dynamics simulations indicate that in general the new potential describes the backbone and ring structures of proline with reasonable accuracy. The last column of Table 8 and 10 gives the averages of the differences for each structural variable. In most cases the averages are close to zero, despite some large discrepancies in some of the structures. In particular, the ring bond angle and dihedral averages are all less than 2°. The proline peptide bond length was adjusted downwards by 0.03 Å from early versions of the parameter set to conform to the crystal minimization average. Otherwise the minimizations and dynamics simulations revealed no other significant differences that might indicate a need to adjust the parameters.

(iii) *Proline dynamics in antamanide*

To explore further the dynamics of the ring puckering, four molecular dynamics simulations of the cyclic decapeptide antamanide were performed starting from different backbone conformations. The primary structure of antamanide is

Pro3-Ala4-Phe5-Phe6-Pro7
| |
Pro2-Val1-Phe10-Phe9-Pro8

where the Pro2-Pro3 and Pro7-Pro8 links are cis peptide bonds. Madi et al. [67] and Schmidt et al. [68] have found by NMR spectroscopy that prolines 2 and 7 of antamanide are in equilibrium between the Cγ-endo and Cγ-exo forms of the rings while prolines 3 and 8 are more or less static in the Cγ-endo conformation. Proline 2 was found to be approximately 35% Cγ-endo and 65% Cγ-exo, while proline 7 was found to 45% Cγ-endo and 55% Cγ-exo. The backbone conformation in chloroform solution was found to be a mixture of conformations with either positive or negative values for the ϕ dihedrals

for phenylalanines 5 and 10, but was found to fluctuate on a very long time scale compared to the proline ring flips [63, 67].

The four simulations were started from the x-ray structure of antamanide [62] (Simulation A), which has positive ϕ dihedrals for phenylalanines 5 and 10, from a solution structure proposed by Kessler et al. [63] (ϕ_5 and ϕ_{10} both negative - Simulation B), and from two other conformations proposed by Brunne et al. [64] from GROMOS simulations with $\phi_5 > 0°$, $\phi_{10} < 0°$ (Simulation C) in one case, and $\phi_5 < 0°$, $\phi_{10} > 0°$ in the other (Simulation D). The dynamics averaged dihedrals, ϕ, ψ, and χ angles are listed in Table 11 for the four simulations, and the simulation averaged temperatures, energies, and time constants for the Cγ-endo and Cγ-exo conformations for each of the rings are listed in Table 12. Similar calculations have been performed and analyzed by Schmidt et al. [68] with an earlier version of the proline parameter set derived in this paper.

The average χ dihedrals and fluctuations for the four prolines in Table 11 indicate which residues are in dynamic equilibrium between the endo and exo forms. In addition, χ_2 for each proline in each simulation is plotted as a function of time in Figure 17. In each of the four simulations, prolines 2 and 7 fluctuate between the two conformations with fluctuations in the χ angles of 15-35°, while prolines 3 and 8 are locked in a Cγ-endo/Cβ-exo conformation with significant fluctuations in each of the χ angles (up to 15°), but not enough to flip the conformation to Cγ-exo for any significant period of time.

The predominant conformation for prolines 2 and 7 is determined by the local backbone conformation – the ϕ dihedral of phenylalanines 5 and 10. The percentages of Cγ-endo and Cγ-exo conformations for each proline in each simulation are listed in Table 13. The presence of a positive ϕ dihedral of the phenylalanine three residues C-terminal to each fluctuating proline (Phe 5 for Pro 2 and Phe 10 for Pro 7) results in a proline residue with higher Cγ-endo populations than Cγ-exo. The presence of a negative backbone dihedral in the phenylalanine skews the population toward Cγ-exo.

The exchange times from the simulations range from 4.2 to 10.6 ps for prolines 2 and 7. The variation in exchange times can be seen in the plots of χ_2 as a function of time for each of the four proline in the four simulations (simulations A, B, C, D in Figures 17a-d respectively). The experimental results indicate a much longer exchange time of approximately 30 ps for Pro 2 and 36 ps for Pro 7 [68], although with very large error margins. The simulations were not performed with an explicit solvent, and the effect of dielectric was not taken into account. Our results and those of Schmidt et al. performed with a developmental version of the present potential (which was somewhat different than the one described here) are significantly better than either simulations performed with the CHARMM version 19 potential [68] or with the GROMOS potential [64]. The CHARMM 19 potential resulted in proline residues which did not pucker back and forth between endo and exo conformations. The GROMOS simulation indicated a larger preference for Cγ-endo for prolines 2 and 7 than the CHARMM 22 simulation, and much shorter exchange times (approximately 1 ps) than the CHARMM 22 simulation (4-10 ps), which are already shorter than the experimental determination (30 ps). In sum, the CHARMM 22 potential is a reasonable compromise between the too stiff CHARMM 19 potential and the too flexible GROMOS potential. Addition of explicit solvent might be expected to bring the CHARMM 22 results closer to the experimental results by the addition of frictional forces due to the solvent.

(iv) *Comparison with NMR experiments*

The structure and dynamics of the proline ring in solution and in solids have been studied by NMR spectroscopy [22, 23, 24, 25, 26, 27]. Modified Karplus equations have been used to calculate prolyl ring geometries from vicinal coupling constants of methylene hydrogens. Bach et al. [22] have found that in $CDCl_3$, prolines in *cyclo*(D-Phe-Pro-Gly-D-Ala-Pro) and *cyclo*(Gly-Pro-Gly-D-Ala-Pro) can either rapidly interconvert between two ring structures (Pro 2 in each peptide) or maintain a relatively

static Cγ-endo conformation (Pro 5 in each peptide), depending on structural constraints on the proline backbone. In both peptides, the value of ϕ determines the ring dynamics: Pro 2 ϕ is -62° and -52° in the two peptides respectively and Pro 5 ϕ is -82° and -86°. These results conform to the limits on ϕ for interconverting and non-interconverting proline rings described by Cung et al.[27] and to the CHARMM potentials shown in Figure 4. Pro 2 in both peptides in the crystal form has a static ring conformation, apparently due to crystal packing forces [22].

Shekar et al. [24] used ^{13}C NMR to determine ring conformations of 26 proline ring systems in H_2O or D_2O. The dihedral angle distribution in solution appears to be similar to the crystal survey results, although the extremes for each dihedral are somewhat reduced in solution. Presumably, certain strained structures appear in crystals as a result of packing forces. They also used their analysis of NMR data to calculate lifetimes for the endo and exo structures of the ring in solution: lifetimes of each state vary from 0.01 ps to 100 ps among the 26 proline-containing molecules studied. Our simulation results on antamanide are well within this broad range. Sarkar et al. [26] used ^{2}H NMR of DL-[4,4-$^{2}H_2$]proline and DL-[4,4-$^{2}H_2$]proline hydrochloride in the crystalline form to measure apparent activation energies, ΔG, of approximately 1.3 and 0.6 kcal/mol to the pseudorotation in proline rings between the predominant endo and exo conformations. *Ab initio* methods matched by the CHARMM potential give an enthalpic activation barrier of 2.2 kcal/mol in going from endo to exo puckers in positively charged proline amide.

The energy difference between cis and trans prolyl residues has also been measured in NMR experiments. This difference is found to be quite small, but may vary significantly with the residues preceding and following proline. In the blocked Gly-Pro dipeptide studied by Stimson et al. [69], the cis minus trans energy difference is measured to be 1.5 kcal/mol in CD_2Cl_2 (dielectric constant = 8.93), and 0.9 kcal/mol in D_2O (dielectric constant = 80.0). Apparently, electrostatic interactions make the cis form less favorable, but these are shielded at high dielectric constant. Roques et al. [70] measured

the free energy and enthalpy differences between cis and trans of L-Pro-L-4-Hyp in D_2O and found a $\Delta G°$ of 0.54 kcal/mol and a $\Delta H°$ of 1.8 kcal/mol with a ΔS of 4 e. u. Higashijima et al. [71] measured ΔH and ΔS between cis and trans $AcProNHCH_3$ to be 2.9 kcal/mol and 15.2 e. u. in CCl_4 (ε=2), 2.0 kcal/mol and 11.5 e. u. in $CDCl_3$ (ε=5), 0.3 kcal/mol and 3.1 e. u. in acetone (ε=21) and -0.3 kcal/mol and 1.1 e. u. in D_2O (ε=78) with temperature dependent ^1H NMR. In the *ab initio* calculations on $AcProNH_2$ with ψ unconstrained, the energy difference was 3.3 kcal/mol and in the CHARMM calculation it was 0.84 kcal/mol at a dielectric constant of 1.0. When ψ was fixed at 150° (the more likely conformation at higher dielectric constants when the C_7 hydrogen bond is not usually formed [71]), the difference was the same in *ab initio* (3.3 kcal/mol), but 1.31 kcal/mol in the vacuum CHARMM calculation (ε=1) (Table 3).

We studied the behavior of the CHARMM potential at various dielectric constants by minimizing the proline dipeptide $AcProNHCH_3$ without constraints on ψ at several values for ε. The minimizations of both cis and trans $AcProNHCH_3$ were started from two different values of ψ – 150° and 0°. The results are listed in Table 14. The cis/trans energy difference, calculated from the lowest cis isomer energy and the lowest trans isomer energy (from the ψ^i=150° and 0° minimizations), falls from 0.6 kcal/mol at ε=1 to -0.5 kcal/mol at ε=2 (compared to ΔH=2.9 kcal/mol in CCl_4 [71]) back to 0.3 kcal/mol at ε=5 (compared to 2.0 kcal/mol in $CDCl_3$), and to 0.3 kcal/mol at ε=20 (compared to 0.3 in acetone). The CHARMM potential underestimates the cis/trans energy difference. The variation as a function of ε come primarily from removing the favorable interaction between the NH group of the C-terminal amide, and the acetyl carbonyl oxygen in the trans isomer. The experiments indicate an intrinsic difference in the cis and trans energies that is lower or nearly absent in water and other high dielectric solvents.

Relevant experimental information on the barrier between cis and trans proline comes from temperature dependent NMR studies of N,N dimethyl acetamide and some

proline containing peptides. Gerig [72] has measured an activation barrier of isomerization for N,N dimethyl acetamide of 19.7 $\pm$ 1.1 kcal/mol in water. Brandts et al. [1] have measured a cis/trans isomerization of the peptide bond in zwitterionic Ala-Pro and obtained a value of 19.8 kcal/mol, while Roques et al. [70] determined a cis/trans barrier in the zwitterionic L-Pro-L-4Hyp in water of 22.3 kcal/mol. In a blocked peptide, N-Acetyl-N´-ethylproline amide in DMSO, Stimson et al. [69] have measured a barrier of approximately 18 kcal/mol. In the folding transition of many proteins, there is an activation barrier of 18-20 kcal/mol, which is interpreted as the energy required to isomerize the peptide bond in X-Pro sequences [1]. Eberhardt et al. [58] have measured the barrier of AcProOMe in dioxane, benzene, and toluene (dielectric constants 2.21, 2.27, and 2.38 respectively) solvents and found values of approximately 18 kcal/mol. Since this value was determined at the low dielectric constants and in non-hydrogen bonding solvents, we used it to adjust the barrier of $AcProNHCH_3$.

Nagaraj et al. [73] have measured the barrier to rotation along the proline Cα-C bond (i.e., the dihedral ψ) in the compounds benzyloxycarbonyl-Pro-N-methylamide and pivaloyl-Pro-N-methylamide with ^1H and ^{13}C NMR spectra as a function of temperature. Peaks for Cγ and Cβ are split into two peaks at lower temperatures for trans prolines, because of the effect of the position on the succeeding peptide group on the resonance frequencies. By measuring the coalescence temperatures, the resonance frequency separation, and site populations of these compounds in the solvent $CDCl_3$, Nagaraj et al. estimated the barrier ($\Delta G^{\ddagger}$) to rotation about ψ between trans' (ψ=150°) and cis'(ψ=-50°) to be 14±1 kcal/mol. In Figure 14a, the energy of $AcProNHCH_3$ is plotted as a function of ψ. In vacuum, with a dielectric of 1.0, the CHARMM potential shows only two minima with a barrier to rotation of 12 kcal/mol. At 6-31g*//3-21g, the barrier for $AcProNHCH_3$ is at least 11.32 kcal/mol (Table 1).

(v) *Raman and IR studies of polyproline*

A number of IR and Raman experiments have been performed on polyproline I and II chains, both in the crystal and in solution [74, 75, 76, 77, 78]. A number of modes in the potential energy distribution analysis of N-acetylproline amide in Table 6 correspond well with the IR and Raman vibrations of polyproline which can be assigned to the ring or the proline backbone. In the all trans PLP II helix [75] which does not contain hydrogen bonds, several experimental frequencies are very close to the CHARMM/*ab initio* values. In Table 15 the Raman and IR frequencies and the assignments based on a normal mode calculation by Gupta et al. are listed, along with frequencies and assignments for the CHARMM normal mode calculation on $AcProNH_2$. In the spectra for PLP I (all cis) IR bands at 350, 1087, 1329, and 1350 cm^{-1} and Raman bands at 363, 1083, 1315 and 1337 cm^{-1} contain components of the C-N stretch (+ N-CD stretch) according to the assignment of Dwivedi and Gupta [74]. In our CHARMM calculation a small component of this stretch is found in the mode at 320 cm^{-1}. In the *ab initio*, there are components of C-N stretch at 341, 1354, 1367, and 1419. Gupta et al. [75] assigned a 100 cm^{-1} IR band of PLP II to the C-N torsion and N-C rock, which we find at 85 cm^{-1} for $AcProNH_2$. Their ring torsion mode at 195 cm^{-1} corresponds to the CHARMM calculated mode at 201 cm^{-1}. A number of other modes correspond rather closely between the experimental PLP II helix (all trans) and the calculated $AcProNH_2$ results. In addition, Deveney et al. [76] have found ring stretching modes at 901 and 985 cm^{-1} for L-proline and at 902 cm^{-1} for pyrrolidine, corresponding to our CHARMM calculated modes at 884 and 987 cm^{-1}.

D. Comparison with other theoretical studies

Other theoretical models of proline have been used to predict both the structures and relative energies of the two minima and transition state of the puckering potential energy surface. Sapse et al. [42] have used STO-3G and 6-31G basis sets to calculate the minimum energy conformations of N-acetylproline amide. Their results indicate a nearly

flat proline ring, which contradicts most experimental results. Most other theoretical models have extrapolated from small molecule experimental results including data on simple (non-cyclic) alkanes and amides to molecular mechanics potentials [28, 43, 46, 47, 51, 79]. For example, Venkatachalam et al. [46] have studied in detail the potential energy surface of the prolyl ring derived from small molecule work. They find a global minimum with χ_5 =-10° (Cδ-N-Cα-Cβ) by using a search on a two dimensional adiabatic surface (as a function of χ_5 and the position of Cγ). The remaining dihedrals χ_1- χ_4 are 27°, -35°, 29°, -11° respectively, which are not far from the trans Cγ-endo/Cβ-exo minimum found here for $AcProNH_2$ (31°, -36°, 27°, -8°, -14°). They find that the fully planar conformation has an energy 3.4 kcal/mol higher than the global minimum, but it is not clear from their calculations whether or not this configuration is necessarily the transition state between the endo and exo minima on their potential energy surface. DeTar and Luthra [28] use a more complicated force field to calculate the puckering of trans N-acetylproline methyl ester. They find two minimum structures with the first (χ_i= 28.7°, -35.0°, 27.7°, -10.3°, and -11.5°) having lower energy by 0.32 kcal/mol than the second (χ_i = -23.1°, 35.1°, -33.3°, 20.0°, and 1.9°), and a transition state between them with χ_2 = 0.0° and a relative energy from the global minimum of 2.7 kcal/mol. They do not state whether this transition state is fully planar or whether N is displaced up or down from the plane of Cαβγδ. Ramachandran et al. [48] find a Cγ-exo conformation 0.5 kcal/mol above the Cγ-endo conformation. For comparison, calculations on $AcProNH_2$ calculated with CHARMM when ψ is 150° (for comparison to the structure used by DeTar and Luthra), give an exo/endo energy difference of 0.9 kcal/mol and a barrier of 2.4 kcal/mol between Cγ-endo and Cγ-exo.

Vasquez et al. [43], using the ECEPP potential, have reported energies of trans $AcProNHCH_3$ as a function of ψ. They find three minima at ψ = 75°, -19°, and 160° with relative energies of 0.0, 0.95, and 1.25 kcal/mol with substantial (> 12 kcal/mol) barriers between 75° and -19° and between -19° and 160°, and a small barrier (0.5

kcal/mol) between the 160° and 75° conformations. The dihedral ϕ was kept fixed. The *ab initio* calculations and the CHARMM potential (Figure 14a) do not show the large barrier between ψ=75° and ψ=-19°. The difference in energy between ψ=75° and ψ=150° is 3.4 kcal/mol in CHARMM, which is substantially larger than the ECEPP difference. The difference between ψ=75° and ψ=-19° is approximately 3 kcal/mol in CHARMM, which is also substantially different than the ECEPP result. The CHARMM results would seem to reflect the importance of the C_7 intramolecular hydrogen bond that is dominant in non-polar solvents to the almost complete exclusion of other conformations [71]. The very large barrier in the ECEPP potential between ψ=75° and -19° as well as that calculated with other parameterizations [47] are probably due to fixing the geometry of the ring and/or part of the backbone.

IV. Discussion

Proline is unique among the amino acids in proteins in having a sidechain which connects to the protein backbone at two points, forming a five membered ring. This presents an unusual challenge in developing new empirical force field parameters, since the backbone and sidechain conformations are highly dependent on one another in a manner not seen in other amino acid sidechains.

In order to determine new parameters for proline, we have performed an extensive series of *ab initio* calculations on $AcProNH_2$ and $AcProNHCH_3$ in a variety of backbone and ring conformations. The conformations were chosen to aid in determining the potential energy surface as a function of the backbone dihedral ψ and the relative energies of the two minimum energy ring conformations at values of ψ commonly found in proteins. An *ab initio* normal mode calculation was used to fit bond, angle, and torsional force constants. Experimental information was used to fit the cis-trans isomerization barrier height and the peptide bond length.

In simulations of the cyclic decapeptide antamanide, the new potential reproduces the experimental results quite well. The NMR experiments of Madi et al. [67] have shown that prolines 2 and 7 are in dynamic equilibrium between the Cγ-endo and Cγ-exo conformations with slight preferences for Cγ-exo. Prolines 3 and 8, however, remain Cγ-endo. In molecular dynamics simulations starting from four different backbone conformations proposed for antamanide in solution, prolines 2 and 7 were in dynamic equilibrium while prolines 3 and 8 were essentially fixed in the Cγ-endo conformation. Our results indicate a faster pseudorotation than is found experimentally (10 ps vs. 30 ps correlation times). The experimental values, however, have large margins of error, and the simulations were performed without explicit solvent with a dielectric constant of 1.0. Despite the discrepancy in correlation times, the force field presented here reproduces the experimental results better than the CHARMM19 force field [68], better than an earlier developmental version of the parameters described here [68], and better than the GROMOS force field [64], which was found to produce much shorter correlation times.

Comparisons with experimental data, including crystal surveys, crystal minimizations, NMR, IR, and Raman experiments, indicate that the new potential reproduces experimentally observed data and promises to be useful in examining the behavior of proline residues in other peptides and proteins. In particular the crystal surveys and minimizations indicate that the potential accurately reflects the structures of proline in various environments.

Acknowledgments

I would like to thank Dr. Diane Joseph-McCarthy for setting up the x-ray minimizations and simulations and Dr. Alexander D. MacKerell, Jr. for helpful discussions.

References

1. J. F. Brandts, H. R. Halvorson, and M. Brennan. Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry*, **14**, 4953-4963 (1975).

2. T. E. Creighton. Possible implications of many proline residues for the kinetics of protein unfolding and refolding. *J. Mol. Biol.*, **125**, 401-406 (1978).

3. Y. Goto and K. Hamaguchi. Unfolding and refolding of the constant fragment of the immunoglobulin light chain. *J. Mol. Biol.*, **156**, 891-910 (1982).

4. L.-N. Lin and J. F. Brandts. Involvement of prolines-114 and -117 in the slow refolding phase of ribonuclease A as determined by isomer-specific proteolysis. *Biochemistry*, **23**, 5713-5723 (1984).

5. F. X. Schmid. Fast and slow folding forms of unfolded proteins. *Methods Enzymol.*, **131**, 70-82 (1986).

6. R. F. Kelley and F. M. Richards. Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry*, **26**, 6765-6774 (1987).

7. L. C. Wood, T. B. White, L. Ramdas, and B. T. Nall. Replacement of a conserved proline eliminates the absorbance-detected slow folding phase of iso-2-cytochrome c. *Biochemistry*, **27**, 8562-8568 (1988).

8. K. Lang and F. X. Schmid. Role of two proline-containing turns in the folding of porcine ribonuclease. *J. Mol. Biol.*, **212**, 185-196 (1990).

9. G. V. Semisotnov, V. N. Uversky, I. V. Sokolovsky, A. M. Gutin, O. I. Razgulyaev, and N. A. Rodionova. Two slow stages in refolding of bovine carbonic anhydrase B are due to proline isomerization. *J. Mol. Biol.*, **213**, 561-568 (1990).

10. T. Kiefhaber, H.-H. Kohler, and F. X. Schmid. Kinetic coupling between protein folding and prolyl isomerization I. *J. Mol. Biol.*, **224**, 217-229 (1992).

11. T. Kiefhaber and F. X. Schmid. Kinetic coupling between protein folding and prolyl isomerization II. *J. Mol. Biol.*, **224**, 231-240 (1992).

12. S. L. Schreiber and G. R. Crabtree. The mechanism of action of cyclosporin A and FK506. *Immunology Today*, **13**, 136-141 (1992).

13. S. L. Schreiber. Chemistry and biology of the immunophilins and their immunosuppressive ligands. *Science*, **251**, 283-287 (1992).

14. P. Y. Chou and G. D. Fasman. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.*, **47**, 45-148 (1978).

15. D. J. Barlow and J. M. Thornton. Helix geometry in proteins. *J. Mol. Biol.*, **201**, 601-619 (1988).

16. G. von Heijne. Proline kinks in transmembrane alpha-helices. *J. Mol. Biol.*, **218**, 499-503 (1991).

17. C. M. Deber, M. Glibowicka, and G. A. Woolley. Conformations of proline residues in membrane environments. *Biopolymers*, **29**, 149-157 (1990).

18. M. Levitt. Conformational preferences of amino acids in globular proteins. *Biochemistry*, **17**, 4277-4284 (1978).

19. C. Chothia and J. Janin. Relative orientaion of close-packed β-pleated sheets in proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **78**, 4146-4150 (1978).

20. G. D. Rose, L. M. Gierasch, and J. A. Smith. Turns in peptides and proteins. *Adv. Protein Chem.*, **37**, 1-109 (1985).

21. M. W. MacArthur and J. M. Thornton. Influence of proline residues on protein conformation. *J. Mol. Biol.*, **218**, 397-412 (1991).

22. A. C. Bach, A. A. Bothner-By, and L. M. Gierasch. Determination of proline ring nonplanarity from proton spin-spin constants: Application to two cyclic pentapeptides. *J. Am. Chem. Soc.*, **104**, 572-576 (1982).

23. S. C. Shekar and K. R. K. Easwaran. Proline ring conformations corresponding to a bistable jump model from ^{13}C spin-lattice relaxation times. *Biopolymers*, **21**, 1479-1487 (1982).

24. S. C. Shekar, M. B. Sankaram, and K. R. K. Easwaran. Pyrrolidine ring conformations in prolyl peptides from ^{13}C spin-lattice relaxation times. *Int. J. Peptide Protein Res.*, **23**, 166-173 (1984).

25. R. E. London. On the interpretation of ^{13}C spin-lattice relaxation resulting from ring puckering in proline. *J. Am. Chem. Soc.*, **100**, 2678-2685 (1978).

26. S. K. Sarkar, P. E. Young, and D. A. Torchia. Ring dynamics of DL-proline and DL-Proline hydrochloride in the solid state: A ^{2}H NMR study. *J. Am. Chem. Soc.*, **108**, 6459-6464 (1986).

27. M. T. Cung, B. Vitoux, and M. Marraud. Flexibility of Pro-Pro sequences: IR and NMR experiments. *New Jour. of Chem.*, **11**, 503-510 (1987).

28. D. F. DeTar and N. P. Luthra. Conformations of proline. *J. Am. Chem. Soc.*, **99**, 1232-1244 (1977).

29. V. N. Balaji, M. J. Rao, S. N. Rao, S. W. Dietrich, and V. Sasisekharan. Geometry of proline and hydroxyproline I. An analysis of X-ray crystal data. *Biochem. Biophys. Res. Comm.*, **140**, 895-900 (1986).

30. V. Madison. Flexibility of the pyrrolidine ring in proline peptides. *Biopolymers*, **16**, 2671-2692 (1977).

31. R. Balasubramanian, A. V. Lakshminarayanan, M. N. Sabesan, G. Tegoni, K. Venkatesan, and G. N. Ramachandran. Conformation of the proline rings as observerd in crystal structures of amino acids and peptides. *Int. J. Protein Research III*, 25-33 (1971).

32. T. Matsuzaki and Y. Iitaka. The crystal structure of acetyl-L-proline-N-methylamide. *Acta. Cryst.*, **B27**, 507-516 (1971).

33. R. B. van Dreele. The crystal structure of cyclo-L-prolyl-glycyl: A refinement of high-angle diffraction data. *Acta Cryst.*, **B31**, 966-970 (1975).

34. I. L. Karle. Crystal structure and conformation of the cyclic dipeptide cyclo-L-prolyl-leucyl. *J. Am. Chem. Soc.*, **94**, 81-84 (1975).

35. Y. C. Leung and R. E. Marsh. The crystal structure of L-leucyl-L-prolyl-glycine. *Acta. Cryst*, **11**, 17-31 (1958).

36. R. E. Marsh. Further refinement of the structure of L-leucyl-L-prolylglycine. *Acta Cryst.*, **B36**, 1265-1267 (1980).

37. R. L. Kayushina and B. K. Vainshtein. X-ray determination of the structure of L-proline. *Sov. Phys. - Crystallogr. (Eng. Trans.)*, **10**, 698 (1966).

38. A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, D. T. Nguyen, T. Ngo, B. Prodhom, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, J. Wiorkiewicz-Kuczera, and M. Karplus. Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. *To be submitted*.

39. A. D. MacKerell Jr., S. Fischer, and M. Karplus. Derivation of empirical potential energy parameters for alkanes. *To be submitted*.

40. L. Kuchnir. Ph. D. dissertation, Harvard University (1993).

41. A. D. MacKerell Jr. and M. Karplus. Derivation of empirical potential energy parameters for amides. *To be submitted*.

42. A. Sapse, L. Mallah-Levy, S. B. Daniels, and B. W. Erickson. The γ turn: *Ab initio* calculations on proline and N-acetylproline amide. *J. Am. Chem. Soc.*, **109**, 3526-3529 (1987).

43. M. Vasquez, G. Nemethy, and H. A. Scheraga. Computed conformational states of the 20 naturally occurring amino acid residues and of the prototype residue α-aminobutyric acid. *Macromolecules*, **16**, 1043-1049 (1983).

44. C. M. Venkatachalam, B. J. Price, and S. Krimm. A theoretical estimate of the energy barriers between stable conformations of the proline dimer. *Biopolymers*, **14**, 1121-1132 (1975).

45. T. Ooi, D. S. Clark, and W. L. Mattice. Conformational properties of poly(γ-hydroxy-L-proline) based on rigid and flexible pyrrolidine rings. *Macromolecules*, **7**, 337-343 (1974).

46. C. M. Venkatachalam, B. J. Price, and S. Krimm. Theoretical analysis of pyrrolidine ring puckering and the conformational energies of proline and 5-methylproline dimers. *Macromolecules*, **7**, 212-220 (1974).

47. W. L. Mattice, K. Nishikawa, and T. Ooi. Conformational properties of poly(L-proline) containing a flexible pyrrolidine ring. *Macromolecules*, **6**, 443-446 (1973).

48. G. N. Ramachandran, A. V. Lakshminarayanan, R. Balasubramanian, and G. Tegoni. Energy calculations of prolyl residue. *Biochim. Biophys. Acta*, **221**, 165-181 (1970).

49. B. R. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, **4**, 187-217 (1983).

50. S. Profeta and N. L. Allinger. Molecular mechanics calculations of the aliphatic amines. *J. Am. Chem. Soc.*, **107**, 1907-1918 (1985).

51. W. L. Jorgensen and J. Tirado-Rives. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, **110**, 1657-1666 (1988).

52. M. J. Frisch, M. Head-Gordon, H. B. Schlegel, K. Raghavachari, J. S. Binckley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J.

Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, E. M. Fluder, S. Topiol, and J. A. Pople. Gaussian 88 (1988).

53. M. J. Frisch, M. Head-Gordon, G. W. Trucks, J. B. Foresman, H. B. Schlegel, K. Raghavachari, M. Robb, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, S. Topiol, and J. A. Pople. Gaussian 90, Revision I (1990).

54. K. Kuczera and J. Wiorkiewicz. Molvib (1991).

55. P. Pulay, G. Fogarasi, F. Pang, and J. E. Boggs. Systematic *ab initio* gradient calculations of molecular geometries, force constants, and dipole moment derivatives. *J. Am. Chem. Soc.*, **101**, 2250-2260 (1979).

56. W. Reiher. Ph. D. Dissertation, Harvard University (1985).

57. W. L. Jorgensen, J. Chandrasekhar, J. P. Madura, R. W. Impey, and M. L. J. Klein. Comparison of simple potential functions for simulating liquid water. *Chemical Physics*, **79**, 926-935 (1983).

58. E. S. Eberhardt, S. N. Loh, A. P. Hinck, and R. T. Raines. Solvent effects on the energetics of prolyl peptide bond isomerizations. *J. Am. Chem. Soc.*, **114**, 5437-5439 (1992).

59. S. Fischer and M. Karplus. Conjugate peak refinement: An algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem. Phys. Letters*, **194,** 252-261 (1992).

60. C. Altona and M. Sundaralingam. Conformational analysis of the sugar ring in nucleosides and nucleotides: A new description using the concept of pseudorotation. *J. Am. Chem. Soc.*, **94**, 8205-8212 (1972).

61. D. O. Harris, G. G. Engerholm, C. A. Tolman, A. C. Luntz, R. A. Keller, H. Kim, and W. D. Gwinn. Ring puckering in five-membered rings. 1. General theory. *J. Chem. Phys.*, **50**, 2438-2445 (1969).

62. I. L. Karle, T. Wieland, D. Schermer, and H. C. J. Ottenheym. Conformation of uncomplexed natural antamanide crystallized from CH_3CN/H_2O. *Proc. Natl. Acad. Sci. USA*, **76**, 1532-1536 (1979).

63. H. Kessler, C. Griesinger, J. Lautz, A. Mueller, W. F. van Gunsteren, and H. J. C. Berendsen. COnformational dynamics detected by nuclear magnetic resonance NOE values and J coupling constants. *J. Am. Chem. Soc.*, **110**, 3393-3396 (1988).

64. R. M. Brunne and W. F. van Gunsteren. Molecular dynamics simulations of antamanide with the GROMOS potential. *To be submitted*.

65. G. Pfafferott, H. Oberhammer, J. E. Boggs, and W. Caminati. Geometric structure and pseudorotational potential of pyrrolidine: An ab initio and electron diffraction study. *J. Am. Chem. Soc.*, **107**, 2305-2309 (1985).

66. H. Guo and M. Karplus. Basis set and polarization function effects on optimized geometries and harmonic frequencies at the second-order Møller-Plesset level. *J. Chem. Phys.*, **91**, 1719-1733 (1989).

67. Z. L. Madi, C. Griesinger, and R. R. Ernst. Conformational dynamics of proline residues in antamanide. J coupling analysis of strongly coupled spin systems based on E. COSY spectra. *J. Am. Chem. Soc.*, **112**, 2908-2914 (1990).

68. J. M. Schmidt, R. Brüschweiler, R. R. Ernst, J. R. L. Dunbrack, D. Joseph-McCarthy, and M. Karplus. Molecular dynamics simulation of the proline conformational equilibrium and dynamics in antamanide using the CHARMM force field. *To be submitted*.

69. E. R. Stimson, S. S. Zimmerman, and H. A. Scheraga. Conformational studies of oligopeptides containing proline and glycine. *Macromolecules*, **10**, 1049-1060 (1977).

70. B. P. Roques, C. Garbay-Jaureguiberry, S. Combrisson, and R. Oberlin. Determination of the rates and barriers to conformational isomerization in the dipeptide L-pro-L-4-Hyp by direct ^{13}C NMR thermal equilibration. *Biopolymers*, **16**, 937-944 (1977).

71. T. Higashijima, M. Tasumi, and T. Miyazawa. ^{1}H NMR studies of N-acetyl-L-proline N-methylamide. Molecular conformations, hydrogen bondings, and thermodynamic quantities in various solvents. *Biopolymers*, **16**, 1259-1270 (1977).

72. J. T. Gerig. The effect of adjacent charges on the kinetics of rotation of the peptide bond. *Biopolymers*, **10**, 2435-2443 (1979).

73. R. Nagaraj, Y. V. Venkatachalapathi, and P. Balaram. Rotational isomerism about the Cα-CO bond in proline derivatives. *Int. J. Peptide Protein Res.*, **16**, 291-298 (1980).

74. A. M. Dwivedi and V. D. Gupta. Dispersion curves of poly-L-proline I chain. *Chem. Phys. Letters*, **16**, 109-113 (1972).

75. V. D. Gupta, R. D. Singh, and A. M. Dwivedi. Vibrational spectra and dispersion curves of poly-L-proline II chain. *Biopolymers*, **12**, 1377-1385 (1973).

76. M. J. Deveney, A. G. Walton, and J. L. Koenig. Raman spectra of imino acids and poly-L-hydroxyproline. *Biopolymers*, **10**, 615-630 (1971).

77. J. L. Koenig. Raman spectroscopy of biological molecules: A review. *J. Polymer Science (Part D)*, 59-177 (1972).

78. D. S. Caswell and T. G. Spiro. Proline signals in ultraviolet resonance Raman spectra of proteins: Cis-trans isomerism in polyproline and ribonuclease A. *J. Am. Chem. Soc.*, **109**, 2796-2800 (1987).

79. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, J. S. Profeta, and P. Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, **106**, 765-784 (1984).

Table 1. ***Ab initio*** **optimizations**

N-formylpyrrolidine					
Peptide	**Ring**[a]	**Constraints**	**3-21g**[b]		**6-31g***[b]
trans	Cγ-endo	none	0.00		
trans	N-exo	$\chi_2 = 0°$	2.37		
N-acetylpyrrolidine					
Peptide	**Ring**	**Constraints**	**3-21g**	**6-31g**	**6-31g***
trans	Cγ-endo	none	0.00	0.00	0.00
trans	N-exo	$\chi_2 = 0°$	2.99	2.59	2.73
perp	Cγ-endo	O_{-1}-C_{-1}-N-Cα=-90°			13.29
N-acetylalanine amide					
Peptide	**Ring**	**Constraints**	**3-21g**		**6-31g***
trans	Cγ-endo	none	0.00		0.00
cis	Cγ-endo	none	6.80		4.95
perp	Cγ-endo	O_{-1}-C_{-1}-N-Cα=-90°	21.80		18.16
N-acetylproline amide					
Peptide	**Ring**	**Constraints**	**3-21g**		**6-31g***
trans	Cγ-endo	none	0.00		0.00
trans	N-exo	$\chi_2 = 0°$	3.34		2.70
trans	Cγ-exo	none	2.22		1.90
cis	Cγ-endo	none	5.51		3.28
perp	Cγ-endo	O_{-1}-C_{-1}-N-Cα=-90°	23.58		19.99
trans	Cg-endo	y=150°			2.51
trans	N-exo	y=150°, c2 = 0°			4.72
trans	Cg-exo	y=150°			3.53
cis	Cg-endo	y=150°			5.88
trans	Cg-endo	y=-50°			6.29
trans	N-endo	y=-50°, c2 = 0°			8.49
trans	Cg-exo	y=-50°			5.97
N-acetylproline N'-methylamide					
Peptide	Ring	Constraints	3-21g		6-31g*//3-21g
trans	Cg-endo	y=-150°	12.28		9.28
trans	Cg-endo	y=-75°	15.97		11.32
trans	Cγ-endo	ψ= 0°	6.27		3.31
trans	Cγ-endo	none (ψ=70°)	0.00		0.00
trans	Cγ-endo	none (ψ=148°)	6.03		2.46
trans	Cγ-endo	ψ= 150°	6.07		2.56
cis	Cγ-endo	ψ=-150°	10.22		7.87

cis	Cγ-endo	ψ=-75°	10.76	7.88
cis	Cγ-endo	none (ψ=-2°)	5.03	2.75
cis	Cγ-endo	ψ= 75°	11.37	9.00
cis	Cγ-endo	ψ= 150°	8.54	5.44
cis	Cγ-endo	none (ψ= 174°)	7.74	5.49
Proline amide$^+$				
	Ring	**Constraints**	**3-21g**	**6-31g***
	Cγ-endo	none	0.00	0.00
	N-exo	$\chi_2 = 0°$		2.16

[a] Cγ-endo indicates that $\chi_2<0°$. Cγ-exo indicates that $\chi_2>0°$. N-endo occurs when $\chi_2=0°$ and $\chi_1<0°$. N-endo occurs when $\chi_2=0°$ and $\chi_1>0°$.

[b] All energies in kcal/mol relative to energy of conformation with energy of 0.00 kcal/mol for each molecule and basis set.

Table 2. *Ab initio* structures: dihedrals

Mol[a]	Fix ψ	Basis Set	χ1	χ2	χ3	χ4	χ5	ω	ϕ	ψ	Impr[b]
ft1		3-21g	30.09	-39.25	32.61	-14.41	-9.92				-179.56
ft2		3-21g	18.81	0.00	-18.86	33.13	-33.10				-177.26
nt1		3-21g	31.81	-39.17							
nt1		6-31g	29.25	-36.72	29.50	-11.61	-11.17	-180.00			-179.81
nt1		6-31g*	29.96	-38.00	31.05	-12.78	-10.87	-179.02			-179.75
nt2		3-21g	17.95	0.00	-17.99	31.43	-31.47	178.66			174.39
nt2		6-31g	15.11	0.00	-15.13	26.45	-26.46	178.34			173.41
nt2		6-31g*	17.93	0.00	-17.94	31.31	-31.32	174.14			162.04
np1		6-31g*	-20.70	-6.88	32.12	-46.54	42.05	88.70			-125.27
at1		3-21g	34.62	-39.54	28.65	-7.08	-17.49	-178.84	-81.33	70.24	-177.79
at1		6-31g*	31.90	-37.82	28.75	-9.02	-14.44	-172.94	-85.69	76.45	-168.41
at1	150	6-31g*	31.58	-35.76	25.95	-6.15	-16.11	176.39	-69.13	150.00	171.30
at1	-50	6-31g*	19.56	-33.59	34.46	-23.53	2.56	-171.11	-69.58	-50.00	-164.52
at2		3-21g	16.41	0.00	-16.41	28.44	-28.56	-176.13	-86.06	68.38	177.01
at2		6-31g*	17.54	0.00	-17.58	30.57	-30.56	-179.47	-83.88	84.04	173.55
at2	150	6-31g*	16.27	0.00	-16.13	28.54	-28.52	172.22	-67.81	150.00	157.88
at4	-50	6-21g*	-17.71	0.00	17.76	-30.75	30.63	-167.47	-54.89	-50.00	-151.44
at3		3-21g	-9.72	29.20	-37.38	33.12	-14.82	-174.86	-83.28	71.18	-172.30
at3		6-31g*	-14.39	31.91	-37.16	29.76	-9.72	-175.12	-81.36	86.37	-167.60
at3	150	6-31g*	-16.47	32.85	-36.42	27.85	-7.26	174.94	-61.57	150.00	172.21
at3	-50	6-31g*	-31.59	38.40	-30.17	10.63	13.25	-173.11	-55.98	-50.00	-167.07
ac1		3-21g	34.03	-39.55	29.59	-8.38	-16.25	1.37	-84.58	-2.63	-174.03
ac1		6-31g*	31.32	-38.01	29.84	-10.50	-13.14	10.56	-90.24	-6.62	-162.39
ac1	150	6-31g*	32.49	-36.58	26.43	-6.01	-16.75	-1.94	-73.10	150.00	174.90
ap1		3-21g	26.67	-40.98	39.69	-23.69	-1.93	89.01	-111.83	102.31	-132.41
ap1		6-31g*	19.99	-37.42	41.26	-29.79	6.01	88.34	-105.18	119.38	-129.77
mt1	-150	3-21g	41.23	-37.13	18.52	8.42	-31.61	172.18	-86.54	-150.00	173.99
mt1	-75	3-21g	22.12	-32.20	29.47	-16.62	-3.42	-172.93	-71.90	-75.00	-166.98
mt1	0	3-21g	34.61	-40.56	30.79	-9.29	-16.06	-172.30	-94.31	0.00	-164.01
mt1		3-21g	34.46	-39.89	29.37	-7.94	-16.84	-173.33	-85.29	69.97	-173.17
mt1		3-21g	34.13	-39.04	28.47	-6.99	-17.36	176.55	-72.14	147.97	171.28
mt1	150	3-21g	34.40	-39.09	28.47	-6.94	-17.45	176.52	-72.30	150.00	170.86
mc1	-150	3-21g	39.38	-40.16	25.04	0.09	-24.96	-4.34	-78.92	-150.00	175.16
mc1	-75	3-21g	25.05	-29.92	22.77	-7.45	-11.10	2.29	-77.61	-75.00	-169.64
mc1		3-21g	33.18	-40.10	31.35	-10.83	-14.15	10.90	-90.85	-1.63	-165.55
mc1	75	3-21g	38.24	-31.92	13.46	11.32	-31.48	5.09	-110.12	75.00	-164.57
mc1	150	3-21g	34.71	-39.65	29.07	-7.32	-17.38	-4.12	-74.96	150.00	173.39
mc1		3-21g	34.31	-40.36	30.65	-9.29	-15.85	-2.97	-70.27	174.12	171.30
p1		3-21g	35.26	-42.24	31.53	-9.79	-15.76			63.78	-120.28
p1		6-31g*	24.19	-39.23	38.60	-24.03	-0.13			172.65	-124.92
p4		6-31g*	-20.58	0.00	20.54	-34.73	34.58			163.84	-124.11

[a]Key: 1st letter = molecule (f=N-formylpyrrolidine; n=N-acetylpyrrolidine; a=N-acetylproline amide; m=N-acetylproline-N′-methylamide); 2nd letter = cis/trans state (t=trans; c=cis; p=perp); number following = conformation of ring (1=Cγ-endo; 2=N-exo; 3=Cγ-exo, 4=N-endo)

[b]Impr=improper dihedral about proline nitrogen=CY-CA-N-CD (see Figure 1 for atom definitions)

Table 3. CHARMM vs *ab initio* energies of $AcProNH_2$ and $ProNH_2$

	ψ[b]		ΔE rel. to Cγ-endo min. of each ω,ψ group		ΔE rel. to trans Cγ-endo with ψ unconstrained	
Mol.[a]	**6-31g***	**CHARMM**	**6-31g***	**CHARMM**	**6-31g***	**CHARMM**
ac1	-50	-50		0.00		4.41
ac2	-50	-50		2.34		6.75
ac3	-50	-50		-0.07		4.35
ac1	(-7)	(-5)		0.00	3.28	0.84
ac2		(-5)		3.11		3.95
ac3		(-12)		1.11		1.95
ac1	150	150		0.00	5.88	4.38
ac2		150		2.76		7.15
ac3		150		0.32		4.71
at1	-50	-50	0.00	0.00	6.29	7.53
at2	-50	-50	2.20	2.31	8.49	9.83
at3	-50	-50	-0.32	-0.24	5.97	7.29
at1	(76)	(74)	0.00	0.00	0.00	0.00
at2	(84)	(81)	2.70	2.88	2.70	2.88
at3	(86)	(86)	1.90	1.09	1.90	1.09
at1	150	150	0.00	0.00	2.51	3.07
at2	150	150	2.21	2.42	4.72	5.48
at3	150	150	1.02	0.92	3.53	3.98
p1	(173)	(174)	0.00	0.00		
p2		(178)		2.40		
p3		(172)		0.23		
p4	(164)	(169)	2.16	2.28		
mt1	-150	-150	9.28	7.92		
mt1	-75	-75	11.32	10.00		
mt1	0	0	3.31	1.83		
mt1	(70)	(70)	0.00	0.00		
mt1	(148)	(165)	2.46	3.02		
mt1	150	150	2.56	3.26		
mc1	-150	-150	5.12	5.72		
mc1	-75	-75	5.13	6.70		
mc1	(-2)	(-4)	0.00	0.00		
mc1	75	75	6.25	6.15		
mc1	150	150	2.69	3.90		
mc1	(174)	(174)	2.74	2.81		

[a] Key: see Table 2

[b] Values not in parentheses are constrained values of ψ. Values in parentheses are from unconstrained local minima.

Table 4. Comparison of CHARMM and *ab initio* structures of $AcProNH_2$

Bond lengths

Mol[a]	Fix ψ	Basis	N-CA	CA-CB	CB-CG	CG-CD	CD-N	N-CY	CAY-CY	CY-OY
at1		6-31g*	1.4687	1.5280	1.5296	1.5299	1.4653	1.3505	1.5128	1.2092
		CHM	1.4601	1.5402	1.5327	1.5284	1.4651	1.3217	1.4872	1.2276
at1	150	6-31g*	1.4461	1.5383	1.5317	1.5322	1.4597	1.3536	1.5115	1.2049
	150	CHM	1.4545	1.5380	1.5336	1.5317	1.4648	1.3201	1.4873	1.2267
at1	-50	6-31g*	1.4628	1.5422	1.5313	1.5266	1.4604	1.3623	1.5139	1.1997
	-50	CHM	1.4596	1.5418	1.5345	1.5300	1.4641	1.3232	1.4876	1.2237
at2		6-31g*	1.4604	1.5346	1.5454	1.5366	1.4579	1.3503	1.5132	1.2090
		CHM	1.4541	1.5453	1.5432	1.5316	1.4589	1.3229	1.4872	1.2278
at2	150	6-31g*	1.4448	1.5447	1.5455	1.5364	1.4569	1.3592	1.5113	1.2038
	150	CHM	1.4520	1.5441	1.5437	1.5322	1.4583	1.3219	1.4874	1.2268
at4	-50	6-21g*	1.4612	1.5418	1.5445	1.5343	1.4593	1.3677	1.5135	1.1993
	-50	CHM	1.4588	1.5430	1.5396	1.5333	1.4618	1.3271	1.4881	1.2237
at3		6-31g*	1.4803	1.5447	1.5460	1.5385	1.4724	1.3472	1.5149	1.2317
		CHM	1.4642	1.5436	1.5290	1.5265	1.4660	1.3250	1.4875	1.2274
at3	150	6-31g*	1.4695	1.5389	1.5317	1.5256	1.4604	1.3506	1.5139	1.2089
	150	CHM	1.4577	1.5427	1.5318	1.5271	1.4621	1.3218	1.4874	1.2267
at3	-50	6-31g*	1.4484	1.5483	1.5326	1.5264	1.4537	1.3537	1.5117	1.2047
	-50	CHM	1.4615	1.5381	1.5306	1.5312	1.4670	1.3258	1.4879	1.2237
ac1		6-31g*	1.4572	1.5387	1.5301	1.5285	1.4665	1.3637	1.5144	1.2006
		CHM	1.4615	1.5389	1.5327	1.5284	1.4634	1.3224	1.4877	1.2248
ac1	150	6-31g*	1.4447	1.5417	1.5308	1.5291	1.4622	1.3580	1.5149	1.2010
	150	CHM	1.4597	1.5390	1.5325	1.5288	1.4618	1.3216	1.4883	1.2261
ap1		6-31g*	1.4771	1.5432	1.5275	1.5200	1.4703	1.4199	1.5073	1.1902
		CHM	1.4619	1.5301	1.5271	1.5336	1.4719	1.3170	1.4849	1.2254

[a] Key: see Table 2

Mol[a]	Fix ψ	Basis	CA-C	C-O	C-NT
at1		6-31g*	1.5365	1.2019	1.3462
		CHM	1.5297	1.2317	1.3559
at1	150	6-31g*	1.5288	1.1976	1.3602
	150	CHM	1.5283	1.2319	1.3584
at1	-50	6-31g*	1.5259	1.1939	1.3686
	-50	CHM	1.5207	1.2297	1.3566
at2		6-31g*	1.5342	1.2015	1.3495
		CHM	1.5302	1.2315	1.3564
at2	150	6-31g*	1.5277	1.1985	1.3587
	150	CHM	1.5291	1.2318	1.3583
at4	-50	6-21g*	1.5248	1.1940	1.3663
	-50	CHM	1.5202	1.2297	1.3565
at3		6-31g*	1.5373	1.2208	1.3439
		CHM	1.5312	1.2314	1.3557
at3	150	6-31g*	1.5383	1.2006	1.3496
	150	CHM	1.5254	1.2310	1.3581
at3	-50	6-31g*	1.5274	1.1964	1.3605
	-50	CHM	1.5194	1.2296	1.3568
ac1		6-31g*	1.5306	1.2002	1.3465
		CHM	1.5213	1.2293	1.3568
ac1	150	6-31g*	1.5309	1.1969	1.3544
	150	CHM	1.5269	1.2298	1.3589
ap1		6-31g*	1.5292	1.1969	1.3563
		CHM	1.5249	1.2300	1.3590

[a] Key: see Table 2

Bond angles

Mol[a]	Fix ψ	Basis	N-CA-CB	CA-CB-CG	CB-CG-CD	CG-CD-N	CD-N-CA	CD-N-CY	CA-N-CY	N-CY-CAY
at1		6-31g*	103.06	103.38	103.43	103.67	112.35	125.41	121.10	117.13
		CHM	102.35	104.29	102.86	104.71	112.73	123.94	123.34	119.20
at1	150	6-31g*	102.88	103.35	103.76	103.63	113.53	127.11	118.68	117.24
	150	CHM	102.54	104.06	103.21	104.70	112.74	124.54	122.34	119.35
at1	-50	6-31g*	103.91	104.71	103.25	102.88	112.48	125.77	119.66	117.10
	-50	CHM	104.44	104.97	103.26	104.55	112.21	123.82	123.24	119.14
at2		6-31g*	103.97	106.20	106.02	104.05	110.49	127.69	121.46	117.41
		CHM	103.51	107.02	105.68	104.83	111.60	124.03	122.70	119.64
at2	150	6-31g*	104.06	105.99	106.28	103.96	111.74	125.84	118.03	117.21
	150	CHM	103.78	107.00	105.80	104.89	111.98	124.05	121.79	119.58
at4	-50	6-21g*	104.11	106.01	105.99	104.49	110.00	123.38	119.48	117.18
	-50	CHM	106.02	107.09	106.35	106.84	112.24	121.91	122.72	119.07
at3		6-31g*	104.66	105.33	103.20	102.10	110.46	127.92	121.09	117.42
		CHM	103.25	104.11	102.01	103.89	111.94	122.87	123.00	119.22
at3	150	6-31g*	104.43	104.99	102.97	102.59	111.01	126.92	120.72	117.83
	150	CHM	103.54	104.37	102.45	103.54	112.51	124.29	122.32	119.29
at3	-50	6-31g*	103.95	104.39	103.13	102.04	112.75	128.06	118.64	117.45
	-50	CHM	104.18	103.47	103.07	104.63	111.53	123.04	123.14	119.24
ac1		6-31g*	103.11	103.15	103.25	103.55	112.66	118.87	126.16	117.67
		CHM	102.49	104.29	102.69	104.57	112.72	120.62	126.63	120.08
ac1	150	6-31g*	102.69	103.17	103.55	103.74	113.33	119.73	126.75	117.52
	150	CHM	102.76	104.06	102.86	104.63	112.62	120.35	126.87	120.18
ap1		6-31g*	105.15	103.61	101.64	103.24	108.86	114.83	114.18	114.46
		CHM	102.64	102.12	102.76	106.23	108.76	116.97	118.88	116.79

[a] Key: see Table 2

Mol[a]	Fix ψ	Basis	N-CY-OY	OY-CY-CAY	CB-CA-C	N-CA-C	CA-C-NT	O-C-NT	C-N-CAT
at1		6-31g*	121.95	120.91	112.28	110.89	114.61	122.10	123.28
		CHM	121.71	119.09	111.84	112.32	115.79	121.54	122.66
at1	150	6-31g*	120.94	121.82	110.53	111.08	114.92	121.98	122.80
	150	CHM	121.36	119.29	111.55	111.02	115.91	121.52	122.58
at1	-50	6-31g*	121.14	121.75	112.92	114.02	115.91	120.93	122.50
	-50	CHM	121.96	118.91	113.47	113.30	116.29	120.43	123.04
at2		6-31g*	121.70	120.88	112.47	109.16	114.49	122.08	123.44
		CHM	121.38	118.98	111.83	111.59	115.79	121.47	122.73
at2	150	6-31g*	120.99	121.79	110.30	110.91	115.19	121.68	122.86
	150	CHM	121.23	119.18	111.03	111.08	115.94	121.42	122.64
at4	-50	6-21g*	121.29	121.50	112.14	114.91	115.78	120.66	122.84
	-50	CHM	122.22	118.71	112.18	113.97	115.76	120.64	123.33
at3		6-31g*	121.49	121.09	111.05	109.75	113.54	122.39	124.07
		CHM	122.01	118.77	111.07	114.55	116.21	121.26	122.53
at3	150	6-31g*	121.71	120.45	112.16	110.42	114.65	122.12	123.23
	150	CHM	121.48	119.23	110.52	112.01	116.15	120.95	122.88
at3	-50	6-31g*	120.69	121.85	109.67	111.13	114.95	121.60	123.03
	-50	CHM	122.07	118.69	112.61	115.00	115.94	120.55	123.24
ac1		6-31g*	120.88	121.45	110.55	115.19	117.21	119.54	123.24
		CHM	120.83	119.09	110.62	112.62	116.93	119.97	123.10
ac1	150	6-31g*	121.14	121.34	110.13	112.89	114.75	122.61	122.63
	150	CHM	121.09	118.73	111.19	112.12	115.87	121.62	122.47
ap1		6-31g*	122.89	122.63	110.48	110.13	115.51	122.19	122.30
		CHM	122.78	120.40	111.88	114.99	116.25	121.04	122.71

[a] Key: see Table 2

Dihedral angles

Mol[a]	Fix ψ	Basis	χ1	χ2	χ3	χ4	χ5	ω	ϕ	ψ	Impr[b]
at1		6-31g*	31.90	-37.82	28.75	-9.02	-14.44	-172.94	-85.69	76.45	-168.41
		CHM	31.92	-36.12	26.18	-6.46	-15.94	178.49	-75.86	74.17	-179.99
at1	150	6-31g*	31.58	-35.76	25.95	-6.15	-16.11	176.39	-69.13	150.00	171.30
	150	CHM	32.70	-35.18	24.00	-3.56	-18.30	171.94	-72.33	150.03	173.28
at1	-50	6-31g*	19.56	-33.59	34.46	-23.53	2.56	-171.11	-69.58	-50.00	-164.52
	-50	CHM	22.02	-32.01	29.87	-17.07	-3.11	-175.71	-68.58	-49.89	-170.56
at2		6-31g*	17.54	0.00	-17.58	30.57	-30.56	-179.47	-83.88	84.04	173.55
		CHM	15.64	-0.03	-15.63	27.58	-27.20	169.97	-72.56	81.41	165.78
at2	150	6-31g*	16.27	0.00	-16.13	28.54	-28.52	172.22	-67.81	150.00	157.88
	150	CHM	14.75	-0.02	-14.73	26.03	-25.72	165.91	-70.18	150.02	163.84
at4	-50	6-21g*	-17.71	0.00	17.76	-30.75	30.63	-167.47	-54.89	-50.00	-151.44
	-50	CHM	-7.05	-0.01	7.08	-12.25	12.14	-173.28	-63.68	-49.90	-160.30
at3		6-31g*	-14.39	31.91	-37.16	29.76	-9.72	-175.12	-81.36	86.37	-167.60
		CHM	-27.55	38.17	-34.12	17.87	6.03	177.26	-69.49	77.08	-163.59
at3	150	6-31g*	-16.47	32.85	-36.42	27.85	-7.26	174.94	-61.57	150.00	172.21
	150	CHM	-24.69	36.19	-33.69	19.39	3.31	174.24	-67.93	150.01	-169.67
at3	-50	6-31g*	-31.59	38.40	-30.17	10.63	13.25	-173.11	-55.98	-50.00	-167.07
	-50	CHM	-29.66	36.30	-29.38	11.23	11.60	-177.07	-61.44	-49.89	-163.22
ac1		6-31g*	31.32	-38.01	29.84	-10.50	-13.14	10.56	-90.24	-6.62	-162.39
		CHM	31.16	-36.56	27.65	-8.58	-14.14	-6.68	-77.12	-4.98	-178.13
ac1	150	6-31g*	32.49	-36.58	26.43	-6.01	-16.75	-1.94	-73.10	150.00	174.90
	150	CHM	31.11	-36.41	27.61	-8.50	-14.19	-1.39	-70.12	150.07	175.46
ap1		6-31g*	19.99	-37.42	41.26	-29.79	6.01	88.34	-105.18	119.38	-129.77
		CHM	-41.30	36.98	-19.49	-6.58	29.90	88.09	-71.16	97.96	-137.22

[a] Key: see Table 2

[b]Impr=improper dihedral about proline nitrogen=CY-CA-N-CD (see Figure 1 for atom definitions)

Table 5. Differences between CHARMM and 6-31g* structures of $AcProNH_2$

Bond lengths

Mol[a]	Fix ψ	N-CA	CA-CB	CB-CG	CG-CD	CD-N	N-CY	CAY-CY	CY-OY
at1		0.0087	-0.0122	-0.0030	0.0014	0.0002	0.0288	0.0256	-0.0184
at1	-50	-0.0084	0.0003	-0.0019	0.0005	-0.0051	0.0335	0.0242	-0.0218
at1	150	0.0032	0.0004	-0.0032	-0.0034	-0.0037	0.0390	0.0263	-0.0240
at2		0.0063	-0.0107	0.0023	0.0050	-0.0011	0.0274	0.0261	-0.0188
at2	-50	-0.0072	0.0006	0.0017	0.0042	-0.0014	0.0373	0.0239	-0.0230
at2	150	0.0023	-0.0012	0.0049	0.0010	-0.0025	0.0406	0.0255	-0.0244
at3		0.0053	-0.0047	0.0027	-0.0008	-0.0056	0.0256	0.0265	-0.0185
at3	-50	-0.0093	0.0057	0.0009	-0.0008	-0.0085	0.0319	0.0243	-0.0220
at3	150	0.0020	-0.0043	-0.0016	-0.0016	-0.0031	0.0334	0.0262	-0.0227
ap1		-0.0043	-0.0002	-0.0026	0.0001	0.0031	0.0413	0.0268	-0.0242
ac1		-0.0150	0.0026	-0.0017	0.0003	0.0004	0.0363	0.0266	-0.0251
ac1	150	0.0152	0.0131	0.0004	-0.0136	-0.0017	0.1029	0.0224	-0.0352

Mol[a]	Fix ψ	CA-C	C-O	C-NT
at1		0.0068	-0.0298	-0.0098
at1	-50	0.0004	-0.0343	0.0018
at1	150	0.0052	-0.0358	0.0119
at2		0.0040	-0.0300	-0.0069
at2	-50	-0.0014	-0.0333	0.0004
at2	150	0.0046	-0.0357	0.0098
at3		0.0071	-0.0308	-0.0061
at3	-50	0.0020	-0.0346	0.0024
at3	150	0.0056	-0.0358	0.0104
ap1		0.0093	-0.0291	-0.0102
ac1		0.0040	-0.0330	-0.0044
ac1	150	0.0042	-0.0331	-0.0028

[a] Key: see Table 2

Bond angles

Mol[a]	Fix ψ	N-CA-CB	CA-CB-CG	CB-CG-CD	CG-CD-N	CD-N-CA	CD-N-CY	CA-N-CY	N-CY-CAY
at1		0.71	-0.90	0.56	-1.04	-0.38	1.47	-2.23	-2.07
at1	-50	0.35	-0.71	0.54	-1.08	0.79	2.57	-3.66	-2.10
at1	150	-0.53	-0.26	0.00	-1.67	0.27	1.95	-3.58	-2.04
at2		0.46	-0.82	0.35	-0.79	-1.11	3.65	-1.23	-2.23
at2	-50	0.28	-1.01	0.48	-0.94	-0.24	1.78	-3.75	-2.38
at2	150	-1.91	-1.09	-0.36	-2.35	-2.24	1.47	-3.24	-1.89
at3		1.18	0.89	0.96	-1.29	-0.93	4.05	-2.28	-1.38
at3	-50	0.41	0.01	0.68	-1.50	0.24	3.77	-3.67	-1.84
at3	150	-0.96	-0.48	0.26	-1.16	0.89	2.86	-2.90	-1.91
ap1		0.62	-1.14	0.56	-1.02	-0.06	-1.75	-0.47	-2.42
ac1		-0.07	-0.89	0.69	-0.88	0.70	-0.63	-0.12	-2.66
ac1	150	2.50	1.49	-1.11	-2.99	0.09	-2.14	-4.70	-2.33

Mol[a]	Fix ψ	N-CY-OY	OY-CY-CAY	CB-CA-C	N-CA-C	CA-C-NT	CA-C-O	O-C-NT
at1		0.25	1.82	0.44	-1.43	-1.18	0.55	0.62
at1	-50	-0.42	2.53	-1.03	0.06	-0.99	0.46	0.23
at1	150	-0.81	2.84	-0.55	0.72	-0.38	0.50	-0.54
at2		0.32	1.90	0.64	-2.43	-1.31	0.60	0.70
at2	-50	-0.24	2.60	-0.74	-0.16	-0.75	0.27	0.22
at2	150	-0.94	2.79	-0.05	0.94	0.02	0.02	-0.49
at3		-0.30	1.68	1.09	-4.13	-1.56	0.86	0.70
at3	-50	-0.79	2.62	-0.85	-0.87	-1.20	0.65	0.15
at3	150	-0.89	2.81	0.44	0.17	0.01	0.09	-0.54
ap1		0.05	2.36	-0.07	2.57	0.28	-0.42	0.14
ac1		0.05	2.61	-1.06	0.77	-1.12	0.99	0.16
ac1	150	0.11	2.23	-1.39	-4.85	-0.74	1.14	-0.41

[a] Key: see Table 2

Dihedral angles

Mol[a]	Fix ψ	χ1	χ2	χ3	χ4	χ5	ω	ϕ	ψ	Impr[b]
at1		0.03	1.70	-2.57	2.56	-1.50	-8.58	9.84	-2.28	-11.57
at1	-50	2.46	1.58	-4.58	6.46	-5.67	-4.60	1.00	0.11	-6.05
at1	150	1.12	0.58	-1.96	2.59	-2.20	-4.46	-3.20	0.03	1.98
at2		-1.91	-0.03	1.95	-2.99	3.36	-10.56	11.33	-2.64	-7.77
at2	-50	10.66	-0.01	-10.68	18.49	-18.49	-5.81	-8.78	0.10	-8.86
at2	150	-1.52	-0.02	1.39	-2.51	2.80	-6.31	-2.37	0.02	5.96
at3		-13.16	6.26	3.04	-11.89	15.74	-7.61	11.88	-9.29	4.01
at3	-50	1.93	-2.10	0.79	0.60	-1.65	-3.96	-5.46	0.11	3.84
at3	150	-8.23	3.33	2.74	-8.46	10.57	-0.70	-6.36	0.01	18.13
ap1		-61.39	74.26	-60.44	22.84	24.20	-0.21	34.39	-20.37	-7.46
ac1		-0.16	1.44	-2.19	1.91	-1.00	-17.24	13.12	1.64	-15.74
ac1	150	-1.38	0.17	1.18	-2.49	2.56	0.55	2.98	0.07	0.56

[a] Key: see Table 2

[b] Impr=improper dihedral about proline nitrogen=CY-CA-N-CD (see Figure 1 for atom definitions)

Table 6. ***Ab initio*** **and CHARMM normal mode potential energy distribution analysis for AcProNH$_2$ (trans Cγ-endo/Cβ-exo, ψ-unconstrained)**

Mode# *Ab intiio* 6-31g*	Mode# CHARMM	Freq (cm^{-1})	Potential Energy Distribution (%)[a]							
1		45.9	A-C tor	89	ring tor'	7				
	1	57.6	A-C tor	82	ring tor'	9	ring tor	3	Y-N tor	3
2		88.6	Y-N tor	54	NAC def	11	Z-Y tor	10	Y-N wag	9
	2	84.7	Y-N tor	62	ring tor'	22	NAC def	9	Y-N wag	3
3		99.3	ring tor'	101	Y-N wag	-12	ring tor	-7	A-C tor	7
	3	110.5	ring tor'	96	Y-N wag	-26	Y-N tor	17	A-C tor	11
4		130.3	Z-Y tor	76	Y-N tor	11	ring tor'	9	Y-N wag	-6
	5	180.2	Z-Y tor	92	Y-N wag	6				
5		153.7	Y-N wag	91	Y-N tor	22	ring tor'	-20	NAC def	6
	4	162.0	Y-N wag	92	ring tor'	-11	ring tor	6	Y-N tor	5
6		187.2	ring tor	46	BAC def	33	ring tor'	-22	Y-N rck	12
	6	200.7	ring tor	61	ring tor'	-33	BAC def	28	Y-N wag	24
7		223.1	Y-N rck	52	ring tor	16	OyYN def	13	Z-Y rck	8
	7	243.9	Y-N rck	43	OyYN def	23	Z-Y rck	19	ring tor	7
8		268.9	NAC def	32	ring tor'	11	BAC def	11	OCNt rck	11
	8	287.6	NAC def	41	Z-Y rck	11	ring tor'	10	OCNt wag	9
9		309.2	OCNt rck	28	BAC def	18	ring tor	18	CA-C str	7
	10	345.1	OCNt rck	32	BAC def	14	ring tor	11	Z-Y rck	10
10		340.7	Z-Y rck	26	ring def	19	OCNt rck	13	Y-N str	9
	9	319.3	BAC def	17	ring def	17	Y-N str	16	Z-Y rck	12
11		410.6	NH2 wag	25	Z-Y rck	18	OCNt sci	10	OCNt rck	9
	11	427.9	OCNt rck	27	A-B str	12	ring def'	11	OCNt sci	10
12		466.4	NH2 wag	59	OCNt rck	6	OyYN def	6	ring def	5
	13	534.3	C-Nt tor	44	NH2 wag	22	Y-N rck	6	Z-Y rck	5
13		496.8	OyYN def	15	NH2 wag	14	N-D str	14	ring def	11
	12	443.1	Y-N rck	23	OyYN def	13	N-D str	9	OCNt wag	9
14		542.3	OCNt sci	35	ring def'	17	HB rck	8	Z-Y rck	8
	14	557.1	ring def'	29	OCNt sci	20	C-Nt tor	15	CA-C str	6
15		596.0	Y wag	37	C-Nt tor	19	Y a-def	7	ring def'	6
	15	585.6	OCNt wag	28	C-Nt tor	28	NH2 wag	19	ring def'	10
16		599.1	OyYN def	24	Z-Y str	17	Y wag	15	ring def	12
	16	592.8	OyYN def	29	Z-Y str	26	Z-Y rck	13	ring def	12
17		612.8	C-Nt tor	62	Y wag	11	NH2 wag	6	OyYN def	4
	19	734.4	Y wag	63	Y a-def	14	NH2 wag	5	Y a-def'	4
18		649.1	OCNt sci	17	ring def'	14	HG rck	11	OCNt rck	10
	17	649.4	OCNt sci	28	ring def'	11	OCNt rck	10	C-Nt str	8
19		704.3	OCNt wag	38	C-Nt tor	14	N-D str	10	A-C str	8
	18	696.0	NH2 wag	34	OCNt wag	16	ring def	9	N-D str	8
20		779.8	A-C str	25	HB rck	14	A-B str	9	Z-Y str	7
	21	786.3	A-C str	18	HD rck	16	HB rck	10	C-Nt str	7
21		854.0	HG rck	22	B-G str	19	ring def'	9	HD rck	8
	20	771.5	HG rck	36	HB rck	18	HD rck	11	ring def'	11
22		856.0	OCNt wag	21	ring def'	14	N-D str	14	NAC def	13
	22	814.0	OCNt wag	19	ring def	16	NAC def	12	HD rck	10
23		894.0	G-D str	28	B-G str	22	HG rck	8	HD rck	8
	24	884.2	B-G str	17	G-D str	13	HG rck	13	A-B str	10
24		895.3	HD rck	20	B-G str	13	HB rck	12	HG twi	10
	23	873.0	HB rck	22	HD rck	21	HB twi	7	A-C str	7

25		941.4	Z-Y str	17	Y a-def'	13	A-B str	12	G-D str	10
	25	961.4	Y a-def'	21	Y a-def	9	N-A str	8	Z-Y str	8
26		971.0	G-D str	28	Z-Y str	9	HG wag	9	N-A str	6
	28	1015.9	G-D str	21	HB rck	16	HD wag	10	B-G str	10
27		1013.8	A-B str	30	Y a-def'	12	HB wag	10	B-G str	9
	26	980.0	HG rck	14	HG wag	12	A-B str	11	B-G str	11
28		1044.1	Y a-def	56	Y wag	26	Y a-def'	14	Y rck'	6
	31	1087.4	Y a-def	52	Y wag	23	Y a-def'	15	Y rck	5
29		1056.2	Y a-def'	27	N-A str	9	HA rck'	6	Z-Y str	4
	27	987.7	Y a-def'	27	HG wag	16	G-D str	9	Y a-def	7
30		1084.8	NH2 rck	18	HB rck	14	HD rck	11	HG rck	10
	30	1072.2	HB wag	13	HA rck	11	HG wag	10	HA rck'	10
31		1121.9	NH2 rck	37	C-Nt str	14	C-O str	10	HD rck	10
	33	1145.6	NH2 rck	47	HB twi	17	C-O str	8	C-Nt str	8
32		1159.9	HB twi	23	HD twi	21	HD rck	21	HG twi	10
	32	1122.7	HB twi	62	NH2 rck	11	C-O str	6	C-Nt str	5
33		1195.3	HG twi	42	HB twi	9	B-G str	7	ring def	6
	29	1057.7	HG twi	84	HG wag	5				
34		1200.1	N-D str	24	N-A str	19	HD twi	15	ring def'	10
	35	1203.1	HD twi	24	HD wag	14	N-A str	13	N-D str	9
35		1244.9	HD twi	45	HB twi	30	HG rck	9	HA rck'	3
	34	1164.1	HD twi	62	HD wag	9	HG wag	6	N-A str	4
36		1269.3	HA rck'	34	C-Nt str	21	NH2 rck	14	OCNt sci	8
	38	1290.4	HA rck'	34	C-Nt str	15	NH2 rck	14	A-B str	8
37		1297.1	HA rck	51	HD wag	10	HG wag	9	HB wag	7
	39	1330.5	HA rck	14	HD sci	12	N-D str	11	OyYN def	10
38		1323.3	HB wag	48	HG wag	10	HG twi	10	A-B str	7
	37	1259.7	HB wag	52	B-G str	20	A-B str	5	ring def	4
39		1336.6	HG wag	53	HG twi	8	HA rck	7	HD twi	6
	36	1229.6	HG wag	39	G-D str	21	B-G str	11	HB wag	7
40		1353.8	HD wag	41	HB wag	14	Y s-def	14	Y-N str	13
	40	1351.1	HD wag	17	HD sci	14	HA rck	13	G-D str	7
41		1367.8	Y s-def	43	HD wag	26	Y-N str	8	HA rck	5
	47	1482.3	HD sci	32	HD wag	14	N-D str	13	HA rck'	6
42		1402.4	HA rck'	29	C-Nt str	12	A-C str	12	Y s-def	11
	48	1529.1	HA rck	30	HA rck'	17	N-A str	14	NH2 sci	14
43		1419.1	Y s-def	25	Y-N str	25	Y rck	9	Z-Y str	8
	41	1393.9	Y s-def	93						
44		1443.7	Y rck	80	Y a-def'	6	Y-N str	5		
	44	1430.5	Y rck	45	HG sci	33	Y rck'	14	Y a-def	4
45		1455.8	Y rck'	93	Y a-def	5				
	42	1417.6	Y rck'	67	Y rck	20	Y a-def'	5		
46		1458.7	HB sci	78	HG sci	18				
	46	1464.0	HB sci	71	HD sci	10				
47		1473.0	HG sci	69	HB sci	15	HD sci	12		
	43	1430.1	HG sci	62	Y rck	24	Y rck'	8		
48		1499.3	HD sci	85	HG sci	10				
	45	1449.4	HD sci	26	HB sci	22	HA rck'	8	ring def	6
49		1610.5	NH2 sci	88	C-Nt str	8	C-O str	3		
	49	1617.3	NH2 sci	61	C-O str	10	C-Nt str	9	A-C str	6
50		1700.5	Y-Oy str	80	Y-N rck	3				
	50	1643.7	Y-Oy str	64	Z-Y str	8	Y-N rck	7	Y-N str	4
51		1751.4	C-O str	79	C-Nt str	5	NH2 rck	4		
	51	1745.8	C-O str	54	C-Nt str	19	OCNt rck	9	NH2 sci	7
52		2865.2	HD s str	77	HD a str	17				
	52	2853.4	HD s str	99						

53	2877.7	HB s str	68	HB a str	27				
54	2899.3	HB s str	52	HG s str	40	HD a str	6	B-G str	4
54	2879.2	HY2 str	41	HY1 str	41	HY3 str	13		
60	2973.5	HY2 str	68	HY3 str	16	HY1 str	16		
55	2903.4	HG s str	99	HG a str	-11	HD a str	6	B-G str	6
55	2906.6	HG s str	41	HB s str	35	HA str	24		
56	2917.1	HD a str	68	HD s str	20	HG s str	8		
53	2892.8	HD a str	91	HG s str	8				
57	2932.9	HA str	97						
56	2907.8	HA str	75	HG s str	14	HB s str	10		
58	2933.1	HY1 str	49	HY2 str	48				
57	2917.1	HY1 str	34	HY3 str	34	HY2 str	32		
59	2949.8	HG a str	103	HG s str	-14	HD a str	6	HB s str	4
59	2942.0	HG a str	60	HB a str	42	HG s str	-3		
60	2970.7	HB a str	71	HB s str	25	HG a str	4		
58	2933.8	HB a str	56	HG a str	47	HG s str	-5		
61	2982.1	HY3 str	87	HY1 str	7	HY2 str	6		
61	2975.4	HY1 str	50	HY3 str	50				
62	3393.9	NH2 s str	111	NH2 a str	-11				
62	3407.1	NH2 s str	104	NH2 a str	-4				
63	3515.3	NH2 a str	111	NH2 s str	-11				
63	3534.4	NH2 a str	104	NH2 s str	-5				

[a]Abbreviations: str=stretch; tor=torsion; rck=rock; def=deformation; twi=twist; s=symmetric; a=asymmetric; Z=atom CAY; Y=CY; A=CA; B=CB; G=CG; D=CD; NH2=NT-HT1,HT2

Table 7. Crystal minimization statistics

Crystal		16 Å[a]	18 Å[a]	20 Å[a]	23 Å[a]	X-ray
APNMA[b]	ENERGY	-11.97	-11.94	-11.98	-11.84	
	RMS	0.01	0.02	0.01	0.00	
	IMGVDW	-19.07	-19.32	-19.38	-19.37	
	IMGELEC	-12.89	-12.55	-12.31	-11.97	
	A	9.81	9.75	9.71	9.72	9.74
	B	13.20	12.97	13.11	13.20	13.20
	C	6.94	7.07	7.02	6.95	7.17
	dA	0.07	0.01	-0.03	-0.02	0.00
	dB	-0.00	-0.23	-0.09	-0.00	0.00
	dC	-0.23	-0.10	-0.15	-0.22	0.00
	Volume	898.32	892.90	893.25	891.94	921.83
	dVolume	-23.51	-28.94	-28.58	-29.89	0.00
CPG	ENERGY	27.12	27.69	27.84	28.33	
	RMS	0.00	0.00	0.00	0.00	
	IMGVDW	-17.38	-17.42	-17.44	-17.57	
	IMGELEC	-11.46	-11.31	-11.45	-11.20	
	A	10.34	10.15	10.10	10.18	9.67
	B	5.54	5.65	5.61	5.53	5.87
	C	12.71	12.91	13.17	13.12	13.07
	dA	0.68	0.48	0.44	0.51	0.00
	dB	-0.33	-0.22	-0.26	-0.34	0.00
	dC	-0.35	-0.16	0.11	0.05	0.00
	Volume	728.85	740.29	747.01	738.82	741.41
	dVolume	-12.57	-1.12	5.59	-2.59	0.00
CPL	ENERGY	16.63	17.07	17.19	17.34	
	RMS	0.00	0.01	0.01	0.00	
	IMGVDW	-20.94	-21.10	-21.24	-21.21	
	IMGELEC	-9.74	-9.46	-9.42	-9.57	
	A	9.05	9.15	9.23	9.32	9.45
	B	19.86	19.85	19.80	19.67	19.59
	C	6.51	6.41	6.41	6.33	6.34
	dA	-0.40	-0.30	-0.22	-0.13	0.00
	dB	0.27	0.27	0.21	0.08	0.00
	dC	0.17	0.07	0.07	-0.01	0.00
	Volume	1169.65	1164.06	1170.77	1161.43	1173.64
	dVolume	-3.99	-9.58	-2.87	-12.21	0.00
LPG	ENERGY	-138.94	-138.48	-139.08	-139.57	
	RMS	0.00	0.00	0.00	0.00	
	IMGVDW	-17.88	-17.24	-18.80	-17.98	
	IMGELEC	-170.83	-169.93	-164.82	-164.33	
	A	9.72	10.08	9.72	9.87	9.44
	B	6.33	6.19	6.35	6.25	6.72

	C	12.11	12.20	12.12	12.15	12.11
	dA	0.28	0.64	0.28	0.43	0.00
	dB	-0.39	-0.53	-0.37	-0.48	0.00
	dC	0.00	0.09	0.02	0.05	0.00
	BETA	101.47	103.71	100.74	101.97	100.18
	dBETA	1.29	3.53	0.56	1.79	0.00
	Volume	745.20	761.00	748.06	749.60	768.52
	dVolume	-23.32	-7.52	-20.47	-18.93	0.00
LPG2	ENERGY	-135.68	-134.68	-135.78	-135.85	
	RMS	0.00	0.01	0.00	0.00	
	IMGVDW	-18.46	-18.72	-18.91	-18.82	
	IMGELEC	-169.01	-165.14	-163.48	-160.99	
	A	9.53	9.67	9.58	9.64	9.44
	B	6.49	6.46	6.48	6.45	6.72
	C	12.11	12.14	12.11	12.14	12.11
	dA	0.08	0.23	0.13	0.19	0.00
	dB	-0.23	-0.26	-0.24	-0.28	0.00
	dC	0.00	0.04	0.00	0.04	0.00
	BETA	101.76	102.73	101.50	102.31	100.18
	dBETA	1.58	2.55	1.32	2.13	0.00
	Volume	748.83	758.97	751.41	754.44	768.52
	dVolume	-19.69	-9.56	-17.11	-14.08	0.00
LPRO	ENERGY	-84.30	-86.28	-85.08	-86.45	
	RMS	0.00	0.00	0.00	0.00	
	IMGVDW	-6.67	-7.12	-6.86	-7.69	
	IMGELEC	-64.27	-64.38	-62.20	-61.96	
	A	12.11	11.92	11.94	12.10	11.55
	B	8.38	8.44	8.41	8.33	9.02
	C	5.32	5.37	5.35	5.44	5.20
	dA	0.56	0.37	0.39	0.55	0.00
	dB	-0.64	-0.58	-0.61	-0.69	0.00
	dC	0.12	0.17	0.15	0.24	0.00
	Volume	540.68	540.46	537.38	548.31	541.74
	dVolume	-1.06	-1.28	-4.36	6.57	0.00

[a] Values of cutim, the image non-bonded cutoff, for four minimizations.

[b] Units and Abbreviations: APNMA: N-acetylproline N′-methylamide; CPG: cyclic prolyl glycine; CPL: cyclic prolyl leucine; LPG: leucyl-prolyl-glycine; LPRO: L-proline (zwitterion); A, B, C: crystal unit cell dimensions (in Å); dA, dB, dC: difference in A, B, C between minimizations and x-ray structure; Volume: unit cell volume (in $Å^3$); dVolume: difference between minimized and x-ray unit cell volumes. Energies in kcal/mol.

Table 8. CHARMM minimized, x-ray, and difference internal coordinates for six peptide crystal structures

CHARMM minimized coordinates[a]

Atom 1	Atom 2	Atom 3	Atom 4	apnma	cpg	cpl	lpg	lpg2	lpro	Aver.
n N	n CA				1.47	1.46	1.48	1.48		1.47
n CA	n C			1.49	1.51	1.53	1.52	1.52		1.51
n C	n O			1.23	1.23	1.23	1.23	1.23		1.23
n C	p N			1.32	1.32	1.32	1.33	1.34		1.33
n CA	n CB					1.54	1.55	1.55		1.55
n CB	n CG					1.54	1.54	1.54		1.54
n CG	n CD1					1.54	1.53	1.53		1.53
n CD2	n CG					1.54	1.53	1.53		1.53
p N	p CA			1.46	1.46	1.46	1.46	1.46	1.50	1.47
p CA	p C			1.52	1.51	1.50	1.52	1.52	1.52	1.51
p C	p O			1.23			1.23	1.23		1.23
p C	p OT1								1.26	1.26
p C	p OT2								1.26	1.26
p C	c N			1.34	1.34	1.34	1.34	1.34		1.34
p CA	p CB			1.54	1.53	1.53	1.54	1.54	1.53	1.53
p CB	p CG			1.53	1.53	1.53	1.53	1.53	1.53	1.53
p CG	p CD			1.52	1.53	1.53	1.52	1.52	1.53	1.53
p CD	p N			1.46	1.46	1.46	1.47	1.47	1.51	1.47
c N	c CA			1.44	1.47	1.46	1.45	1.45		1.45
c CA	c C				1.51	1.53	1.52	1.52		1.52
c C	c OT1						1.26	1.26		1.26
c C	c OT2						1.26	1.26		1.26
n N	n CA	n C			113.4	112.7	111.2	111.4		112.2
n CA	n C	n O		119.6	122.7	124.0	118.7	118.0		120.6
n CA	n C	p N		118.9	113.5	113.3	120.6	121.5		117.5
n O	n C	p N		121.5	123.7	122.7	120.7	120.6		121.8
n N	n CA	n CB				110.7	109.8	110.1		110.2
n CB	n CA	n C				110.6	110.6	110.3		110.5
n CA	n CB	n CG				115.6	114.7	114.1		114.8
n CB	n CG	n CD1				109.5	111.9	111.2		110.9
n CB	n CG	n CD2				112.6	109.7	110.1		110.8
n C	p N	p CA		122.9	123.7	123.9	122.8	122.6		123.2
n C	p N	p CD		123.5	122.8	122.5	125.0	124.9		123.7
p N	p CA	p C		113.6	109.8	109.7	111.9	113.3	110.5	111.5
p CA	p C	c N		119.9	114.7	115.4	117.8	118.5		117.3
p CB	p CA	p C		110.8	115.6	115.1	111.0	109.4	111.9	112.3
p N	p CA	p CB		103.1	102.3	102.2	102.8	103.6	103.4	102.9
p CB	p CG	p CD		102.8	103.3	103.4	102.5	102.5	102.5	102.8
p CG	p CD	p N		104.2	104.4	104.0	104.8	103.8	104.7	104.3
p CA	p CB	p CG		104.8	103.0	102.3	104.4	104.7	104.1	103.9
p CD	p N	p CA		112.8	112.4	112.4	112.2	111.7	109.4	111.8
p CA	p C	p O		118.4			120.3	119.8		119.5
p CA	p C	p OT1							120.6	120.6
p CA	p C	p OT2							117.9	117.9
p C	c N	c CA		122.2	121.9	122.2	121.0	120.2		121.5
c N	c CA	c C			113.4	112.7	112.8	113.7		113.1
c CA	c C	c OT1					121.2	120.9		121.1
c CA	c C	c OT2					117.2	117.3		117.3
n N	n CA	n C	p N		27.5	25.1	146.9	151.4		87.7
n CA	n C	p N	p CA	-178.5	12.9	16.7	-175.4	-177.9		-100.4
n CA	n C	p N	p CD	-9.1	179.7	-176.8	4.7	-9.0		-2.1
n O	n C	p N	p CA	1.8	-169.4	-164.5	3.2	1.8		-65.4
n C	p N	p CA	p C	-77.6	-46.4	-47.3	-74.5	-68.8		-62.9
p N	p CA	p C	c N	-8.4	37.0	33.2	165.1	163.4		78.1
p N	p CA	p C	p OT2						150.2	150.2
p CA	p C	c N	c CA	-178.9	1.0	6.5	-177.6	178.3		-34.2

p C	c N	c CA	c C		-35.1	-37.5	-178.4	-176.9		-107.0
c N	c CA	c C	c OT2				-176.7	-174.9		-175.8
n N	n CA	n CB	n CG			-67.2	-81.1	-81.6		-76.6
n CA	n CB	n CG	n CD1			175.7	-66.6	-68.8		13.4
n C	p N	p CA	p CB	162.4	-169.6	-169.9	166.4	172.7		32.4
n C	p N	p CD	p CG	175.6	-167.6	-166.4	170.7	-149.8		-27.5
p CD	p N	p CA	p C	111.9	145.6	145.0	105.5	121.0	106.5	122.6
p N	p CA	p CB	p CG	26.9	-35.8	-36.8	31.3	-24.2	33.1	-0.9
p CA	p CB	p CG	p CD	-35.4	36.6	38.4	-36.9	36.3	-40.2	-0.2
p CB	p CG	p CD	p N	30.0	-23.2	-25.0	28.3	-34.2	31.5	1.2
p CG	p CD	p N	p CA	-14.0	0.5	1.6	-9.2	20.1	-11.4	-2.1
p CD	p N	p CA	p CB	-8.1	22.4	22.4	-13.7	2.5	-13.4	2.0
c N	n C	* n CA	n CB			124.5				124.5
n N	n C	* n CA	n CB				122.3	122.6		122.4
n CD1	n CB	* n CG	n CD2			122.8	-120.8	-120.7		-39.6
p N	n CA	* n C	n O	179.7	-177.7	-178.8	-178.6	-179.6		-107.0
n C	p CA	* p N	p CD	-170.4	-168.0	-167.7	179.9	-170.2		-99.3
p N	p C	* p CA	p CB	115.5	115.1	114.6	114.2	115.1	114.6	114.8
c N	p CA	* p C	p O	-179.5			-180.0	178.4		-60.4
p OT2	p CA	* p C	p OT1						177.4	177.4
c OT2	c CA	* c C	c OT1				179.2	-179.9		-0.4

X-ray crystallographic coordinates

Atom 1	Atom 2	Atom 3	Atom 4	apnma	cpg	cpl	lpg	lpg2	lpro	Aver.
n N	n CA				1.45	1.46	1.48	1.48		1.47
n CA	n C			1.49	1.51	1.52	1.53	1.53		1.52
n C	n O			1.24	1.24	1.24	1.23	1.23		1.24
n C	p N			1.34	1.33	1.34	1.33	1.33		1.33
n CA	n CB					1.54	1.53	1.53		1.54
n CB	n CG					1.53	1.53	1.53		1.53
n CG	n CD1					1.52	1.52	1.52		1.52
n CD2	n CG					1.53	1.52	1.52		1.52
p N	p CA			1.47	1.46	1.46	1.47	1.47	1.53	1.48
p CA	p C			1.53	1.51	1.51	1.53	1.53	1.53	1.52
p C	p O			1.23			1.22	1.22		1.22
p C	p OT1								1.28	1.28
p C	p OT2								1.26	1.26
p C	c N			1.32	1.35	1.36	1.34	1.34		1.34
p CA	p CB			1.53	1.53	1.52	1.50	1.63	1.52	1.54
p CB	p CG			1.50	1.53	1.49	1.52	1.52	1.54	1.52
p CG	p CD			1.53	1.52	1.51	1.54	1.53	1.53	1.53
p CD	p N			1.48	1.47	1.48	1.48	1.48	1.48	1.48
c N	c CA			1.45	1.45	1.46	1.45	1.45		1.45
c CA	c C				1.51	1.52	1.51	1.51		1.52
c C	c OT1						1.25	1.25		1.25
c C	c OT2						1.26	1.26		1.26
n N	n CA	n C			112.0	110.7	106.9	106.9		109.1
n CA	n C	n O		122.9	122.2	124.0	119.9	119.9		121.8
n CA	n C	p N		117.0	114.2	113.9	117.1	117.1		115.9
n O	n C	p N		120.1	123.6	122.1	123.0	123.0		122.4
n N	n CA	n CB				110.1	111.3	111.3		110.9
n CB	n CA	n C				110.3	110.1	110.1		110.2
n CA	n CB	n CG				114.6	116.0	116.0		115.5
n CB	n CG	n CD1				107.4	110.4	110.4		109.4
n CB	n CG	n CD2				110.7	109.4	109.4		109.9
n C	p N	p CA		121.4	122.4	123.4	120.7	120.7		121.7
n C	p N	p CD		125.5	124.2	124.0	127.2	127.2		125.6
p N	p CA	p C		114.3	110.8	110.9	110.6	110.6	106.4	110.6
p CA	p C	c N		117.9	113.6	114.0	114.1	114.1		114.8
p CB	p CA	p C		111.5	116.7	115.3	117.0	104.1	111.5	112.7
p N	p CA	p CB		103.4	102.8	103.2	105.1	99.7	106.3	103.4
p CB	p CG	p CD		104.2	105.1	107.2	105.9	98.4	103.7	104.1
p CG	p CD	p N		102.8	103.3	101.5	101.2	103.9	105.3	103.0

p CA	p CB	p CG		104.7	103.2	103.0	101.8	103.8	100.8	102.9
p CD	p N	p CA		112.2	112.9	112.5	112.1	112.1	107.1	111.5
p CA	p C	p O		117.7			122.5	122.5		120.9
p CA	p C	p OT1							119.0	119.0
p CA	p C	p OT2							120.8	120.8
p C	c N	c CA		121.0	122.7	123.1	120.2	120.2		121.4
c N	c CA	c C			112.0	110.7	112.7	112.7		112.0
c CA	c C	c OT1					119.3	119.3		119.3
c CA	c C	c OT2					115.9	115.9		115.9
n N	n CA	n C	p N		32.5	33.8	152.2	152.2		92.7
n CA	n C	p N	p CA	-177.0	7.2	6.2	175.6	175.6		37.5
n CA	n C	p N	p CD	-8.4	178.6	-176.7	-2.6	-2.6		-2.3
n O	n C	p N	p CA	4.3	-172.8	-173.3	-2.8	-2.8		-69.5
n C	p N	p CA	p C	-76.3	-43.8	-41.6	-66.7	-66.7		-59.0
p N	p CA	p C	c N	-15.9	38.2	33.7	160.7	160.7		75.5
p N	p CA	p C	p OT2						170.6	170.6
p CA	p C	c N	c CA	-177.3	0.7	6.4	175.9	175.9		36.3
p C	c N	c CA	c C		-37.4	-41.6	-175.8	-175.8		-107.6
c N	c CA	c C	c OT2				179.6	179.6		179.6
n N	n CA	n CB	n CG			-72.4	-82.0	-82.0		-78.8
n CA	n CB	n CG	n CD1			177.7	-68.5	-68.5		13.6
n C	p N	p CA	p CB	162.3	-169.2	-165.6	166.1	-175.9		-36.5
n C	p N	p CD	p CG	176.3	-168.8	-172.9	170.6	-155.9		-30.1
p CD	p N	p CA	p C	113.7	143.9	141.0	111.7	111.7	105.3	121.2
p N	p CA	p CB	p CG	27.0	-32.3	-31.6	32.0	-30.3	33.6	-0.3
p CA	p CB	p CG	p CD	-36.2	35.4	36.0	-37.6	45.0	-41.0	0.3
p CB	p CG	p CD	p N	30.7	-24.0	-25.1	28.0	-43.3	33.6	0.0
p CG	p CD	p N	p CA	-14.2	3.4	4.4	-7.7	25.8	-12.4	-0.1
p CD	p N	p CA	p CB	-7.7	18.5	17.0	-15.5	2.6	-13.7	0.2
c N	n C	* n CA	n CB			122.1				122.1
n N	n C	* n CA	n CB				121.0	121.0		121.0
n CD1	n CB	* n CG	n CD2			120.6	-121.3	-121.3		-40.7
p N	n CA	* n C	n O	178.7	179.9	179.6	178.4	178.4		179.0
n C	p CA	* p N	p CD	-170.0	-172.3	-177.4	178.4	178.4		-32.6
p N	p C	* p CA	p CB	116.8	117.2	116.7	120.3	106.2	115.4	115.5
c N	p CA	* p C	p O	-178.9			176.6	176.6		58.1
p OT2	p CA	* p C	p OT1						-177.4	-177.4
c OT2	c CA	* c C	c OT1				177.7	177.7		177.7

CHARMM - x-ray coordinates

Atom 1	Atom 2	Atom 3	Atom 4	apnma	cpg	cpl	lpg	lpg2	lpro	Aver.
n N	n CA				0.01	0.00	0.00	0.00		0.00
n CA	n C				0.00	0.01	-0.01	-0.01		0.00
n C	n O			-0.02	-0.01	-0.01	0.00	0.00		-0.01
n C	p N			-0.02	-0.01	-0.02	0.00	0.01		-0.01
n CA	n CB					0.01	0.02	0.01		0.01
n CB	n CG					0.01	0.01	0.01		0.01
n CG	n CD1					0.01	0.01	0.01		0.01
n CD2	n CG					0.01	0.01	0.01		0.01
p N	p CA			-0.01	0.00	-0.01	-0.01	0.00	-0.03	-0.01
p CA	p C			-0.01	0.00	0.00	-0.01	-0.01	-0.01	-0.01
p C	p O			0.00	0.00	0.00	0.01	0.01	0.00	0.01
p C	p OT1								-0.01	-0.01
p C	p OT2								0.00	0.00
p C	c N			0.03	-0.01	-0.02	0.00	0.00		0.00
p CA	p CB			0.01	0.00	0.02	0.04	-0.09	0.01	0.00
p CB	p CG			0.03	0.01	0.05	0.01	0.01	-0.01	0.01
p CG	p CD			-0.01	0.01	0.02	-0.02	-0.01	-0.01	0.00
p CD	p N			-0.01	-0.01	-0.03	-0.01	-0.01	0.02	-0.01
c N	c CA			-0.01	0.01	0.00	-0.01	0.00	0.00	0.00
c CA	c C					0.01	0.01	0.01	0.00	0.01
c C	c OT1						0.01	0.01	0.00	0.01
c C	c OT2						0.00	0.01	0.00	0.00

n N	n CA	n C			1.4	2.0	4.3	4.5		3.1
n CA	n C	n O		-3.2	0.5	0.0	-1.3	-1.9		-1.5
n CA	n C	p N		1.9	-0.7	-0.7	3.5	4.4		1.7
n O	n C	p N		1.4	0.2	0.7	-2.2	-2.4		-0.5
n N	n CA	n CB				0.6	-1.4	-1.2		-0.7
n CB	n CA	n C				0.3	0.4	0.1		0.3
n CA	n CB	n CG				1.0	-1.3	-1.8		-0.7
n CB	n CG	n CD1				2.1	1.5	0.8		1.5
n CB	n CG	n CD2				1.8	0.3	0.6		0.9
n C	p N	p CA		1.5	1.3	0.5	2.1	1.9		1.4
n C	p N	p CD		-2.0	-1.4	-1.5	-2.2	-2.3		-1.9
p N	p CA	p C		-0.7	-1.0	-1.2	1.2	2.7	4.0	0.8
p CA	p C	c N		2.0	1.1	1.4	3.7	4.4		2.5
p CB	p CA	p C		-0.7	-1.1	-0.2	-6.0	5.3	0.4	-0.4
p N	p CA	p CB		-0.3	-0.5	-1.0	-2.4	4.0	-2.8	-0.5
p CB	p CG	p CD		-1.4	-1.8	-3.9	-3.4	4.0	-1.2	-1.3
p CG	p CD	p N		1.4	1.1	2.5	3.5	-0.1	-0.6	1.3
p CA	p CB	p CG		0.1	-0.1	-0.8	2.5	0.9	3.3	1.0
p CD	p N	p CA		0.6	-0.5	-0.1	0.1	-0.3	2.4	0.4
p CA	p C	p O		0.8	0.0	0.0	-2.1	-2.7		-1.3
p CA	p C	p OT1							1.6	1.6
p CA	p C	p OT2							-2.9	-2.9
p C	c N	c CA		1.2	-0.7	-0.8	0.8	-0.1		0.1
c N	c CA	c C			1.4	2.0	0.1	0.9		1.1
c CA	c C	c OT1					1.9	1.6		1.7
c CA	c C	c OT2					1.3	1.4		1.4
n N	n CA	n C	p N		-5.1	-8.6	-5.3	-0.8		-5.0
n CA	n C	p N	p CA	-1.6	5.8	10.5	9.0	6.5		6.0
n CA	n C	p N	p CD	-0.7	1.1	-0.1	7.3	-6.4		0.2
n O	n C	p N	p CA	-2.5	3.4	8.8	5.9	4.5		4.0
n C	p N	p CA	p C	-1.3	-2.5	-5.7	-7.7	-2.1		-3.9
p N	p CA	p C	c N	7.4	-1.3	-0.5	4.4	2.8		2.6
p N	p CA	p C	p OT2						-20.4	-20.4
p CA	p C	c N	c CA	-1.6	0.3	0.1	6.5	2.4		1.5
p C	c N	c CA	c C		2.3	4.0	-2.6	-1.1		0.6
c N	c CA	c C	c OT2				3.7	5.5		4.6
n N	n CA	n CB	n CG			5.1	0.9	0.5		2.2
n CA	n CB	n CG	n CD1			-1.9	1.8	-0.3		-0.1
n C	p N	p CA	p CB	0.1	-0.5	-4.3	0.3	-11.4		-3.2
n C	p N	p CD	p CG	-0.7	1.2	6.6	0.1	6.1		2.7
p CD	p N	p CA	p C	-1.8	1.8	4.0	-6.3	9.3	1.3	1.4
p N	p CA	p CB	p CG	-0.1	-3.5	-5.2	-0.8	6.1	-0.5	-0.7
p CA	p CB	p CG	p CD	0.8	1.2	2.4	0.7	-8.7	0.7	-0.5
p CB	p CG	p CD	p N	-0.8	0.8	0.1	0.2	9.1	-2.1	1.2
p CG	p CD	p N	p CA	0.3	-2.8	-2.8	-1.5	-5.7	1.0	-1.9
p CD	p N	p CA	p CB	-0.4	3.9	5.4	1.8	-0.1	0.3	1.8
c N	n C	* n CA	n CB			2.4				2.4
n N	n C	* n CA	n CB				1.3	1.6		1.4
n CD1	n CB	* n CG	n CD2			2.2	0.5	0.6		1.1
p N	n CA	* n C	n O	1.0	2.3	1.7	3.0	2.0		2.0
n C	p CA	* p N	p CD	-0.5	4.3	9.6	1.5	11.4		5.3
p N	p C	* p CA	p CB	-1.3	-2.1	-2.2	-6.2	8.8	-0.8	-0.6
c N	p CA	* p C	p O	-0.6			3.4	1.7		1.5
p OT2	p CA	* p C	p OT1						-5.2	-5.2
c OT2	c CA	* c C	c OT1				1.5	2.4		1.9

[a] Bond lengths between atoms 1 and 2. Bond angles between atoms 1, 2, and 3. Dihedral angles between atoms 1, 2, 3, and 4. Improper dihedrals are marked with a * on the third atom.

Table 9. Crystal minimization: Selected intermolecular distances

	Atom 1	Atom 2	X-ray distance	Min (23Å) distance	X-ray-min difference
APNMA	n O	c HN	2.05	1.92	0.13
	n CA	n O	3.42	3.40	0.02
	n C	n O	3.39	3.40	-0.01
	p N	n O	3.26	3.31	-0.05
	p CD	n O	3.38	3.42	-0.04
	p O	n CA	3.23	3.27	-0.04
	p O	n C	3.43	3.42	0.01
	c N	n O	2.88	2.87	0.01
CPG	p CA	c O	3.29	3.42	-0.13
	p C	c O	3.35	3.33	0.02
	p O	p CB	3.44	3.31	0.13
	p O	p CG	3.42	3.47	-0.05
	p O	c CA	3.49	3.46	0.03
	p O	p N	3.18	3.41	-0.23
	p O	c CA	3.34	3.40	0.06
	p O	c C	3.02	3.07	-0.05
	p O	c O	3.40	3.29	0.11
	c N	c O	2.83	2.80	0.03
	c O	p CD	3.22	3.27	-0.05
	c O	p CG	3.44	3.43	0.01
CPL	p CA	c O	3.44	3.49	-0.05
	p CB	c O	3.36	3.48	-0.12
	p O	p CA	3.35	3.39	-0.04
	p O	c CA	3.26	3.31	-0.05
	c N	c O	2.97	2.89	0.08
LPG	n N	c C	3.49	3.18	0.31
	n N	c OT2	2.83	2.69	0.14
	n N	OH2	3.09	3.39	-0.30
	n N	c OT2	2.90	2.68	0.22
	n CA	OH2	3.37	3.34	0.03
	n CA	c OT2	3.50	3.45	0.05
	p CA	n O	3.37	3.45	-0.08
	p CB	n O	3.17	3.46	-0.29
	p CG	p O	3.37	3.34	0.03
	c N	n O	3.19	2.98	0.21
	c OT1	n N	2.83	2.70	0.13
	c OT1	n O	3.41	3.28	0.13
	c OT1	OH2	2.81	2.70	0.11
LPG2	n N	c C	3.49	3.19	0.30
	n N	c OT2	2.83	2.69	0.14
	n N	OH2	3.09	3.00	0.09

	n N	c OT2	2.90	2.69	0.21
	n CA	OH2	3.37	3.41	-0.04
	n CA	c OT2	3.50	3.39	0.11
	p CG	p O	3.21	3.07	0.14
	c N	n O	3.19	3.10	0.09
	c OT1	n N	2.83	2.70	0.13
	c OT1	n O	3.41	3.40	0.01
	c OT1	OH2	2.81	2.71	0.10
LPRO	p N	p OT2	2.71	2.77	-0.06
	p OT1	p N	2.69	2.71	-0.02
	p OT1	p CD	3.15	3.17	-0.02
	p OT1	p CA	3.42	3.22	0.20
	p OT1	p CB	3.24	3.31	-0.07

Table 10 Crystal molecular dynamics simulations: Internal coordinates, fluctuations, and differences from x-ray structures

CHARMM dynamics averaged internal coordinates

Atom 1	Atom 2	Atom 3	Atom 4	apnma	cpg	cpl	lpg	lpg2	lpro	Aver.
n N	n CA				1.46	1.46	1.49	1.49		1.47
n CA	n C			1.49	1.51	1.53	1.52	1.52		1.51
n C	n O			1.23	1.23	1.23	1.23	1.23		1.23
n C	p N			1.32	1.32	1.32	1.34	1.34		1.33
n CA	n CB					1.55	1.55	1.55		1.55
n CB	n CG					1.54	1.54	1.54		1.54
n CG	n CD1					1.54	1.54	1.53		1.54
n CD2	n CG					1.54	1.53	1.53		1.53
p N	p CA			1.46	1.45	1.46	1.46	1.46	1.49	1.46
p CA	p C			1.52	1.51	1.51	1.52	1.52	1.51	1.52
p C	p O			1.23	1.23	1.23	1.23	1.23		1.23
p C	p OT1								1.26	1.26
p C	p OT2								1.26	1.26
p C	c N			1.34	1.34	1.34	1.34	1.34		1.34
p CA	p CB			1.54	1.53	1.54	1.54	1.54	1.53	1.54
p CB	p CG			1.53	1.53	1.54	1.53	1.53	1.52	1.53
p CG	p CD			1.53	1.53	1.53	1.53	1.53	1.53	1.53
p CD	p N			1.46	1.46	1.46	1.47	1.47	1.50	1.47
c N	c CA			1.44	1.46	1.46	1.45	1.45		1.45
c CA	c C				1.51	1.53	1.53	1.53		1.52
c C	c OT1						1.26	1.26		1.26
c C	c OT2						1.26	1.26		1.26
n N	n CA	n C			113.3	112.5	111.0	111.0		112.0
n CA	n C	n O		119.0			118.1	118.1		118.4
n CA	n C	p N		119.6	113.6	113.1	121.0	121.1		117.7
n O	n C	p N			123.8	122.7				123.2
n N	n CA	n CB				111.0	110.1	110.2		110.4
n CB	n CA	n C				110.5	109.7	109.7		110.0
n CA	n CB	n CG				115.7	114.9	114.8		115.1
n CB	n CG	n CD1				109.5	112.0	112.1		111.2
n CB	n CG	n CD2				112.5	109.3	109.3		110.4
n C	p N	p CA		122.5	123.1	123.8	122.6	122.6		122.9
n C	p N	p CD		123.5	122.4	122.1	124.8	124.8		123.5
p N	p CA	p C		113.6	109.8	109.6	112.1	112.1	109.8	111.2
p CA	p C	c N		119.5	114.6	115.2	117.8	117.8		117.0
p CB	p CA	p C		111.0	115.7	115.3	111.6	111.5	112.1	112.9
p N	p CA	p CB		103.2	102.3	102.9	102.9	102.9	103.6	102.9
p CB	p CG	p CD		102.9	103.5	103.8	102.6	102.6	103.0	103.1
p CG	p CD	p N		104.4	104.2	103.8	104.8	104.8	104.3	104.4
p CA	p CB	p CG		104.9	102.9	103.8	104.8	104.8	104.2	104.2
p CD	p N	p CA		112.4	112.0	112.3	112.0	111.9	109.4	111.7
p CA	p C	p O		118.4	122.0	121.4	120.4	120.4		120.5
p CA	p C	p OT1							120.5	120.5
p CA	p C	p OT2							117.7	117.7
p C	c N	c CA		122.5	121.8	122.1	121.1	121.1		121.7
c N	c CA	c C			113.3	112.5	112.8	112.8		112.9
c CA	c C	c OT1					120.9	120.9		120.9
c CA	c C	c OT2					117.2	117.2		117.2
n N	n CA	n C	p N		26.8	25.5	149.0	149.1		87.6
n CA	n C	p N	p CA	-179.3	13.3	16.0	-179.0	-178.9		-101.6
n CA	n C	p N	p CD		179.1	-172.8				3.2
n O	n C	p N	p CA		-168.6	-165.3				-167.0
n C	p N	p CA	p C	-76.5	-46.1	-46.6	-72.2	-72.1		-62.7
p N	p CA	p C	c N	-11.2	36.6	32.9	166.4	166.2		78.2
p N	p CA	p C	p OT2						164.1	164.1
p CA	p C	c N	c CA	-178.7	1.0	6.4	-177.9	-177.9		-105.4
p C	c N	c CA	c C		-34.5	-37.5	-179.9	-179.8		-107.9

c N	c CA	c C	c OT2				-172.9	-172.9		-172.9
n N	n CA	n CB	n CG			-65.8	-80.4	-80.3		-75.5
n CA	n CB	n CG	n CD1			176.6	-66.4	-66.7		14.5
n C	p N	p CA	p CB		-169.7	-169.8				-169.8
n C	p N	p CD	p CG	176.8	-168.4	179.1	170.6	172.5		106.1
p CD	p N	p CA	p C						107.3	107.3
p N	p CA	p CB	p CG	24.6	-35.3	-19.8	28.8	26.4	31.2	9.3
p CA	p CB	p CG	p CD	-32.2	35.1	15.1	-35.0	-31.6	-38.0	-14.4
p CB	p CG	p CD	p N	27.3	-21.3	-4.5	27.7	24.6	29.7	13.9
p CG	p CD	p N	p CA	12.6	-1.2	-8.8	-10.3	-8.7	-10.7	-8.7
p CD	p N	p CA	p CB	-7.5	23.1	18.2	-11.5	-11.0	-12.6	-0.2
c N	n C	* n CA	n CB			124.9				124.9
n N	n C	* n CA	n CB				122.1	122.1		122.1
n CD1	n CB	* n CG	n CD2			123.1	-120.8	-120.8		-39.5
p N	n CA	* n C	n O	179.5			-179.4	-179.4		-59.8
n C	p CA	* p N	p CD	-170.7	-167.2	-172.0	-179.2	-178.9		-173.6
p N	p C	* p CA	p CB	115.8	115.3	115.6	114.9	114.9	114.6	115.2
c N	p CA	* p C	p O	-179.6	179.8	-179.6	-179.7	-179.8		-107.8
p OT2	p CA	* p C	p OT1						178.1	178.1
c OT2	c CA	* c C	c OT1				-179.9	-179.9		-179.9

CHARMM dynamics rms fluctuations

Atom 1	Atom 2	Atom 3	Atom 4	apnma	cpg	cpl	lpg	lpg2	lpro	Aver.
n N	n CA				0.03	0.03	0.03	0.03		0.03
n CA	n C			0.03	0.03	0.03	0.03	0.03		0.03
n C	n O			0.02	0.03	0.02	0.02	0.02		0.02
n C	p N			0.03	0.04	0.04	0.03	0.03		0.03
n CA	n CB					0.03	0.03	0.03		0.03
n CB	n CG					0.03	0.03	0.03		0.03
n CG	n CD1					0.03	0.03	0.03		0.03
n CD2	n CG					0.03	0.03	0.03		0.03
p N	p CA			0.03	0.03	0.03	0.02	0.02	0.03	0.03
p CA	p C			0.03	0.03	0.03	0.03	0.03	0.03	0.03
p C	p O			0.02	0.02	0.02	0.02	0.02		0.02
p C	p OT1								0.02	0.02
p C	p OT2								0.02	0.02
p C	c N			0.03	0.03	0.03	0.02	0.02		0.03
p CA	p CB			0.03	0.03	0.03	0.03	0.03	0.04	0.03
p CB	p CG			0.03	0.03	0.03	0.03	0.05	0.03	0.03
p CG	p CD			0.03	0.03	0.03	0.03	0.03	0.03	0.03
p CD	p N			0.04	0.03	0.19	0.02	0.02	0.03	0.06
c N	c CA			0.03	0.03	0.03	0.02	0.02		0.03
c CA	c C				0.03	0.03	0.03	0.03		0.03
c C	c OT1						0.02	0.02		0.02
c C	c OT2						0.02	0.02		0.02
n N	n CA	n C			3.0	2.7	2.7	2.8		2.8
n CA	n C	n O		3.2			2.3	2.4		2.6
n CA	n C	p N		3.3	2.8	2.5	2.4	2.6		2.7
n O	n C	p N			3.0	2.8				2.9
n N	n CA	n CB				3.3	2.6	2.7		2.9
n CB	n CA	n C				3.1	2.6	2.7		2.8
n CA	n CB	n CG				3.0	2.5	2.7		2.7
n CB	n CG	n CD1				3.5	2.7	2.8		3.0
n CB	n CG	n CD2				3.5	2.7	2.8		3.0
n C	p N	p CA		2.6	2.8	2.6	2.2	2.3		2.5
n C	p N	p CD		2.7	2.7	3.6	2.3	2.4		2.7
p N	p CA	p C		3.2	2.9	2.8	2.6	2.7	3.0	2.9
p CA	p C	c N		2.7	2.3	2.2	2.2	2.3		2.3
p CB	p CA	p C		3.2	3.1	3.0	2.7	2.8	3.4	3.0
p N	p CA	p CB		2.4	2.5	2.5	2.0	2.0	2.6	2.3

p CB	p CG	p CD		2.8	2.7	2.6	2.3	2.8	2.7	2.7
p CG	p CD	p N		2.8	2.5	6.4	2.1	2.1	2.5	3.1
p CA	p CB	p CG		2.7	2.6	3.1	2.2	2.5	2.7	2.6
p CD	p N	p CA		2.3	2.3	3.8	1.7	1.8	2.2	2.4
p CA	p C	p O		2.7	2.7	2.6	2.2	2.3		2.5
p CA	p C	p OT1							2.7	2.7
p CA	p C	p OT2							2.8	2.8
p C	c N	c CA		3.3	2.8	2.7	2.7	2.6		2.8
c N	c CA	c C			3.0	2.7	2.8	2.8		2.8
c CA	c C	c OT1					2.4	2.5		2.4
c CA	c C	c OT2					2.4	2.4		2.4
n N	n CA	n C	p N		8.6	6.6	5.1	5.7		6.5
n CA	n C	p N	p CA	7.6	8.7	7.9	5.0	4.9		6.8
n CA	n C	p N	p CD		9.1	10.1				9.6
n O	n C	p N	p CA		9.3	8.5				8.9
n C	p N	p CA	p C	7.5	7.3	7.4	5.3	5.5		6.6
p N	p CA	p C	c N	10.0	6.3	5.8	6.1	6.1		6.9
p N	p CA	p C	p OT2						7.8	7.8
p CA	p C	c N	c CA	6.4	7.8	7.4	6.0	6.0		6.7
p C	c N	c CA	c C		8.7	7.6	6.8	7.2		7.6
c N	c CA	c C	c OT2				7.1	7.3		7.2
n N	n CA	n CB	n CG			8.5	5.3	5.7		6.5
n CA	n CB	n CG	n CD1			7.8	5.8	6.1		6.6
n C	p N	p CA	p CB		7.5	7.3				7.4
n C	p N	p CD	p CG	8.9	8.4	18.6	6.3	10.6		10.6
p CD	p N	p CA	p C						9.3	9.3
p N	p CA	p CB	p CG	10.0	5.9	21.3	5.5	12.7	7.5	10.5
p CA	p CB	p CG	p CD	11.4	7.1	28.9	5.7	16.3	7.7	12.8
p CB	p CG	p CD	p N	10.4	9.1	26.3	6.7	14.8	9.3	12.8
p CG	p CD	p N	p CA	8.7	10.1	16.2	7.8	10.1	10.6	10.6
p CD	p N	p CA	p CB	8.2	8.3	9.9	7.0	7.6	9.6	8.4
c N	n C	* n CA	n CB			4.6				4.6
n N	n C	* n CA	n CB				3.5	3.5		3.5
n CD1	n CB	* n CG	n CD2			4.8	3.8	3.7		4.1
p N	n CA	* n C	n O	5.0			3.9	3.9		4.3
n C	p CA	* p N	p CD	10.3	11.7	12.7	8.7	9.0		10.5
p N	p C	* p CA	p CB	3.7	4.1	4.0	2.9	3.0	3.5	3.5
c N	p CA	* p C	p O	4.7	5.2	5.2	4.1	3.9		4.6
p OT2	p CA	* p C	p OT1						5.7	5.7
c OT2	c CA	* c C	c OT1				5.0	5.0		5.0

CHARMM dynamics averaged - x-ray internal coordinates

Atom 1	Atom 2	Atom 3	Atom 4	apnma	cpg	cpl	lpg	lpg2	lpro	Aver.
n N	n CA				0.01	0.00	0.00	0.00		0.00
n CA	n C				-0.01	0.01	-0.01	-0.01		0.00
n C	n O			-0.02	0.00	0.00	0.00	0.00		0.00
n C	p N			-0.01	-0.04	-0.04	0.01	0.01		-0.02
n CA	n CB					0.01	0.02	0.02		0.02
n CB	n CG					0.01	0.01	0.01		0.01
n CG	n CD1					0.01	0.01	0.01		0.01
n CD2	n CG					0.01	0.01	0.01		0.01
p N	p CA			-0.01	-0.01	-0.01	0.00	0.00	-0.03	-0.01
p CA	p C			-0.01	0.00	0.00	-0.01	-0.01	-0.01	0.00
p C	p O			0.00	0.00	0.00	0.01	0.01		0.00
p C	p OT1								-0.01	-0.01
p C	p OT2								0.00	0.00
p C	c N			0.03	-0.01	-0.02	0.00	0.00		0.00
p CA	p CB			0.01		0.02	0.04	-0.09	-0.03	-0.01
p CB	p CG			0.03	0.01	0.05	-0.01	-0.03	-0.01	0.01
p CG	p CD			0.00	0.01	0.01	-0.02	0.00	0.00	0.00

p CD	p N			-0.04	-0.03	-0.22	-0.01	-0.01	0.02	-0.05
c N	c CA			-0.01	0.01	0.00	0.00	0.00		0.00
c CA	c C				-0.01	0.01	0.01	0.01		0.01
c C	c OT1						0.01	0.01		0.01
c C	c OT2						0.01	0.01		0.01
n N	n CA	n C			1.3	1.9	4.1	4.2		2.9
n CA	n C	n O		-3.8			-1.8	-1.8		-2.5
n CA	n C	p N		2.6	0.0	-0.3	3.9	4.0		2.0
n O	n C	p N			1.4	1.6				1.5
n N	n CA	n CB				1.0	-1.0	-1.0		-0.3
n CB	n CA	n C				0.1	-0.4	-0.4		-0.2
n CA	n CB	n CG				1.1	-1.1	-1.2		-0.4
n CB	n CG	n CD1				2.1	1.7	1.7		1.8
n CB	n CG	n CD2				1.8	-0.1	-0.1		0.5
n C	p N	p CA		1.1	1.3	0.9	1.9	1.9		1.4
n C	p N	p CD		-1.9	-1.8	-4.4	-1.8	-1.8		-2.3
p N	p CA	p C		-0.8	-1.0	-1.3	1.5	1.5	3.3	0.5
p CA	p C	c N		1.6	0.9	1.2	3.6	3.7		2.2
p CB	p CA	p C		-0.5	-1.0	-0.0	-5.4	7.4	0.7	0.2
p N	p CA	p CB		-0.1	-0.4	-0.2	-2.1	3.3	-1.9	-0.2
p CB	p CG	p CD		-1.3	-1.6	-3.4	-2.9	5.7	-0.7	-0.7
p CG	p CD	p N		2.5	1.6	8.2	3.6	0.9	-1.0	2.6
p CA	p CB	p CG		0.2	-0.2	0.8	3.2	2.2	3.9	1.7
p CD	p N	p CA		1.0	-0.4	2.9	-0.1	-0.1	2.3	0.9
p CA	p C	p O		0.8	-1.1	-1.4	-2.1	-2.1		-1.2
p CA	p C	p OT1							1.5	1.5
p CA	p C	p OT2							-3.1	-3.1
p C	c N	c CA		1.5	-0.9	-1.0	0.9	0.9		0.3
c N	c CA	c C			1.3	1.9	0.1	0.1		0.8
c CA	c C	c OT1					1.6	1.6		1.6
c CA	c C	c OT2					1.3	1.3		1.3
n N	n CA	n C	p N		-5.7	-8.3	-3.2	-3.2		-5.1
n CA	n C	p N	p CA	-2.3	6.3	10.0	5.5	5.5		5.0
n CA	n C	p N	p CD		0.5	3.1				1.8
n O	n C	p N	p CA		4.3	8.3				6.3
n C	p N	p CA	p C	-0.0	-2.7	-5.3	-5.4	-5.3		-3.8
p N	p CA	p C	c N	4.7	-1.6	-0.8	5.8	5.6		2.7
p N	p CA	p C	p OT2						-6.4	-6.4
p CA	p C	c N	c CA	-1.4	0.3	0.0	6.3	6.2		2.3
p C	c N	c CA	c C		2.9	4.1	-4.1	-4.0		-0.3
c N	c CA	c C	c OT2				7.4	7.5		7.5
n N	n CA	n CB	n CG			6.6	1.7	1.8		3.3
n CA	n CB	n CG	n CD1			-1.1	2.1	1.8		0.9
n C	p N	p CA	p CB		-0.8	-4.5				-2.7
n C	p N	p CD	p CG	0.1	0.7	-7.8	-0.0	-31.6		-7.7
p CD	p N	p CA	p C						2.1	2.1
p N	p CA	p CB	p CG	-2.4	-2.9	11.7	-3.0	57.3	-1.8	9.8
p CA	p CB	p CG	p CD	3.9	-0.3	-20.9	2.3	-77.5	2.4	-15.0
p CB	p CG	p CD	p N	-2.9	2.6	20.7	-0.1	68.6	-3.9	14.2
p CG	p CD	p N	p CA	1.1	-4.5	-14.1	-2.5	-34.5	1.6	-8.8
p CD	p N	p CA	p CB	0.4	4.7	2.3	4.0	-13.5	0.7	-0.2
c N	n C	* n CA	n CB			2.8				2.8
n N	n C	* n CA	n CB				1.1	1.1		1.1
n CD1	n CB	* n CG	n CD2			2.5	0.5	0.5		1.2
p N	n CA	* n C	n O	0.8			2.2	2.2		1.7
n C	p CA	* p N	p CD	-0.7	5.5	6.8	2.4	2.7		3.3
p N	p C	* p CA	p CB	-1.0	-1.9	-1.1	-5.4	8.7		-0.1
c N	p CA	* p C	p O	-0.7	0.1	-1.1	3.7	3.6		1.1
p OT2	p CA	* p C	p OT1						-4.5	-4.5
c OT2	c CA	* c C	c OT1				2.4	2.4		2.4

Table 11. Molecular dynamics of Antamanide: Average internal coordinates and fluctuations

Simulation A (+,+)[a]

	ϕ	ψ	χ_1	χ_2	χ_3	χ_4	χ_5	Impr[b]
Val 1	-74.4	151.2	-63.9					
	11.2	9.7	9.0					
Pro 2	-67.3	152.4	2.9	-2.2	0.7	1.1	-2.5	-175.6
	7.3	9.9	26.6	34.4	30.1	17.1	12.0	15.4
Pro 3	-83.4	12.5	31.1	-32.3	20.9	-1.4	-18.6	-176.4
	8.9	10.2	12.4	15.8	15.1	11.8	9.0	33.0
Ala 4	-86.1	-45.2						
	10.3	8.4						
Phe 5	72.8	-48.6	-62.0	-75.8				
	7.3	9.7	8.8	11.5				
Phe 6	-74.7	153.0	-72.2	-18.2				
	11.8	13.5	26.4	88.1				
Pro 7	-66.4	152.4	3.1	-3.2	1.9	-0.0	-2.0	-175.8
	7.3	9.2	26.6	34.4	30.2	17.1	12.0	8.3
Pro 8	-82.1	10.4	30.2	-31.6	20.8	-1.9	-17.7	-177.0
	8.6	9.7	13.7	17.6	16.5	12.1	9.1	11.7
Phe 9	-83.8	-49.2	-52.8	-46.5				
	9.8	7.8	9.2	62.2				
Phe 10	72.8	-45.4	-61.5	-76.8				
	7.1	9.6	8.5	11.5				

[a] (+,+) refers to signs of ϕ_5 and ϕ_{10} respectively. Average values of dihedrals appear on each line with amino acid sequence designation. RMS values appear below each dihedral.

[b] Impr=nitrogen improper dihedral: C_{i-1}-Cα-N-Cδ.

Simulation B (-,-)

	ϕ	ψ	χ_1	χ_2	χ_3	χ_4	χ_5	Impr
Val 1	-111.5	114.5	-39.2					
	14.5	17.7	149.4					
Pro 2	-67.5	143.9	-3.8	4.5	-3.6	1.3	1.6	-170.7
	8.8	10.0	27.5	34.0	28.7	15.5	13.7	2.0
Pro 3	-84.2	5.6	29.9	-30.8	19.7	-0.9	-18.1	-175.4
	9.3	9.8	14.1	17.5	16.1	11.9	9.7	2.3
Ala 4	-90.7	79.9						
	11.2	18.3						
Phe 5	-85.3	72.5	-68.1	-59.3				
	14.9	16.3	24.9	41.1				
Phe 6	-111.9	108.2	-63.6	8.5				
	13.9	17.5	10.5	90.3				
Pro 7	-65.5	142.8	-7.6	8.4	-6.0	1.3	3.9	-170.2
	8.7	10.2	26.9	33.2	28.2	15.8	13.6	9.1
Pro 8	-83.0	4.8	30.5	-33.1	22.7	-3.6	-16.8	-175.0
	8.9	9.6	11.4	14.3	13.8	11.3	9.1	5.3
Phe 9	-88.6	79.4	-57.0	46.9				
	9.7	15.4	9.7	87.5				
Phe 10	-84.9	71.7	-71.2	36.5				
	11.9	14.5	29.1	93.4				

Simulation C (+,-)

	ϕ	ψ	χ_1	χ_2	χ_3	χ_4	χ_5	Impr
Val 1	-124.4	150.0	-70.4					
	18.0	9.2	10.2					
Pro 2	-70.5	154.8	14.9	-15.6	10.2	-0.9	-8.8	-179.0
	7.3	12.2	23.0	29.7	26.2	15.6	11.3	10.2
Pro 3	-83.7	12.9	33.1	-33.3	20.3	0.6	-21.1	-179.1
	9.2	12.2	9.0	11.4	12.0	11.1	8.7	15.1
Ala 4	-85.2	-57.2						
	12.1	10.0						
Phe 5	66.7	-16.2	-60.3	-72.9				
	9.9	30.1	11.0	12.2				
Phe 6	-79.2	132.3	-134.5	-85.2				
	18.9	11.5	70.7	42.2				
Pro 7	-64.9	145.8	-3.5	1.4	1.2	-3.6	4.4	-170.5
	7.8	8.8	26.1	33.7	29.6	17.1	12.2	15.9
Pro 8	-83.0	7.3	31.4	-34.0	23.4	-3.8	-17.2	-175.5
	8.9	9.1	9.7	12.1	12.2	10.7	8.6	10.0
Phe 9	-89.8	70.6	-56.3	-69.5				
	9.1	8.6	9.5	14.9				
Phe 10	-76.4	48.3	-60.8	-70.1				
	8.3	17.3	9.5	14.3				

Simulation D (-,+)

	ϕ	ψ	χ_1	χ_2	χ_3	χ_4	χ_5	Impr
Val 1	-76.5	146.1	-65.0					
	13.2	12.6	9.6					
Pro 2	-66.0	147.0	-5.0	4.5	-2.3	-0.9	3.7	-170.4
	7.7	8.7	25.8	33.5	29.8	17.5	12.0	11.7
Pro 3	-83.3	5.1	29.2	-31.0	20.6	-2.3	-16.8	-174.8
	9.1	10.2	13.8	17.5	16.2	12.1	9.6	5.7
Ala 4	-96.1	70.1						
	11.3	9.2						
Phe 5	-76.9	50.1	-35.9	-73.1				
	8.5	14.5	45.8	31.9				
Phe 6	-147.9	156.3	44.8	-94.8				
	21.0	14.6	18.8	12.2				
Pro 7	-69.8	154.0	6.9	-8.1	6.2	-2.0	-3.1	-173.8
	7.5	9.5	24.6	32.6	29.4	17.8	11.3	0.2
Pro 8	-83.1	8.2	32.0	-34.3	23.3	-3.3	-17.9	-176.3
	9.0	10.2	9.0	11.1	11.5	10.6	8.5	4.8
Phe 9	-84.9	-57.8	-58.8	107.5				
	10.7	9.9	9.7	15.9				
Phe 10	71.7	-35.1	-60.8	-76.4				
	8.1	14.6	9.1	11.8				

Table 12. Antamanide molecular dynamics simulations: Average and RMS temperatures, energies

Simulation[a]	A(++)	B(--)	C(+-)	D(-+)
Ave. T	290.9	295.5	298.4	303.0
RMS T	14.6	14.0	15.0	15.2
Ave. Total E	472.8	474.6	477.6	478.4
RMS Total E	0.07	0.07	0.07	0.07
Ave. PE	356.6	356.6	358.4	357.4
RMS PE	5.9	6.0	6.0	6.1
Ave. KE	116.2	118.1	119.2	121.0
RMS KE	5.8	5.9	6.0	6.1

[a] Abbreviations: T: temperature; E: Energy; PE: Potential energy; KE: Kinetic energy. Temperature in degrees K and energies in kcal/mol.

Table 13. Antamanide molecular dynamics simulations: Cγ-Endo/Cγ-exo populations and correlation times

Simulation[a]	A(++)		B(--)		C(+-)		D(-+)	
	%	τ	%	τ	%	τ	%	τ
Pro 2								
endo	54	19.6	44	19.2	73	15.8	44	9.7
exo	46	16.3	56	23.9	27	5.8	56	12.7
τ_{ex}		8.9		10.6		4.2		5.5
Pro 3								
endo	96	75.8	94	128.1	97	71.7	94	81.3
exo	4	3.7	6	7.6	3	1.4	6	5.0
τ_{ex}		3.5		7.1		1.4		4.7
Pro 7								
endo	55	21.0	39	11.6	48	9.3	62	12.6
exo	45	17.0	61	18.1	52	9.9	38	7.7
τ_{ex}		9.4		7.1		4.8		4.8
Pro 8								
endo	94	59.4	96	131.2	98	132.6	98	155.7
exo	6	3.9	4	4.6	2	3.1	2	2.7
τ_{ex}		3.7		4.4		3.0		2.6

[a] Abbreviations: τ: residence times; τ_{ex}: exchange times; endo: Cγ-endo/Cβ-exo conformation (χ_2<0); exo: Cγ-exo/Cβ-endo conformation (χ_2>0). Times in picoseconds.

Table 14. Calculated energy differences between cis and trans AcProNHCH$_3$ as a function of dielectic constant and backbone dihedral ψ

ε[a]	E_{cis}	E_{trans}	ΔE	ψ_{cis}^{i}	ψ_{trans}^{i}	ψ_{cis}^{f}	ψ_{trans}^{f}
1	19.4	19.1	0.4	150	150	174.2	164.8
1	16.6	19.1	-2.5	0	150	-4.6	164.8
1	19.4	16.0	3.4	150	0	174.2	70.3
1	**16.6**	**16.0**	**0.6**	**0**	**0**	**-4.6**	**70.3**
2	18.3	18.1	0.3	150	150	172.6	168.1
2	17.5	18.1	-0.6	0	150	-3.5	168.1
2	18.3	18.0	0.3	150	0	172.6	4.1
2	**17.5**	**18.0**	**-0.5**	**0**	**0**	**-3.5**	**4.1**
5	**17.4**	**17.2**	**0.3**	**150**	**150**	**171.6**	**170.0**
5	17.9	17.2	0.7	0	150	-2.7	170.0
5	17.4	17.9	-0.5	150	0	171.6	0.6
5	17.9	17.9	0.0	0	0	-2.7	0.6
20	**16.9**	**16.6**	**0.3**	**150**	**150**	**171.1**	**171.0**
20	18.0	16.6	1.4	0	150	-2.3	171.0
20	16.9	17.8	-0.9	150	0	171.1	-0.8
20	18.1	17.8	0.3	0	0	-2.3	-0.8
50	**16.8**	**16.5**	**0.3**	**150**	**150**	**171.0**	**171.2**
50	18.1	16.5	1.5	0	150	-2.2	171.2
50	16.8	17.7	-0.9	150	0	171.0	-1.0
50	18.1	17.7	0.3	0	0	-2.2	-1.0
78	**16.8**	**16.5**	**0.3**	**150**	**150**	**171.0**	**171.2**
78	18.1	16.5	1.6	0	150	-2.1	171.2
78	16.8	17.7	-1.0	150	0	171.0	-1.1
78	18.1	17.7	0.4	0	0	-2.1	-1.1
∞	**16.7**	**16.5**	**0.3**	**150**	**150**	**171.0**	**171.3**
∞	18.1	16.5	1.6	0	150	-2.1	171.3
∞	16.7	17.7	-1.0	150	0	171.0	-1.2
∞	18.1	17.7	0.4	0	0	-2.1	-1.2

[a] Abbreviations: ε: dielectric constant; E_{trans}: energy of trans AcProNHCH$_3$; E_{cis}: energy of cis AcProNHCH$_3$; ΔE: cis - trans energies; ψ_{trans}^{i}: value of ψ at start of trans AcProNHCH$_3$ minimization; ψ_{cis}^{i}: value of ψ at start of cis AcProNHCH$_3$ minimization; ψ_{trans}^{f}: value of ψ at end of trans AcProNHCH$_3$ minimization; ψ_{cis}^{f}: value of ψ at end of cis AcProNHCH$_3$ minimization. Calculations with lowest energy of trans and lowest energy of cis AcProNHCH$_3$ for given value of dielectric constant are printed in bold type.

Table 15. CHARMM vs. experimental frequencies for proline-containing peptides

Exp.	IR/ Raman	Freq.	Assignment[d]	CHARMM Freq.	Assignment[d]
PLP II[a]	IR	100	N-Y rck; Y-N tor	85	Y-N tor; ring tor'
	IR	195	ring tor; BAC def	201	ring tor; ring tor'; BAC def
	IR	210	Y-N tor; ring tor	244	N-Y rck; OYN def; ring tor
	IR	260	NAC def; CA-C tor	288	NAC def; N-Y rck
	Raman	285	A-C tor; NAC def; ring tor	288	NAC def; N-Y rck
	IR	314	NAC def; N-Y rck	319	BAC def; ring def; Y-N str; N-Y rck
	Raman	314		319	BAC def; ring def; Y-N str; N-Y rck
	IR	493,495	OYN def; NAC def; N-Y rck	443	N-Y rck; OYN def
	IR	870	B-G str; A-C str	873	HB/HD rck; A-C str
	IR	917	B-G str; NAC def	884	B-G str; G-D str
	IR	974	G-D str; B-G str	980	HG rck; A-B str; B-G str
	Raman	1000	G-D str; B-G str	987	Y def; HG wag; G-D str
	IR	1093	A-B str; A-C str	1016	G-D str; HB rck; B-G str
PLP I[b]	IR	350	Y-N str	320	BAC def; ring def; Y-N str
	Raman	363			
Pro[c]	IR	901, 985	ring str	884, 980	B-G str; G-D str
Pyrrol.[c]	IR	902	ring str	884	B-G str; G-D str

[a] Polyproline II. Ref. 75

[b] Polyproline I. Ref. 74

[c] Proline and pyrrolidine. Ref. 76

[d] Assignment abbreviations are the same as in Table 6. Frequencies in cm^{-1}.

Appendix: CHARMM Parameters for Proline

Bonds

Atom 1	Atom 2	K_b	b_o
C	N	260.0	1.300
C	NH1	370.0	1.345
CC	NH2	430.0	1.360
C	O	620.0	1.230
CC	O	650.0	1.230
CP1	C	250.0	1.490
CP1	CC	250.0	1.490
CT3	C	250.0	1.490
CP1	CP2	222.5	1.527
CP2	CP2	222.5	1.537
CP2	CP3	222.5	1.537
CP1	HB	330.0	1.080
CP2	HA	309.0	1.111
CP3	HA	309.0	1.111
CT3	HA	322.0	1.111
N	CP1	320.0	1.434
N	CP3	320.0	1.455
NP	CP1	320.0	1.485
NP	CP3	320.0	1.502
NH1	CT3	320.0	1.430
NH1	H	440.0	0.997
NH2	H	480.0	1.000
NP	HC	460.0	1.006

Angles

Atom 1	Atom 2	Atom 3	K_θ	θ_o	K_{ub}	s_o
CP2	CP1	C	52.00	112.30		
CP2	CP1	CC	52.00	112.30		
CP1	CP2	CP2	70.00	108.50		
CP2	CP2	CP3	70.00	108.50		
CP3	NP	CP1	100.00	111.00		
CP3	N	CP1	100.00	114.20		
N	CP1	CP2	70.00	110.80		
N	CP3	CP2	70.00	110.50		
N	CP1	C	50.00	108.20		
N	CP1	CC	50.00	108.20		
NP	CP1	C	50.00	106.00		
NP	CP1	CC	50.00	106.00		
NP	CP1	CP2	70.00	108.50		
NP	CP3	CP2	70.00	108.50		
C	N	CP3	60.00	117.00		
CT3	C	N	20.00	112.50		
CP1	CC	NH2	80.00	112.50		
CP1	C	NH1	80.00	116.50		
O	C	N	80.00	122.50		
O	C	NH1	80.00	122.50		
O	CC	NH2	75.00	122.50	50.0	2.370
O	C	CP1	80.00	118.00		
O	CC	CP1	80.00	118.00		
O	C	CT3	80.00	121.00		
H	NH2	H	23.00	120.00		
HC	NH2	HC	39.00	106.50		
H	NH2	CC	50.00	120.00		
HC	NP	HC	51.00	107.50		
HB	CP1	N	48.00	112.00		
HB	CP1	NP	51.50	107.50		
HA	CP3	N	48.00	108.00		
HA	CP3	NP	51.50	109.15		
HA	CT3	NH1	51.50	109.50		
HA	CP2	CP2	26.50	110.10	22.53	2.179
HA	CP2	CP3	26.50	110.10	22.53	2.179
HA	CP3	CP2	26.50	110.10	22.53	2.179
HA	CP2	CP1	26.50	110.10	22.53	2.179
HB	CP1	CP2	35.00	118.00		
HA	CT3	C	33.00	109.50	30.00	2.163
HB	CP1	C	50.00	112.00		
HB	CP1	CC	50.00	112.00		
HC	NP	CP1	33.00	109.50	4.00	2.056
HC	NP	CP3	33.00	109.50	4.00	2.056
H	NH1	CT3	35.00	117.00		
H	NH1	C	34.00	123.00		
HA	CT3	HA	35.50	108.40	5.40	1.802
HA	CP2	HA	35.50	109.00	5.40	1.802
HA	CP3	HA	35.50	109.00	5.40	1.802

Dihedrals

Atom 1	Atom 2	Atom 3	Atom 4	K_ϕ	n	δ
CT3	C	N	CP1	0.30	4	0.0
CT3	C	N	CP3	0.30	4	0.0
O	C	N	CP1	0.30	4	0.0
O	C	N	CP3	0.30	4	0.0
CT3	C	N	CP1	2.75	2	180.0
CT3	C	N	CP3	2.75	2	180.0
O	C	N	CP1	2.75	2	180.0
O	C	N	CP3	2.75	2	180.0
CP1	C	NH1	CT3	1.60	1	0.0
CP1	C	NH1	CT3	2.50	2	180.0
CP1	C	NH1	H	2.50	2	180.0
O	C	NH1	CT3	2.50	2	180.0
O	C	NH1	H	2.50	2	180.0
CP1	CC	NH2	H	2.50	2	180.0
O	CC	NH2	H	1.40	2	180.0
C	N	CP1	C	0.80	3	0.0
C	N	CP1	CC	0.80	3	0.0
C	N	CP1	CP2	0.80	3	0.0
C	N	CP1	HB	0.80	3	0.0
CP3	N	CP1	C	0.10	3	0.0
CP3	N	CP1	CC	0.10	3	0.0
CP3	N	CP1	CP2	0.10	3	0.0
CP3	N	CP1	HB	0.10	3	0.0
CP3	NP	CP1	C	0.08	3	0.0
CP3	NP	CP1	CC	0.08	3	0.0
CP3	NP	CP1	CP2	0.08	3	0.0
CP3	NP	CP1	HB	0.08	3	0.0
HC	NP	CP1	C	0.08	3	0.0
HC	NP	CP1	CC	0.08	3	0.0
HC	NP	CP1	CP2	0.08	3	0.0
HC	NP	CP1	HB	0.08	3	0.0
N	CP1	C	NH1	0.30	1	0.0
N	CP1	CC	NH2	0.30	1	0.0
CP2	CP1	C	NH1	0.40	1	0.0
CP2	CP1	CC	NH2	0.40	1	0.0
CP2	CP1	C	O	0.40	1	180.0
CP2	CP1	CC	O	0.40	1	180.0
HB	CP1	C	NH1	0.40	1	180.0
HB	CP1	CC	NH2	0.40	1	180.0
HB	CP1	C	O	0.40	1	0.0
HB	CP1	CC	O	0.40	1	0.0
NP	CP1	CC	NH2	0.40	1	0.0
CP2	CP1	C	NH1	0.60	2	0.0
CP2	CP1	CC	NH2	0.60	2	0.0
CP2	CP1	C	O	0.60	2	0.0
CP2	CP1	CC	O	0.60	2	0.0
HB	CP1	C	NH1	0.60	2	0.0
HB	CP1	CC	NH2	0.60	2	0.0
HB	CP1	C	O	0.60	2	0.0
HB	CP1	CC	O	0.60	2	0.0
N	CP1	C	NH1	-0.30	4	0.0
N	CP1	CC	NH2	-0.30	4	0.0
N	CP1	C	O	-0.30	4	0.0
N	CP1	CC	O	-0.30	4	0.0

X	CP1	CP2	X	0.14	3	0.0
X	CP2	CP2	X	0.16	3	0.0
X	CP2	CP3	X	0.14	3	0.0
CP2	CP3	N	CP1	0.10	3	0.0
HA	CP3	N	CP1	0.10	3	0.0
CP2	CP3	NP	CP1	0.08	3	0.0
CP2	CP3	NP	HC	0.08	3	0.0
HA	CP3	NP	CP1	0.08	3	0.0
HA	CP3	NP	HC	0.08	3	0.0

Improper Dihedrals

Atom 1	Atom 2	Atom 3	Atom 4	K_ω	ω_0
NH1	X	X	H	20.00	0.00
NH2	X	X	H	4.00	0.00
O	CP1	NH2	CC	45.00	0.00
O	NH2	CP1	CC	45.00	0.00
O	X	X	C	120.00	0.00

Improper Dihedrals Used on Model Compounds (by Atom Names)

$AcProNH_2$			
CY	CAY	N	OY
C	CA	NT	O
C	NT	CA	O
NT	C	HT2	HT1
$ProNH_2$			
C	CA	NT	O
C	NT	CA	O
NT	C	HT2	HT1
$AcProNHCH_3$			
CY	CAY	N	OY
C	CA	NT	O
C	NT	CA	O
NT	C	HT2	HT1
NT	C	CAT	HNT

van der Waals Parameters

Atom	ε_{ij}	$Rmin_{ij}$	ε_{ij} (1,4)	$Rmin_{ij}$(1,4)
C	-0.110	2.0000		
CC	-0.070	2.0000		
CP1	-0.020	2.2750	-0.0100	1.9000
CP2	-0.055	2.1750	-0.0100	1.9000
CP3	-0.055	2.1750	-0.0100	1.9000
CT3	-0.080	2.0600	-0.0100	1.9000
H	-0.046	0.2245		
HA	-0.022	1.3200		
HB	-0.022	1.3200		
HC	-0.046	0.2245		
N	-0.200	1.8500	-0.0001	1.8500
NH1	-0.200	1.8500	-0.2000	1.5500
NH2	-0.200	1.8500		
NP	-0.200	1.8500		
O	-0.120	1.7000	-0.1200	1.4000

Charges on Model Compounds $AcProNH_2$, $AcProNHCH_3$, and $ProNH_2$

$AcProNH_2$			$AcProNHCH_3$			$ProNH_2$		
Name	**Type**	**Charge**	**Name**	**Type**	**Charge**	**Name**	**Type**	**Charge**
CAY	CT3	-0.27	CAY	CT3	-0.27	HN1	HC	0.24
HY1	HA	0.09	HY1	HA	0.09	HN2	HC	0.24
HY2	HA	0.09	HY2	HA	0.09			
HY3	HA	0.09	HY3	HA	0.09			
CY	C	0.51	CY	C	0.51			
OY	O	-0.51	OY	O	-0.51			
N	N	-0.29	N	N	-0.29	N	NP	-0.07
CA	CP1	0.02	CA	CP1	0.02	CA	CP1	0.16
CB	CP2	-0.18	CB	CP2	-0.18	CB	CP2	-0.18
CG	CP2	-0.18	CG	CP2	-0.18	CG	CP2	-0.18
CD	CP3	0.00	CD	CP3	0.00	CD	CP3	0.16
HA	HB	0.09	HA	HB	0.09	HA	HB	0.09
HB1	HA	0.09	HB1	HA	0.09	HB1	HA	0.09
HB2	HA	0.09	HB2	HA	0.09	HB2	HA	0.09
HG1	HA	0.09	HG1	HA	0.09	HG1	HA	0.09
HG2	HA	0.09	HG2	HA	0.09	HG2	HA	0.09
HD1	HA	0.09	HD1	HA	0.09	HD1	HA	0.09
HD2	HA	0.09	HD2	HA	0.09	HD2	HA	0.09
C	CC	0.51	C	C	0.51	C	CC	0.51
O	O	-0.51	O	O	-0.51	O	O	-0.51
NT	NH2	-0.62	NT	NH1	-0.47	NT	NH2	-0.62
HT1	H	0.31	HNT	H	0.31	HT1	H	0.31
HT2	H	0.31	CAT	CT3	-0.11	HT2	H	0.31
			HT1	HA	0.09			
			HT2	HA	0.09			
			HT3	HA	0.09			

Chapter 2

Catalysis of X-Pro peptide bond cis-trans isomerization: *Ab initio* calculations on the effect of ammonium ion hydrogen bonded to amide nitrogen

Abstract

Ab initio calculations have been performed to test the influence of positively charged sidechains on the cis-trans isomerization of the X-Pro peptide bond in proteins. This isomerization is catalyzed by two families of enzymes — the cyclophilins and FK506 binding proteins. The recently solved x-ray structure of cyclophilin complexed with a tetrapeptide revealed the presence of an arginine and a histidine sidechain within hydrogen bonding distance of the proline ring nitrogen. Electronic structure arguments can be used to predict that the presence of a positively charged hydrogen bond donor on proline nitrogen will reduce the barrier to rotation by withdrawing charge from the peptide bond and reducing its double bond character. The prediction is confirmed by *ab initio* calculations on formamide and N,N-dimethyl acetamide that show that the barrier can be reduced from 13-18 kcal/mol in the uncomplexed molecules to free rotation in the presence of ammonium ion. These results are used to rationalize the catalytic activity of cyclophilin as well as recent evidence from site-directed mutagenesis [1]that dihydrofolate reductase uses its Arg-44 sidechain to self-catalyze the cis-trans isomerization of Pro-66 during protein folding.

I. Introduction

Many proteins have been found to have one or more slow refolding pathways most likely caused by the cis-trans isomerization of the peptide bond preceding proline residues [2] (reviewed by Schmid [3]). In unfolded proteins, prolines are predominantly trans [4]. Proteins with cis prolines therefore require proline isomerization in order to fold properly. Proteins with such slow-refolding pathways include ribonuclease A [5, 6, 7], ribonuclease T1 [8], chymotrypsinogen [9], cytochrome *c* [10, 11], β-lactoglobulin [12], metmyoglobin [13], *E. coli* thioredoxin [14], T4 thioredoxin [15], carp parvalubumin [16], and bovine carbonic anhydrase B [17]. In some instances, the slow folding pathways have been eliminated by mutagenesis of particular proline residues, eliminating the need for isomerization about the peptide bond. Examples include iso-1-cytochrome *c* [18], thioredoxin [19], iso-2-cytochrome *c* [20, 21], calbindin D_{9k} [22], and ribonuclease T1 [23].

The cis-trans isomerization free energy barrier has been found to be approximately 20 kcal/mol in peptides [24, 25, 26, 27, 28], proteins [2], and in formamide, N-methyl formamide, N,N-dimethyl formamide, acetamide, N-methyl acetamide, and N,N-dimethyl acetamide [29, 30, 31, 32]. The large barrier and presence of proline in many protein sequences have led to speculation about the role of catalytic activity in protein folding by such cellular proteins as cyclophilin and FKBP [33] or by other means, such as catalysis by other functional groups in the folding protein [1, 34]. In this paper, we consider one such catalytic mechanism – hydrogen bonding by positively charged ionic groups to the imide nitrogen of proline. But first we review what is known about the catalyzed and uncatalyzed isomerizations.

Harrison and Stein have measured $\Delta H^{\ddagger}$ for a number of proline-containing tripeptides and tetrapeptides in water (pH 7.8), and found an average value of 20 ± 1 kcal/mol and an average $-T\Delta S^{\ddagger}$ of -0.3 ± 1.5 kcal/mol. One exception found was the

peptide Gly-Gly-Lys-Phe-Pro, which had a $\Delta H^{\ddagger}$ of 16.1 kcal/mol, and a $-T\Delta S^{\ddagger}$ of 5.6 kcal/mol [26]. The effects of solvent, salt, temperature, and pH on isomerization of C-N amide bonds have been studied [24, 27, 35]. Eberhardt et al. [27] measured $\Delta G^{\ddagger}$ of several Pro-containing peptides in protic and aprotic solvents, and found that $\Delta G^{\ddagger}$ in aprotic solvents such as dioxane, benzene, and toluene, was 1-2 kcal/mol less than in protic solvents, such as alcohols and water. There was little correlation with solvent dielectric constant alone.

Cis-trans isomerization can be catalyzed in acidic solution by protonation on the ring nitrogen or hydrogen bonding by positively charged solvent molecules to the ring nitrogen [24, 35]. $\Delta G^{\ddagger}$ and $\Delta H^{\ddagger}$ for the cis/trans isomerization of N,N dimethyl acetamide in water is lowered from 19.3 and 19.0 kcal/mol to 16.3 and 16.4 kcal/mol on changing pH from 7 to 1.8 [24]. Berger et al. [29] found in NMR experiments that on acidifying water solutions of N,N-dimethyl acetamide the C-N peptide bond rotates freely. They argue that protonation occurs mostly on the carbonyl oxygen, but also occurs appreciably on nitrogen, leading to free rotation of the amide C-N bond. From temperature-dependent NMR experiments, Fraenkel and Franconi [30] calculated an activation energy of 7.3 kcal/mol for N,N-dimethyl acetamide in 0.1 M HCl (pH 2).

A number of enzymes catalyze the conversion of cis to trans proline and vice versa. These fall into two broad groups: cyclophilins which bind the immunosuppressant cyclosporin A (CsA) and proteins having homology to the human FK506 binding protein (FKBP), which binds the immunosuppressants rapamycin (for a review see Ref. [36]) and FK506 (for reviews, see Refs. [37, 38, 39, 40, 41]). Both sets of proteins have been found in a large number of species [42], as diverse as humans [43], cows [44], *Neurospora crassa* [45, 46], *Escherichia coli* [47] and *Saccharomyces cerevisiae* [45, 48]. In *in vitro* experiments, cyclophilin has been found to catalyze the refolding of denatured Type III collagen and ribonuclease A [49], the immunoglobulin light chain, the S-protein fragment of bovine RNase A, and porcine pancreatic RNase [50], RNase T1

[42], and carbonic anhydrase [51]. FKBP has been shown to catalyze the folding of RNase T1 [46]. Their exact roles in cellular physiology are still unknown. Yeast cells can survive without cyclophilin and yeast FKBP [45, 52, 53], but are highly sensitive to the effects of cyclophilin and FK506 when the proteins are present. FKBP has been found to have activity not only as a cis-trans peptidyl-proline isomerase (PPIase) [54], but also as a folding chaperone, preventing the aggregation of carbonic anhydrase at an early stage of folding, while PPIase activity was found to take place at a later stage [51]. In human T-cells, cyclophilin binds cyclosporin A and inhibits T cell activation at a concentration much lower than that which inhibits PPIase activity. Similarly, rapamycin and FK506 prevent T-cell activation by mitogens at concentrations below that which inhibits the PPIase activity of FKBP. While initially it was thought that inhibition of PPIase activity was directly related to the immunosuppressive effects of these drugs, it was later found that both the cyclophilin/cyclosporin A and the FKBP/FK506 complexes bind to the human T-cell phosphatase calcineurin [37, 55, 56], indicating that another mechanism may be at work. Neither the proteins alone nor the uncomplexed drugs bind to the phosphatase, indicating that only the complexes contain the structure necessary for binding to and inhibiting the activity of calcineurin. The targets of calcineurin are still unknown, but are likely to be phosphorylated proteins which interact either directly or indirectly with gene activating sequences in chromosomal DNA [39, 40]. Both FK506 and cyclosporin A prevent the nuclear translocation of a cytoplasmic portion of NF-AT, a transcription factor active in T-cell activation [57].

The mechanism of PPIase activity of cyclophilin and FKBP has been studied by a number of groups [28, 33, 58, 59, 60, 61, 62], and the structures of cyclophilin complexed with CsA and a tetrapeptide have been solved by x-ray crystallography and NMR [63]. The structures of FKBP [64, 65] and FKBP complexed with FK506 [66] have also been solved. For both proteins, nucleophilic catalysis by sulfhydryl and other groups has been ruled out by mutagenesis of nearby residues that might be able to attack

the carbonyl carbon of the amide bond [61, 67]. The substrate binding site of FKBP is quite hydrophobic, containing a number of aromatic sidechains. The site contains few hydrogen-bonding groups, and these have been mutated to non-hydrogen bonding sidechains and found to be unimportant in catalysis [61]. FKBP catalyzes isomerization in peptides with hydrophobic sidechains in the residue C-terminal to proline, while having little effect on substrates with charged or polar sidechains. One hypothesis is that FKBP catalyzes cis-trans isomerization of peptide-proline amide bonds by binding more tightly to twisted amide groups than planar cis or trans amide groups, thereby stabilizing the transition state of cis-trans isomerization [59, 60, 61]. The hydrophobic environment has been suggested to lower the activation barrier in analogy to the lowering of isomerization rates in non-polar solvents compared to water. Eberhardt et al. [27] have measure this change to be at most 1-2 kcal/mol. The peptide binding site of cyclophilin is less restrictive to the identity of the residue preceding proline, and contains a number of hydrogen bonding sidechains including two positively charged residues – His 126 and Arg 55.

Harrison and Stein [33] have measured the thermodynamics of catalyzed and uncatalyzed cis-trans isomerization in the proline containing peptides Suc-Ala-Xaa-Pro-Phe-para-nitroaniline, where Xaa was Gly, Ala, Leu, or Trp. As described above, for the uncatalyzed reaction, an average $\Delta H^{\ddagger}$ over these four peptides in water (pH 7.8) was found to be 20 ± 1 kcal/mol and $-T\Delta S^{\ddagger}$ was -0.3 ± 1.5 kcal/mol. For the isomerization of these peptides catalyzed by cyclophilin, values for $\Delta H^{\ddagger}$ (in kcal/mol) of 3.2 (Xaa=Gly), 4.3 (Ala), Trp (7.5), and 7.9 (Leu) were measured. Values for $-T\Delta S^{\ddagger}$ (in kcal/mol) were 13.3 (Gly), 11.6 (Ala), 9.6 (Trp), and 8.2 (Leu). This results in a lowering of $\Delta G^{\ddagger}$ over the uncatalyzed reaction of only 3.1 ± 0.4 kcal/mol across the four peptides, because of compensation between the enthalpy and entropy terms of free energy. The results for FKBP differ from cyclophilin by having

substantially larger $\Delta H^{\ddagger}$ values (in kcal/mol: 12 (Xaa=Ala), 11 (Trp), and 15 (Leu)) and smaller $-T\Delta S^{\ddagger}$ values (in kcal/mol: 5.9 (Ala), 7.1 (Trp), and 2.4 (Leu)) [33] .

Texter et al. [1] have shown evidence from mutagenesis experiments that an arginine residue (Arg-44) in dihydrofolate reductase (DHFR) self-catalyzes the cis-trans isomerization of Pro-66 and therefore the folding of DHFR's which initially have the wrong cis-trans isomer of Pro-66. In the crystal structure of DHFR, the arginine sidechain NH_2 is 2.99Å from the proline ring nitrogen and 3.01Å from the carbonyl oxygen. From electronic structure arguments, it would seem that the hydrogen bond to nitrogen would lower the barrier to isomerization but the bond to the oxygen would not. This evidence indicates that positively charged ions can hydrogen bond to proline nitrogen and significantly lower the barrier.

The uncatalyzed isomerization of formamide has been studied by *ab initio* methods [68, 69, 70, 71]. From 6-31g* and 6-31g** calculations with MP2 and MP3 electron correlation, Wiberg et al. argue that the conventional picture of amide resonance, shown in Figure 1, involving a transfer of charge from oxygen to nitrogen is inaccurate. Instead, they find a transfer of charge from nitrogen to carbon during isomerization from the planar to the saddle point structures. The movement of electronic charge can be described as a transfer of electrons from π orbitals on nitrogen to σ orbitals on carbon [71].

Besides the experimental information described above, little is known however about the effects of hydrogen bonding on the isomerization process of proline residues. Protonation of the amide nitrogen results in free rotation in acidic solutions of DMA [24, 29, 30], but the presence of hydrogen bonding to the amide nitrogen from a direction approximately perpendicular to the plane determined by nitrogen and its substituents could also lower the barrier to rotation by decreasing amide resonance. The hydrogen bond donor might be the NH group of the succeeding residue or hydrogen bonding sidechains elsewhere in the peptide or protein. One might expect a stronger effect from

positively charged sidechains, such as the Gly-Gly-Lys-Phe-Pro peptide studied by Lin and Brandts [26].

In this paper, we present *ab initio* calculations on the isomerization of formamide and N,N-dimethyl acetamide at the 3-21g and 6-31g* levels, both alone and with an ammonium ion hydrogen bonded to the amide bond nitrogen. These molecules are used as models for proline, since they contain all of the atoms likely to be involved electronically in the isomerization. The planar conformations are compared with saddle point conformers on the isomerization energy surface. The calculations demonstrate that in the presence of ammonium ion, the barrier is reduced from 13-18 kcal/mol to nearly free rotation. We discuss the relevance of these results on the isomerization of proline in proteins and solvent, the self-catalysis of protein folding by positively charged residues, and the catalytic mechanism of cyclophilin and FKBP.

II. Calculations

Ab initio calculations were performed with the Gaussian 88 [72] and Gaussian 90 [73] programs at the 3-21g and 6-31g* Hartree-Fock levels. In all of the *ab initio* calculations reported, the structures were fully minimized with the Berny minimizer to the default tolerances specified in Gaussian 88 and Gaussian 90. These consist of a maximum force limit of 0.00045 mdyne/Å, a maximum RMS force of 0.00030 mdyne/Å, a maximum displacement per step of 0.0018 Å, and a maximum RMS displacement per step of 0.0012 Å. The chemical structures of formamide and N,N-dimethyl acetamide are shown in Figure 2, along with the atom names used in this paper. The structure of N-acetylproline-N′-methylamide is shown for comparison.

Locating saddle points on potential energy surfaces can be quite difficult in systems with more than one rotatable group [74]. To locate the cis-trans saddle points for N,N-dimethyl acetamide in both *ab initio* and CHARMM calculations, we tried initially to fix the dihedral CAY-CY-N-CA (see Figure 1 for definition of atom names). This

dihedral is the same as the peptide bond dihedral ω defined for amino acids in proteins. When the dihedral was fixed at successive values from 180° to 90° the energy increased as expected and the nitrogen center became tetrahedral. But beyond 90°, the energy fell precipitously as the nitrogen center inverted (cf. inversion of ammonia). In another CHARMM calculation, we used the routine "TRAVEL" which employs a conjugate peak refinement method to locate saddle points between defined end-points of isomerization [75]. This calculation resulted in a saddlepoint of nearly ideal tetrahedral geometry with ω equal to 120° without the inversion about the nitrogen center. Steepest descent calculations from the saddlepoint defined a simple path from cis to trans (or vice versa) that did not include an inversion of the tetrahedral center near the transition state.

Since we wanted to perform *ab initio* calculations on amides and since the Gaussian programs have not implemented the conjugate peak refinement method or any similar procedure, we were forced to choose another single variable which when varied would change the structure smoothly from cis to the saddlepoint to trans without an inversion of the nitrogen center. Fischer [74] defined a "virtual dihedral" angle ζ CAY-OY-CD-CA that was found to be orthogonal (or nearly so) to the nitrogen inversion coordinate (the improper dihedral about the nitrogen center – η or CY-CA-N-CD). This meant that at points near the saddlepoint, fixing ζ alone resulted in structures that were in a deep potential energy well with respect to η without the possibility of inversion (which would require variation of ζ which was not allowed). This was in contrast to using ω alone, where an inversion could take place without changing the value of ω. In preliminary calculations, the saddlepoint was found to be located at values of ζ between 80° and 90° (or between -80° and -90°).

In this paper we are interested in the isomerization of formamide and dimethyl formamide about the C-N bond. We first calculated the planar minimum structures at 3-21g and 6-31g* by optimizing the structures with no constraints. To locate the two cis-trans isomerization saddle point structures (through ζ=+90° and -90°), we fixed the

virtual dihedrals (ζ) HY-OY-HD-HA in formamide and CAY-OY-CD-CA in dimethyl acetamide to 90° and -90°, with starting values of the improper dihedral η about nitrogen (CY-HA-N-HA or CY-CA-N-CD) near -120°. The structures were then fully minimized at 3-21g and 6-31g* using the Gaussian program.

We then calculated the structures of a complex of formamide and ammonium ion and a complex of N,N-dimethyl acetamide and ammonium ion. A major obstacle to these calculations was found to be that during optimization, an ammonium ion in a hydrogen-bonded geometry to the formamide (or dimethyl acetamide) nitrogen moved toward the carbonyl oxygen, which is a stronger hydrogen bond acceptor than the amide nitrogen. A hydrogen bond to oxygen is likely to increase the barrier to cis-trans isomerization by shifting electrons into the peptide bond, while a hydrogen bond to nitrogen is likely to lower it by rehybridizing the π and sp^2 orbitals on nitrogen to sp^3-like orbitals. This results in a C-N peptide bond with more σ character and less π character, i.e. essentially a single bond. To keep the ammonium ion on nitrogen, a number of constraints were tried. Rather than fixing the nitrogen-nitrogen distance or nitrogen-hydrogen distance, the angles H1-N-HA and NA-N-HD angles were required to be equal (i.e. the hydrogen bonds consists of ammonium atoms NA and H1 and the amide N atom; HA and HD are part of the amide molecule). This produced an energy minimum with a hydrogen bond between the ammonium nitrogen and hydrogen and the formamide nitrogen. This was effective for the ζ=-90° and ζ=90° structures of both formamide and dimethyl acetamide, as well as the 3-21g optimization of the planar molecules. However, for the 6-31g* minimization of the planar molecules, the pull on the ammonium by the oxygen was still too large, and the ammonium drifted away from the nitrogen toward the carbonyl oxygen. Rather than fixing the distance between the two molecules, we fixed the angle H1-N-HA (=NA-N-HD) to 108°, which is approximately the angle of a tetrahedral arrangement of the nitrogen lone pair and was close to the value found in the 3-21g minimization of the formamide/NH_4^+ complex. Movement of the NH_4^+ ion away

from a hydrogen bonding geometry to the amide N would require a large change in this angle away from 108°. This procedure was successful in producing a fully optimized hydrogen-bonded structure in the context of the constraints applied.

III. Results

A. Formamide without ligands

To calculate the barrier to isomerization about the N-C bond in formamide, we constrained the virtual dihedral ζ (HY-OY-HD-HA) to -90° and +90°, and minimized the energy of formamide with Gaussian 90 with the 3-21g and 6-31g* basis sets. The results are listed in Table 1 along with the planar conformations at 3-21g and 6-31g*.

The planar molecule at 3-21g is essentially flat with the improper dihedral angle about nitrogen (η) equal to 180.00°. At 6-31g*, there is a very small pyramidization with η equal to -178.37°. Apparently the addition of d orbitals affects the relationship of nitrogen to its substituents.

Because of the symmetry of formamide, there are only two saddle point structures in the isomerization about the N-CY bond. These can be described as having ζ = -90° and 90° with values of the improper dihedral about nitrogen (CY-HA-N-HD) consistent with a tetrahedral geometry (approximately -132°). The same structures can also be described with {ζ,η} values of {±90°,+120°} or {+90°,±120°} or {-90°,±120°}. In the calculations described here, we started with the {-90°,-120°} and {+90°,-120°} structures with ζ fixed, and used Gaussian 90 to minimize the energy. All four combinations of positive and negative values of ζ and η are shown in Figure 3. From the figure it is clear that there are only two different structures to be considered. Since HA and HD are equivalent substituents, switching them does not result in a new structure. Switching them does results in reversing the sign of both ζ and η.

The most significant difference in the results from the two basis sets was that starting at ζ=-90°, η=-132°, the 3-21g optimization ended up in a {-90°,+132}

conformation by inverting the nitrogen center. This is identical to the {90°,-132°} structure. At 6-31g*, a {-90°,-120°} saddle point exists, and is 2.5 kcal/mol higher in energy than the {90°,-117°} structure. At both 3-21g and 6-31g*, the N-CY bond has lengthened by 0.07 Å and the bond angles centered on nitrogen have decreased from 120° to 105°-114° (see Table 1). The improper dihedral η is consistent with a tetrahedral arrangement of the substituents of N. Similar results have been found by Wiberg and Breneman [71] with the 6-31g* and 6-31g** basis sets. Using Møller-Plesset level 2 electron correlation calculations, they found an increase in the ζ=90° energy of 1.0 kcal/mol and in the ζ=-90° energy a change of +0.75 kcal/mol over the 6-31g* calculations alone.

B. N,N-dimethyl acetamide without ligands

We used the virtual dihedral CAY-OY-CD-CA (ζ) to calculate the barrier height in isomerization in N,N-dimethyl acetamide by fixing ζ to 90° and -90° with the initial value of the improper dihedral CY-CA-N-CD set to -140°, but allowing it to relax as the minimization progressed. The values for various internal coordinates and the energies relative to the unconstrained ground state (initial value of ζ = 180°) are given in Table 2. In contrast to formamide, there were two saddle point conformations for N,N-dimethyl acetamide in both 3-21g and 6-31g*. The barrier heights were found to be 18.8 and 15.6 kcal/mol with the 3-21g basis and 17.0 and 13.3 kcal/mol at 6-31g* for the ζ = 90° and -90° saddle point structures respectively. The results demonstrate that in both formamide and dimethyl acetamide the lower energy saddle point involves a trans arrangement of the nitrogen "lone pair" and the oxygen atom. This arrangement reduces the dipole of the molecule considerably, and the dipole dipole repulsion of the amino and carbonyl groups [69]. The energy difference is higher in dimethyl acetamide than in formamide and larger at 6-31g* than at 3-21g.

The ζ=-90° and 90° structures have fairly similar bond lengths and angles, except for slightly smaller angles involving nitrogen as the central atom when ζ=+90°. The value of the nitrogen improper η for the planar structure at 6-31g* is -165°, compared to 180° at 3-21g, indicating that the addition of d-orbitals increases the tetrahedral nature of the nitrogen and its substituents, including the lone pair orbital. The same is true for the saddle point structures where the 6-31g* structures are closer to tetrahedral (η=±120°) than the 3-21g structures by 7.9° (ζ=-90°) and 3.4° (ζ=90°).

C. Formamide + NH_4^+

To investigate the effect of interaction of a positively charged sidechain or other ion with the peptide bond, we performed *ab initio* calculations on complexes of formamide and N,N-dimethyl acetamide with ammonium ion.

The results for the complex of formamide and NH_4^+ are listed in Table 3. The planar formamide complex, the ζ=-90° complex, and the ζ=+90° complex (all at 6-31g*) are depicted in Figures 4, 5, and 6 respectively. The energies, ΔE, in Table 3 are the differences in energy between each structure and the planar formamide/NH_4^+ complex. ΔE(interaction) is the difference in energy between each complex and the sum of the uncomplexed formamide (from Table 1) and uncomplexed NH_4^+. At 3-21g, the energy of the saddle point conformations has gone from 18.3 kcal/mol (ζ=90°,η=-132°) to 0.3 kcal/mol (ζ=-90°,η=-125°) and 9.1 kcal/mol (ζ=90°,η=-118°). It should be noted that even though the ζ=-90° with η<0 saddle point did not exist in uncomplexed formamide, the complexed structure does exist and has considerably lower energy than the ζ=+90° conformation. This occurs because in the +90° conformation the NH_4^+ is closer to the carbonyl oxygen than in the -90° conformation. Even though the NH_4^+ is hydrogen-bonded to the nitrogen, it is close enough to the carbonyl oxygen for there to be a large favorable electrostatic interaction with the carbonyl (see Figure 5). The isomerization has been reduced to nearly free rotation about the N-CY bond. Rotating in

the opposite direction occurs through a 9 kcal/mol barrier, which is half the original barrier. In this case, the NH_4^+ is on the opposite side of the C-N bond as the carbonyl (Figure 6) which is less favorable than being on the same side.

The 6-31g* results yield relatively similar energies, with the barrier equal to 1.3 kcal/mol when ζ=-90° and 10.8 kcal/mol when ζ=90°. The energies of interaction indicate that the ammonium makes a much stronger hydrogen bond to the rotated structures than to the planar structure, and a much stronger bond to the ζ=-90° than to the ζ=90° structure. The planar interaction energies are -14.2 and -10.6 kcal/mol, while the ζ=-90° interaction energies are -32.2 and -27.3 kcal/mol. The ζ=+90° interaction energies are -23.4 and -15.3 kcal/mol. The reason is most likely that since the nitrogen electrons in the transition state structures have already rehybridized to an sp^3 configuration, the lone pair electrons are in a good geometry for hydrogen bonding. In the planar conformation, the electrons not involved in the planar sp^2 orbitals must rearrange to interact with the ammonium ion but at a considerable cost in energy. Even then, the hydrogen bond may not be as strong, leading to a less favorable interaction energy with the ammonium ion.

The stabilization due to NH_4^+ at 6-31g* for the ζ=+90° conformation is only 4.7 kcal/mol, which is the difference between the interaction energy of ammonium with planar formamide and ζ=-90° formamide (-10.6 and -15.3 kcal/mol). The difference for ζ=-90° is 16.7 kcal/mol (-10.6 - -27.3 kcal/mol). In the ζ=-90° conformation, the carbonyl carbon-oxygen bond dipole is anti-parallel to the lone pair-ammonium hydrogen bond, which is a much more favorable arrangement than the parallel dipoles in the ζ=90° structures (see Figure 3).

There are some significant differences in geometry between the two saddle point conformations of the complexes. While the bond lengths are relatively similar, the bond angles N-CY-HY, HD-N-HA, HA-N-CY and HD-N-CY are smaller when ζ=90° than when ζ=-90°, while the N-CY-OY and HY-CY-OY bond angles are larger. The

improper dihedrals (η) indicate that the "planar" structures (ζ=180°) are nearly tetrahedral about the nitrogen (η=-131° and -129° at 3-21g and 6-31g*), compared to the nearly planar arrangement in the uncomplexed molecules (η=180° and -165°). The ζ=90° structures are bent beyond tetrahedral, such that η is -118° and -112° at 3-21g and 6-31g*. This is considerably more bent than the ζ=-90° structures, where η is -125° and -119°. A difference in the hydrogen bond length arises from the two different basis sets used. The bond is approximately 0.2 Å longer in each 6-31g* structure compared to the relevant 3-21g structure. In the planar molecules the H1-N distance increases from 1.73 to 1.94 Å with the larger basis set. Similar differences are found for the saddle point structures.

D. N,N-dimethyl acetamide + NH_4^+

The structural parameters and energy differences for N,N-dimethyl acetamide complexed with ammonium ion are listed in Table 6. The 6-31g* planar, ζ=-90°, and ζ=+90° complexes are shown in Figures 7, 8, and 9 respectively. The energetics of the interaction of NH_4^+ with dimethyl acetamide are similar to the case of formamide, with the exception of the 3-21g calculation of the ζ=90° complex. Both of the ζ=-90° calculations have energies close to the planar complexes; the 3-21g energy is actually lower than the planar complex by 1.9 kcal/mol, and the 6-31g* energy is 0.2 kcal/mol lower than the planar case. The 6-31g* calculation of the ζ=90° complex yields an energy difference of 9.8 kcal/mol, which is close to the formamide value of 10.8 kcal/mol. However, the 3-21g energy is 2.5 kcal/mol lower than the 3-21g planar complex energy. In formamide, the 3-21g ζ=90° energy was 9.1 kcal/mol. The reason for the difference is that the ammonium ion has protonated the nitrogen of the dimethyl acetamide (N-H1 bond distance 1.0 Å), and the ammonia has stayed in place with an N-NA distance of 2.77 Å.

The 6-31g* calculation of the ζ=90° complex results in only a 3.5 kcal/mol stabilization over the uncomplexed transition between planar conformations of dimethyl acetamide. The ζ=-90° complex, however, is essentially equivalent in energy to the planar form, and is therefore a reduction of the barrier from 17.0 kcal/mol (ζ=-90°) and 13.3 kcal/mol (ζ=90°) to free rotation. The 6-31g* ζ=90° calculation shows some of the distortions of the tetrahedral geometry that were found in the corresponding formamide complex — a tightening of the X-N-X angles and nitrogen improper η closer to -120°. The dimethyl acetamide saddle point structures are all further from ideal tetrahedral geometry than the formamide structures, because of steric repulsion of the methyl groups.

IV. Discussion

The isomerization of Xaa-Pro peptide bonds to the correct isomer has been shown to be the slow step in the folding of a large number of proteins. The presence of proline in a protein sequence should indicate the likelihood of a slow folding phase as incorrect isomers change to the native protein conformation. Exceptions are likely to arise only when the folded protein can accommodate either isomer (e.g. BPTI [76, 77]) or when the isomerization can be catalyzed by other proteins (e.g. cyclophilin and FKBP) or sidechains within the protein itself (e.g., DHFR).

It can be inferred from the kinetic data that cyclophilin may have a different catalytic mechanism than the FKBP related proteins. Harrison and Stein [33] have found that cyclophilin reduces $\Delta H^{\ddagger}$ from 19-21 kcal/mol to 3-8 kcal/mol in a number of proline-containing peptides. However, $T\Delta S^{\ddagger}$ was generally large and negative (-9 to -15 kcal/mol) resulting in $\Delta G^{\ddagger}$ of 16-17 kcal/mol. For the same series of peptides, FKBP lowered $\Delta H^{\ddagger}$ to 11-15 kcal/mol with entropies of -3 to -8 kcal/mol. The crystal structure of cyclophilin complexed with a tetrapeptide has shown an arginine sidechain (Arg 55) Nη1 atom 4.5 Å from the proline nitrogen (and 3.0 Å from the proline carbonyl oxygen). FKBP has no positively charged sidechains near the catalytic site of the protein. We may

speculate that at least part of the catalytic activity of cyclophilin is due to the positively charged hydrogen bond donor of the arginine sidechain, withdrawing electron density from the C-N peptide bond in a manner similar to the NH_4^+/formamide and NH_4^+/dimethyl acetamide complexes demonstrated here.

To search for other candidate proteins, in addition to DHFR, which may catalyze their own folding by isomerization of proline by hydrogen bonding to the imide nitrogen, we performed a search of Protein Databank structures which have hydrogen bond donors within 3.5 Å of proline N. The list is presented in Table 5. Texter et al. [1] mention that in a survey of 42 high-resolution x-ray structures they found eight cases of positively charged residues (Arg, Lys, His) within 4 Å of proline nitrogen. An additional 23 charged residues were found hydrogen bonded to proline oxygen, 12 neutral donors to nitrogen, and 22 neutral donors to oxygen. In Table 5, we list only those sidechains which have hydrogen bond donating atoms within 3.5 Å of proline nitrogen, and where the donor/Pro N distance is less than the donor/Pro O distance. There are 6 arginines, 3 lysines, and 26 histidines as well as 14 asparagines, 4 cysteines, 3 glutamines, 36 serines, and 27 threonines. In total, there are 119 examples, some of which are from homologous proteins, which demonstrates that the interaction may be preserved in evolution. It is unlikely that all of these are involved in hydrogen bonds in the folded proteins, since some of the distances are rather large. In addition, many of them may not at all be involved in cis-trans isomerization of the proline residue during protein folding, but their proximity to the proline nitrogen makes them candidates for further study. It should be noted that a majority of these close contacts are between proline N and sidechain γ and δ atoms in residues immediately preceding the proline residue. These can not be immediately ruled out as candidates in assisting protein folding. In such cases, the transition structures are likely to have the nitrogen lone pair *trans* to the carbonyl oxygen preceding proline. In this configuration, the Cα atom of the residue previous to proline is

cis to the lone pair, and any γ or δ hydrogen bond donor of the sidechain is likely to be closer to the lone pair in the transition state than if the lone pair was trans to Cα.

One of the more interesting aspects of the NH_4^+/amide complexes is the much smaller barrier when ζ=-90° than when ζ=+90°. The barriers for formamide and dimethyl acetamide at both 3-21g and 6-31g* are approximately 9-10 kcal/mol when ζ=+90°, but near 0 kcal/mol when ζ=-90° in the presence of NH_4^+. Without NH_4^+, the +90° barriers are 2-3 kcal/mol lower (at 13-18 kcal/mol) than when ζ=-90°. While the effect of solvent on the two saddle point conformations is not known, the calculations on the complex indicate that cyclophilin and self-catalyzed isomerization of proline is much more likely to occur via the ζ=-90° barrier while uncatalyzed isomerization is slightly more likely to occur via the ζ=+90° conformation.

References

1. F. L. Texter, D. B. Spencer, R. Rosenstein, and C. R. Matthews. Intramolecular catalysis of a proline isomerization reaction in the folding of dihydrofolate reductase. *Biochemistry*, **31**, 5687-5691 (1992).

2. J. F. Brandts, H. R. Halvorson, and M. Brennan. Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry*, **14**, 4953-4963 (1975).

3. F. X. Schmid. Kinetics of unfolding and refolding of single-domain proteins. In *Protein Folding*. T. E. Creighton ed. (New York: W. H. Freeman, 1993).

4. D. E. Stewart, A. Sarkar, and J. E. Wampler. Occurrence and role of cis peptide bonds in protein structures. *J. Mol. Biol.*, **214**, 253-260 (1990).

5. T. Y. Tsong, R. L. Baldwin, and E. L. Elson. Properties of the refolding and unfolding reactions of ribonuclease A. *Proc. Natl. Acad. Sci. USA*, **69**, 1809-1812 (1972).

6. L.-N. Lin and J. F. Brandts. Involvement of prolines-114 and -117 in the slow refolding phase of ribonuclease A as determined by isomer-specific proteolysis. *Biochemistry*, **23**, 5713-5723 (1984).

7. L.-N. Lin and J. F. Brandts. Evidence for the existence of three or more slow phases in the refolding of ribonuclease A and some characteristics of the phases. *Biochemistry*, **26**, 3537-3543 (1987).

8. T. Kiefhaber, R. Quaas, U. Hahn, and F. X. Schmid. Folding of ribonuclease T1. 1. Existence of multiple unfolded states created by proline isomerization. *Biochemistry*, **29**, 3053-3061 (1990).

9. T. Y. Tsong and R. L. Baldwin. *J. Mol. Biol.*, **69**, 145 (1972).

10. T. Y. Tsong. Detection of three kinetic phases in the thermal unfolding of ferricytochrome c. *Biochemistry*, **12**, 2209-2214 (1973).

11. A. Ikai, W. Fish, and C. Tanford. Kinetics of unfolding and refolding of proteins II. Results for cytochrome c. *J. Mol. Biol.*, **73**, 165-184 (1973).

12. A. Ikai. Ph. D. Dissertation, Duke University (1971).

13. M. R. Summers and P. McPhie. The mechanism of unfolding of globular proteins. *Biochem. Biophys. Res. Comm.*, **47**, 831-837 (1972).

14. R. F. Kelley and E. Stellwagen. Conformational transitions in thioredoxin in guanidine hydrochloride. *Biochemistry*, **23**, 5095-5102 (1984).

15. K. L. B. Borden and F. M. Richards. Folding kinetics of phage T4 thioredoxin. *Biochemistry*, **29**, 3071-3077 (1990).

16. L.-N. Lin and J. F. Brandts. Further evidence suggesting that the slow phase in protein folding and refolding is due to proline isomerization: A kinetic study of carp parvalbumins. *Biochemistry*, **17**, 4102-4110 (1978).

17. G. V. Semisotnov, V. N. Uversky, I. V. Sokolovsky, A. M. Gutin, O. I. Razgulyaev, and N. A. Rodionova. Two slow stages in refolding of bovine carbonic anhydrase B are due to proline isomerization. *J. Mol. Biol.*, **213**, 561-568 (1990).

18. L. Ramdas and B. T. Nall. Folding/unfolding kinetics of mutant forms of iso-1-cytochrome c with replacement of proline-71. *Biochemistry*, **25**, 6959-6964 (1986).

19. R. F. Kelley and F. M. Richards. Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry*, **26**, 6765-6774 (1987).

20. T. B. White, P. B. Berget, and B. T. Nall. Changes in conformation and slow refolding kinetics in mutant iso-2-cytochrome c with replacement of a conserved proline residue. *Biochemistry*, **26**, 4358-4366 (1987).

21. L. C. Wood, T. B. White, L. Ramdas, and B. T. Nall. Replacement of a conserved proline eliminates the absorbance-detected slow folding phase of iso-2-cytochrome c. *Biochemistry*, **27**, 8562-8568 (1988).

22. J. Kördel, S. Forsén, T. Drakenberg, and W. J. Chazin. The rate and structural consequences of proline cis-trans isomerization in calbindin D_{9k}: NMR studies of the minor (cis-Pro43) isoform and the Pro43Gly mutant. *Biochemistry*, **29**, 4400-4409 (1990).

23. T. Kiefhaber, H.-P. Grunert, U. Hahn, and F. X. Schmid. Replacement of a cis proline simplifies the mechanism of ribonuclease T1 folding. *Biochemistry*, **29**, 6475-6480 (1990).

24. J. T. Gerig. The effect of adjacent charges on the kinetics of rotation of the peptide bond. *Biopolymers*, **10**, 2435-2443 (1979).

25. R. Deslauriers, J. M. Becker, A. S. Steinfeld, and F. Naider. Steric effects of cis-trans isomerism on neighboring residues in proline oligopeptides: A ^{13}C-NMR

study of conformational heterogeneity in linear tripeptides. *Biopolymers*, **18**, 523-538 (1979).

26. L.-N. Lin and J. F. Brandts. Determination of cis-trans proline isomerization by trypsin proteolysis. Application to a pentapeptide and to oxidized ribonuclease A. *Biochemistry*, **22**, 553-559 (1983).

27. E. S. Eberhardt, S. N. Loh, A. P. Hinck, and R. T. Raines. Solvent effects on the energetics of prolyl peptide bond isomerization. *J. Am. Chem. Soc.*, **114**, 5437-5439 (1992).

28. R. K. Harrison and R. L. Stein. Mechanistic studies of peptidyl prolyl cis-trans isomerase: Evidence for catalysis by distortion. *Biochemistry*, **29**, 1684-1689 (1990).

29. A. Berger, A. Loewenstein, and S. Meiboom. Nuclear magnetic resonance study of the protolysis and ionization of N-methylacetamide. *J. Am. Chem. Soc.*, **81**, 62-67 (1959).

30. G. Fraenkel and C. Franconi. Protonation of amides. *J. Am. Chem. Soc.*, **82**, 4478-4483 (1960).

31. R. C. J. Neuman, W. R. Woolfenden, and V. Jonas. The effect of hydrogen bonding on the barrier to rotation about amide bonds. *J. Phys. Chem.*, **73**, 3177-3180 (1969).

32. T. Drakenberg, K.-I. Dahlqvist, and S. Forsén. The barrier to internal rotation in amides. IV. N,N-dimethylamides: Substituent and solvent effects. *J. Phys. Chem.*, **76**, 2178-2183 (1972).

33. R. K. Harrison and R. L. Stein. Mechanistic studies of enzymic and nonenzymic prolyl cis-trans isomerization. *J. Am. Chem. Soc.*, **114**, 3464-3471 (1992).

34. T. E. Creighton. Possible implications of many proline residues for the kinetics of protein unfolding and refolding. *J. Mol. Biol.*, **125**, 401-406 (1978).

35. I. Z. Steinberg, W. F. Harrington, A. Berger, M. Sela, and E. Katchalski. The configurational changes of poly-L-proline in solution. *J. Am. Chem. Soc.*, **82**, 5263-5279 (1960).

36. R. A. Morris. Rapamycin: FK506's fraternal twin or distant cousin ? *Immunology Today*, **12**, 137-140 (1991).

37. F. McKeon. Immunosuppressants meet protein phosphatases. *Cell*, **66**, 823-826 (1991).

38. S. L. Schreiber. Chemistry and biology of the immunophilins and their immunosuppressive ligands. *Science*, **251**, 283-287 (1992).

39. S. L. Schreiber. Immunophilin-sensitive protein phosphatase action in cell signaling pathways. *Cell*, **70**, 365-368 (1992).

40. S. L. Schreiber and G. R. Crabtree. The mechanism of action of cyclosporin A and FK506. *Immunology Today*, **13**, 136-141 (1992).

41. M. K. Rosen and S. L. Schreiber. Natural products as probes of cellular function: Studies of immunophilins. *Angew. Chem. Int. Ed. Engl.*, **31**, 384-400 (1992).

42. E. R. Schönbrunner, S. Mayer, M. Tropschug, G. Fischer, N. Takahashi, and F. X. Schmid. Catalysis of protein folding by cyclophilins from different species. *J. Biol. Chem.*, **266**, 3630-3635 (1991).

43. R. F. Standaert, A. Galat, G. L. Verdine, and S. L. Schreiber. Molecular cloning and overexpression of the human FK506-binding protein FKBP. *Nature*, **346**, 671-677 (1990).

44. W. S. Lane, A. Galat, M. W. Harding, and S. L. Schreiber. *J. Protein Chem.*, **10**, 151-160 (1991).

45. M. Tropschug, I. B. Barthelmess, and W. Neupert. Sensitivity to cyclosporin A is mediated by cyclophilin in *Neurospora crassa* and *Saccharomyces cerevisiae. Nature*, **342**, 953-955 (1989).

46. M. Tropschug, E. Wachter, S. Mayer, E. R. Schönbrunner, and F. X. Schmid. Isolation and sequence of an FK506-binding protein from *N. crassa* which catalyses protein folding. *Nature*, **346**, 674-677 (1990).

47. J. Liu and C. T. Walsh. Peptidyl-prolyl cis-trans isomerase from *Escherichia coli*: A periplasmic homolog of cyclophilin that is not inhibited by cyclosporin. *Proc. Natl. Acad. Sci. USA*, **87**, 4028-4032 (1990).

48. Y. Koltin, L. Faucette, D. J. Bergsma, M. A. Levy, R. Cafferkey, P. L. Koser, R. K. Johnson, and G. P. Livi. Rapamycin sensitivity in *Saccharomyces cerevisiae* is mediated by a peptidyl-prolyl cis-trans isomerase related to human FK506-binding protein. *Mol. Cell. Biol.*, **11**, 1718-1723 (1991).

49. H. P. Bächinger. The influence of peptidyl-prolyl cis-trans isomerase on the in vitro folding of Type III collagen. *J. Biol. Chem.*, **262**, 17144-17148 (1987).

50. K. Lang, F. X. Schmid, and G. Fischer. Catalysis of protein folding by prolyl isomerase. *Nature*, **329**, 268-270 (1987).

51. P.-O. Freskgård, N. Bergenhem, B.-H. Jonnson, M. Svensson, and U. Carlsson. Isomerase and chaperone activity of prolyl isomerase in the folding of carbonic anhydrase. *Science*, **258**, 466-468 (1992).

52. J. Heitman, N. R. Movva, and M. N. Hall. Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. *Science*, **253**, 905-909 (1991).

53. J. Heitman, N. R. Movva, P. C. Hiestand, and M. N. Hall. FK506-binding protein proline rotamase is a target for the immunosuppressant agent FK506 in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA*, **88**, 1948-1952 (1991).

54. J. J. Siekerka, S. H. Y. Hung, M. Poe, C. L. Lin, and N. H. Sigal. A cytosolic binding protein for the immunosuppressant FK506 has peptidyl-prolyl isomerase activity but is distinct from cyclophilin. *Nature*, **341**, 755-757 (1989).

55. J. Liu, J. D. Farmer, W. S. Lane, J. Friedman, I. Weissman, and S. L. Schreiber. *Cell*, **66**, 807-815 (1991).

56. R. A. Aldape, O. Futer, M. T. DeCenzo, B. P. Jarrett, M. A. Murcko, and D. J. Livingston. Charged surface residues of FKBP12 participate in formation of the FKBP12-FK506-calcineurin complex. *J. Biol. Chem.*, **267**, 16029-16032 (1992).

57. W. M. Flanagan, B. Corthésy, R. J. Bram, and G. R. Crabtree. Nuclear association of a T-cell transcription factor blocked by FK-506 and cyclosporin A. *Nature*, **352**, 803-807 (1991).

58. R. K. Harrison and R. L. Stein. Substrate specificities of the peptidyl prolyl cis-trans isomerase activities of cyclophilin and FK-506 binding protein: Evidence for the existence of a family of distinct enzymes. *Biochemistry*, **29**, 3813-3816 (1990).

59. M. K. Rosen, R. F. Standaert, A. Galat, M. Nakatsuka, and S. L. Schreiber. Inhibition of FKBP rotamase activity by immunosuppressant FK506: Twisted amide surrogate. *Science*, **248**, 863-866 (1990).

60. M. W. Albers, C. T. Walsh, and S. L. Schreiber. Substrate specificity for the human rotamase FKBP: A view of FK506 and rapamycin as leucine-(twisted amide)-proline mimics. *J. Org. Chem.*, **55**, 4984-4986 (1990).

61. S. T. Park, R. A. Aldape, O. Futer, M. T. DeCenzo, and D. J. Livingston. PPIase catalysis by human FK506-binding protein proceeds through a conformational twist mechanism. *J. Biol. Chem.*, **267**, 3316-3324 (1992).

62. J. L. Kofron, P. Kuzmic, V. Kishore, E. Colón-Bonilla, and D. H. Rich. Determination of kinetic constants for peptidyl prolyl cis-trans isomerases by an improved spectrophotometric assay. *Biochemistry*, **30**, 6127-6134 (1991).

63. J. Kallen, C. Spitzfaden, M. G. M. Zurini, G. Wider, H. Widmer, K. Wüthrich, and M. D. Walkinshaw. Structure of human cyclophilin and its binding site for cyclosporin A determined by X-ray crystallography and NMR spectroscopy. *Nature*, **353**, 276-279 (1991).

64. M. K. Rosen, S. W. Michnick, M. Karplus, and S. L. Schreiber. Proton and nitrogen sequential assignments and secondary structure determination of the human FK506 and rapamycin binding protein. *Biochemistry*, **30**, 4774-4789 (1991).

65. S. W. Michnick, M. K. Rosen, T. J. Wandless, M. Karplus, and S. L. Schreiber. Solution structure of FKBP, a rotamase enzyme and receptor for FK506 and rapamycin. *Science*, **252**, 836-839 (1991).

66. G. D. Van Duyne, R. F. Standaert, P. A. Karplus, S. L. Schreiber, and J. Clardy. Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. *Science*, **252**, 839-842 (1991).

67. J. Liu, M. W. Albers, C. M. Chen, S. L. Schreiber, and C. T. Walsh. Cloning, expression, and purification of human cyclophilin in *Escherichia coli* and assessment of the catalytic role of cysteines by site-directed mutagenesis. *Proc. Natl. Acad. Sci. USA*, **87**, 2304-2308 (1990).

68. R. F. Nalewajski. Optimized geometries of the saddle-point rotamers of formamide. *J. Am. Chem. Soc.*, **100**, 41-46 (1978).

69. K. B. Wiberg and K. E. Laidig. Barriers to rotation adjacent to double bonds. 3. The C-O barrier in formic acid, methyl formate, acetic acid, and methyl acetate. The origin of ester and amide "resonance". *J. Am. Chem. Soc.*, **109**, 5935-5943 (1987).

70. K. B. Wiberg, C. M. Hadad, C. M. Breneman, K. E. Laidig, M. A. Murcko, and T. J. LePage. The response of electrons to structural changes. *Science*, **252**, 1266-1272 (1991).

71. K. B. Wiberg and C. M. Breneman. Resonance interactions in acyclic systems. 3. Formamide internal rotation revisited. Charge and energy redistribution along the C-N bond rotational pathway. *J. Am. Chem. Soc.*, **114**, 831-840 (1992).

72. M. J. Frisch, M. Head-Gordon, H. B. Schlegel, K. Raghavachari, J. S. Binckley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, E. M. Fluder, S. Topiol, and J. A. Pople. Gaussian 88 (1988).

73. M. J. Frisch, M. Head-Gordon, G. W. Trucks, J. B. Foresman, H. B. Schlegel, K. Raghavachari, M. Robb, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, S. Topiol, and J. A. Pople. Gaussian 90, Revision I (1990).

74. S. Fischer. PhD dissertation, Harvard University (1992).

75. S. Fischer and M. Karplus. Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem. Phys. Letters*, **194**, 256-261 (1992).

76. M. Levitt. Effect of proline residues on protein folding. *J. Mol. Biol.*, **145**, 251-263 (1981).

77. M. R. Hurle, S. Anderson, and I. D. Kuntz. Confirmation of the predicted source of a slow folding reaction: proline 8 of bovine pancreatic trypsin inhibitor. *Protein Engineering*, **4**, 451-455 (1991).

Table 1
Formamide

	Planar		ζ=-90°		ζ=90°	
Basis	3-21g	6-31g*	3-21g	6-31g*	3-21g	6-31g*
N-HD[a]	0.9952	0.9927	1.0078	1.0035	1.0076	1.0066
N-HA	0.9978	0.9955	1.0044	1.0058	1.0046	1.0039
N-CY	1.3536	1.3488	1.4288	1.4203	1.4287	1.4263
CY-OY	1.2122	1.1928	1.2034	1.1795	1.2034	1.1833
CY-HY	1.0833	1.0905	1.0788	1.0937	1.0786	1.0869
HD-N-HA[b]	118.65	118.95	111.73	106.99	111.72	105.83
HA-N-CY	121.92	121.79	114.69	110.60	114.69	108.94
HD-N-CY	119.43	119.23	113.98	109.16	113.98	108.31
N-CY-HY	112.29	112.72	113.08	116.09	113.08	113.26
N-CY-OY	125.28	124.92	125.34	123.34	125.39	125.13
HY-CY-OY	122.43	122.36	121.57	120.56	121.53	121.60
HD-N-CY-HY	-0.01	1.02	107.41	46.67	-107.43	-116.51
HD-N-CY-OY	-180.00	-179.04	-73.45	-134.40	73.44	64.33
HA-N-CY-HY	179.99	179.35	-122.00	-70.78	122.00	128.83
HA-N-CY-OY	0.00	-0.72	57.13	108.15	-57.13	-50.33
CY-HA-N-CD (η)	180.00	178.37	132.03	-119.71	-132.02	-116.68
N-HY-CY-OY	179.99	-179.94	-179.17	-178.96	179.17	179.19
HY-OY-HD-HA (ζ)	179.99	-179.77	-90.00	-90.00	90.00	90.00
ΔE[c]	0.000	0.000	18.312	18.004	18.312	15.506

[a] Bond lengths in Å.

[b] Bond angles and dihedral angles in degrees.

[c] Energies in kcal/mol.

Table 2

N,N-Dimethyl acetamide

	Planar		ζ=-90°		ζ=90°	
Basis	3-21g	6-31g*	3-21g	6-31g*	3-21g	6-31g*
N-CD	1.4568	1.4464	1.4644	1.4502	1.4779	1.4577
N-CA	1.4631	1.4497	1.4681	1.4531	1.4752	1.4562
N-CY	1.3615	1.3631	1.4259	1.4276	1.4398	1.4338
CY-OY	1.2217	1.2019	1.2043	1.1859	1.2061	1.1882
CY-CY	1.5193	1.5173	1.5180	1.5165	1.5031	1.5049
HA1-CA	1.0852	1.0878	1.0902	1.0920	1.0865	1.0878
HA2-CA	1.0852	1.0845	1.0823	1.0834	1.0819	1.0834
HA3-CA	1.0763	1.0761	1.0818	1.0825	1.0828	1.0840
HD1-CD	1.0772	1.0784	1.0823	1.0834	1.0833	1.0844
HD2-CD	1.0855	1.0836	1.0823	1.0833	1.0817	1.0832
HD3-CD	1.0855	1.0885	1.0899	1.0919	1.0869	1.0882
HY1-CAY	1.0832	1.0845	1.0842	1.0855	1.0841	1.0852
HY2-CAY	1.0832	1.0848	1.0841	1.0845	1.0837	1.0850
HY3-CAY	1.0784	1.0794	1.0793	1.0805	1.0795	1.0810
CD-N-CA	115.09	115.34	115.41	113.51	112.89	111.99
CA-N-CY	125.97	124.09	117.18	115.31	112.55	110.97
CD-N-CY	118.94	118.62	114.78	112.29	112.99	111.43
N-CY-CY	117.38	117.79	115.56	117.65	112.56	113.87
N-CY-OY	121.99	122.08	122.06	120.79	122.61	122.77
CY-CY-OY	120.63	120.12	122.36	121.54	124.83	123.36
HA1-CA-N	110.23	111.31	112.85	113.46	112.42	112.89
HA2-CA-N	110.23	109.44	108.88	108.91	108.45	108.70
HA3-CA-N	108.36	109.47	109.39	109.63	109.49	109.77
HD1-CD-N	111.65	111.62	109.72	109.90	109.57	109.87
HD2-CD-N	110.15	109.10	108.43	108.53	108.57	108.84
HD3-CD-N	110.16	111.90	113.48	113.99	112.16	112.68
HY1-CAY-CY	111.28	111.53	110.78	110.52	108.79	109.14
HY2-CAY-CY	111.31	111.55	109.90	111.05	109.47	109.74
HY3-CAY-CY	106.97	107.38	109.51	109.31	110.81	110.66
CD-N-CY-CAY	0.01	-12.01	57.77	53.31	-110.20	-113.03
CD-N-CY-OY	-179.99	169.14	-124.24	-128.28	70.40	67.51
CA-N-CY-CAY	-179.99	-175.35	-82.44	-78.82	120.48	121.44
CA-N-CY-OY	0.01	5.80	95.55	99.59	-58.93	-58.03
CY-CA-N-CD (η)	180.00	-164.77	-140.93	-133.02	-129.15	-124.97
N-CAY-CY-OY	179.99	178.87	-177.98	-178.40	179.39	179.46
CAY-OY-CD-CA (ζ)	-179.99	176.65	-90.00	-90.00	90.00	90.00
HA1-CA-N-CY	120.03	98.26	76.30	72.84	66.40	65.76
HA2-CA-N-CY	-120.05	-142.00	-163.31	-166.88	-172.65	-173.18
HA3-CA-N-CY	-0.01	-22.13	-45.32	-49.04	-54.66	-55.39
HD1-CD-N-CY	0.04	29.90	44.42	48.91	54.32	55.68
HD2-CD-N-CY	120.49	148.68	162.04	166.35	172.41	173.54
HD3-CD-N-CY	-120.41	-91.95	-77.63	-73.31	-66.71	-65.44
HY1-CAY-CY-N	-60.84	-58.50	-64.93	-71.26	-60.53	-60.47
HY2-CAY-CY-N	60.67	62.46	53.99	47.77	56.19	56.28
HY3-CAY-CY-N	179.93	-178.05	174.67	169.07	178.13	178.16
ΔE	0.000	0.000	18.795	17.045	15.576	13.272

Table 3

Formamide/NH_4^+

	Planar		ζ=-90°		ζ=90°	

Basis	3-21g	6-31g*	3-21g	6-31g*	3-21g	6-31g*
N-HD	1.0087	1.0030	1.0128	1.0076	1.0185	1.0109
N-HA	1.0113	1.0048	1.0137	1.0090	1.0164	1.0089
N-CY	1.4252	1.4030	1.4541	1.4323	1.4814	1.4563
CY-OY	1.1971	1.1803	1.2012	1.1835	1.1902	1.1723
CY-HY	1.0784	1.0878	1.0774	1.0862	1.0750	1.0851
HD-N-HA	111.91	111.06	109.05	105.38	106.44	103.65
HA-N-CY	115.08	114.72	112.42	110.28	109.40	106.78
HD-N-CY	112.45	111.43	112.41	109.42	108.99	106.30
N-CY-HY	113.07	114.13	116.71	117.46	112.85	113.86
N-CY-OY	122.34	122.10	120.33	120.56	122.58	122.75
HY-CY-OY	124.54	123.66	122.95	121.98	124.57	123.38
HD-N-CY-HY	-23.68	-20.60	49.81	45.57	-115.67	-119.29
HD-N-CY-OY	158.92	163.06	-129.36	-134.09	64.68	61.18
HA-N-CY-HY	-153.42	-147.88	-73.70	-69.91	128.32	130.53
HA-N-CY-OY	29.19	35.78	107.13	110.43	-51.32	-49.00
CY-HA-N-HD (η)	-131.35	-129.24	-125.37	-118.57	-117.89	-112.37
N-HY-CY-OY	177.33	176.28	179.14	179.65	179.64	179.52
HY-OY-HD-HA (ζ)	-177.15	-170.68	-90.00	-90.00	90.00	90.00
N-H1	1.7269	1.9427	1.7569	1.9875	1.6017	1.8380
N-NA	2.7992	2.9758	2.7839	2.9583	2.7113	2.8861
NA-H2	1.0179	1.0114	1.0158	1.0102	1.0172	1.0111
NA-H3	1.0189	1.0125	1.0213	1.0137	1.0171	1.0111
NA-H4	1.0176	1.0112	1.0156	1.0100	1.0170	1.0110
H1-N-HA	105.73	108.00[a]	113.81	118.87	107.11	108.57
H2-NA-N	110.06	111.09	117.06	118.60	108.44	108.42
H3-NA-N	107.54	105.88	91.66	89.69	110.62	110.77
H4-NA-N	111.77	112.46	116.20	116.92	110.84	111.01
H1-N-HA-HD	-113.41	-116.21	-123.03	-129.73	-114.80	-115.75
NA-N-HD-HA	113.86	117.55	131.81	139.84	113.68	114.64
H2-NA-N-CY	163.62	158.06	123.90	125.46	178.10	177.46
H3-NA-N-CY	44.95	39.87	9.95	12.48	58.72	58.14
H4-NA-N-CY	-74.86	-79.23	-103.83	-99.96	-62.39	-63.03
ΔE	0.000	0.000	0.282	1.327	9.122	10.802
ΔE interaction	-14.172	-10.636	-32.201	-27.313	-23.361	-15.340

[a] Constrained angle H1-N-HA = 108°

Table 4

N,N-Dimethyl acetamide/NH_4^+

	Planar		ζ=-90°		ζ=90°	
Basis:	3-21g	6-31g*	3-21g	6-31g*	3-21g	6-31g*
N-CD	1.4963	1.4631	1.4939	1.4696	1.5200	1.4789
N-CA	1.5001	1.4714	1.4995	1.4737	1.5194	1.4778
N-CY	1.4478	1.4115	1.4626	1.4439	1.5142	1.4645
CY-OY	1.2022	1.1965	1.2082	1.1922	1.1892	1.1785
CY-CY	1.5079	1.5076	1.5033	1.5062	1.4943	1.5041
HA1-CA	1.0818	1.0853	1.0831	1.0855	1.0775	1.0820
HA2-CA	1.0806	1.0823	1.0803	1.0818	1.0786	1.0834
HA3-CA	1.0765	1.0785	1.0811	1.0824	1.0798	1.0834
HD1-CD	1.0786	1.0816	1.0818	1.0833	1.0799	1.0835
HD2-CD	1.0804	1.0821	1.0805	1.0819	1.0786	1.0833
HD3-CD	1.0825	1.0858	1.0829	1.0856	1.0778	1.0824
HY1-CAY	1.0829	1.0849	1.0853	1.0854	1.0841	1.0867
HY2-CAY	1.0839	1.0836	1.0813	1.0826	1.0841	1.0865
HY3-CAY	1.0791	1.0795	1.0793	1.0798	1.0791	1.0796
CD-N-CA	111.42	114.47	111.73	111.52	110.22	110.10
CA-N-CY	116.64	119.44	113.83	113.63	109.21	108.88
CD-N-CY	112.42	115.00	111.50	110.90	109.70	109.30
N-CY-CY	116.77	119.55	117.99	119.56	113.75	115.25
N-CY-OY	119.13	118.04	117.93	118.01	117.10	120.45
CY-CY-OY	124.04	122.35	124.08	122.43	129.15	124.30
HA1-CA-N	110.08	110.95	111.25	112.31	108.30	111.48
HA2-CA-N	108.55	109.33	108.42	108.69	108.28	109.01
HA3-CA-N	109.41	109.58	109.46	109.83	108.83	109.80
HD1-CD-N	110.75	110.75	109.79	110.08	108.92	109.86
HD2-CD-N	108.33	108.34	108.24	108.46	108.35	109.08
HD3-CD-N	110.74	113.03	111.76	112.85	108.17	111.32
HY1-CAY-CY	112.03	111.19	110.16	109.64	110.30	110.01
HY2-CAY-CY	110.83	111.06	111.68	111.91	110.41	110.27
HY3-CAY-CY	107.66	108.23	108.69	108.87	109.03	109.78
CD-N-CY-CAY	-32.34	16.80	49.85	49.69	-116.24	-116.86
CD-N-CY-OY	150.25	-160.53	-129.82	-130.07	63.58	63.22
CA-N-CY-CAY	-162.82	-124.64	-77.67	-76.87	122.87	122.84
CA-N-CY-OY	19.78	58.03	102.66	103.38	-57.32	-57.08
CY-CA-N-CD (η)	-133.08	-143.49	-128.64	-127.72	-120.28	-119.55
N-CAY-CY-OY	177.27	177.22	179.65	179.74	-179.79	179.92
CAY-OY-CD-CA (ζ)	173.79	-132.10	-90.00	-90.00	90.00	90.00
HA1-CA-N-CY	70.65	81.82	69.11	70.11	57.76	60.90
HA2-CA-N-CY	-169.74	-158.70	-171.21	-170.07	178.06	-178.77
HA3-CA-N-CY	-48.77	-38.91	-51.99	-51.36	-62.12	-59.62
HD1-CD-N-CY	56.56	46.23	52.47	52.19	61.37	59.59
HD2-CD-N-CY	176.14	164.71	171.44	170.58	-178.76	178.75
HD3-CD-N-CY	-64.89	-75.79	-68.93	-69.55	-58.57	-61.00
HY1-CAY-CY-N	-48.62	-68.72	-76.69	-77.19	-60.85	-61.11
HY2-CAY-CY-N	72.54	51.11	43.53	42.30	58.98	58.42
HY3-CAY-CY-N	-169.00	171.88	164.84	164.16	179.11	178.79
N-H1	1.4855	1.8000	1.6162	1.8904	1.0661	1.7017
N-NA	2.6415	2.9042	2.6963	2.9032	2.7653	2.7763
NA-H2	1.0158	1.0101	1.0144	1.0093	1.0113	1.0102
NA-H3	1.0167	1.0136	1.0200	1.0127	1.0110	1.0103
NA-H4	1.0153	1.0096	1.0143	1.0097	1.0110	1.0011
H1-N-HA	104.92	108.00[a]	111.55	113.52	107.40	105.13
H2-NA-N	108.39	117.77	115.90	116.71	106.23	108.28

H3-NA-N	109.35	96.65	94.36	94.25	111.27	109.00
H4-NA-N	111.98	112.92	114.93	115.22	112.70	113.61
H1-N-HA-HD	-112.71	-115.23	-121.54	-125.13	-118.08	-113.85
NA-N-HD-HA	111.68	121.35	128.18	132.08	113.60	111.38
H2-NA-N-CY	155.76	148.72	126.44	127.97	183.75	227.15
H3-NA-N-CY	37.22	32.35	11.17	12.78	65.50	109.17
H4-NA-N-CY	-83.86	-82.23	-103.77	-101.63	-57.23	-12.11
ΔE	0.000	0.000	-1.906	-0.152	-2.534	9.767
ΔE interaction	-19.024	-13.338	-39.725	-30.535	-37.137	-16.842

[a] Constrained angle H1-N-HA = 108°

Table 5

Proline nitrogen-hydrogen bond donor distances

PDB CODE	Resol(Å)	Protein Name	Pro Res#-Chain	Donor Res. type	Donor residue #	Donor atom	Pro N-D dist. (Å)	Pro O-D dist. (Å)
8ADH	2.40	ALCOHOL DEHYDROGENASE	30	ARG	37	NH2	3.043	5.249
1FDH	2.50	HEMOGLOBIN F	124-G	ARG	31-A	NH2	3.322	4.815
1OVA	1.95	OVALBUMIN	175-A	ARG	154-A	NH1	3.358	6.115
1FDL	2.50	LYSOZYME	215-H	ARG	191-H	NH2	3.387	6.144
3BLM	1.90	BACTERIOCHLOR.-A PROT.	283	ARG	230	NH2	3.471	5.542
2HHB	1.74	HEMOGLOBIN	119-A	ARG	30-B	NH1	3.494	5.034
1SNC	1.65	STAPH. NUCLEASE	47	HIS	46	ND1	2.696	4.561
4TIM	2.40	TRIOSE PHOSPHATE ISOM.	58-B	HIS	57-B	ND1	2.983	4.979
2HIP	2.50	HIPIP	51-A	HIS	50-A	ND1	3.002	4.768
3IL8	2.00	INTERLEUKIN-8	19	HIS	18	ND1	3.025	4.992
2PKA	2.05	KALLIKREIN A	92-A	HIS	91-A	ND1	3.063	4.996
4TPI	2.20	TRYPSIN INHIBITOR	92-Z	HIS	91-Z	ND1	3.077	4.925
2SNI	2.10	CHYMOTRYPSIN INHIB. 2	40-E	HIS	39-E	ND1	3.083	5.082
1CSE	1.20	SUBTILISIN CARLSBERG	40-E	HIS	39-E	ND1	3.164	5.163
1TON	1.80	TONIN	92	HIS	91	ND1	3.171	5.212
1MEE	2.00	EGLIN-C	40-A	HIS	39-A	ND1	3.174	5.001
1TPP	1.40	TRYPSIN BETA	92	HIS	91	ND1	3.197	5.073
1TGS	1.80	PIG PANC.SECR TRYP INH	92-Z	HIS	91-Z	ND1	3.208	5.205
2TEC	1.98	EGLIN C	46-E	HIS	45-E	ND1	3.225	5.049
3B5C	1.50	CYTOCHROME B5	81	HIS	80	ND1	3.226	5.072
4FD1	1.90	FERREDOXIN	36	HIS	35	ND1	3.244	5.199
2GBP	1.90	D-GAL/D-GLUC BINDING P	153	HIS	152	ND1	3.262	5.155
1CY3	2.50	CYTOCHROME C3	108	HIS	96	ND1	3.284	3.365
1S01	1.70	SUBTILISIN BPN'	40	HIS	39	ND1	3.294	5.173
1TGN	1.65	TRYPSINOGEN	92	HIS	91	ND1	3.325	5.289
2FBJ	1.95	IGA FAB FRAGMENT	53-H	HIS	52-H	ND1	3.330	5.249
1TIM	2.50	TRIOSE PHOSPHATE ISOM.	70-A	HIS	115-A	NE2	3.336	5.282
8ADH	2.40	ALCOHOL DEHYDROGENASE	106	HIS	105	ND1	3.377	5.297
1TEC	2.20	THERMITASE	46-E	HIS	45-E	ND1	3.393	5.290
1TAB	2.30	BOWMAN-BIRK INHIBITOR	92-E	HIS	91-E	ND1	3.405	5.165
1THB	1.50	HEMOGLOBIN	44-A	HIS	45-A	ND1	3.468	4.964
2HIP	2.50	HIPIP	23-A	HIS	22-A	ND1	3.484	5.382
3FAB	2.00	IGG1 FAB' FRAGMENT	127-H	LYS	213-H	NZ	2.954	5.209
1OVA	1.95	OVALBUMIN	391-D	LYS	216-D	NZ	3.159	5.193
2ABX	2.50	ALPHA BUNGAROTOXIN	53-B	LYS	26-B	NZ	3.420	4.949
8CAT	2.50	CATALASE	295-A	ASN	294-A	ND2	2.842	4.886
1COX	1.80	CHOLESTEROL OXIDASE	354	ASN	353	ND2	3.010	5.049
4ENL	1.90	ENOLASE	327	ASN	326	ND2	3.042	4.882
3ENL	2.25	ENOLASE	265	ASN	264	ND2	3.050	3.490
1COX	1.80	CHOLESTEROL OXIDASE	383	ASN	382	ND2	3.096	5.209
1MEE	2.00	EGLIN-C	225-A	ASN	123-A	ND2	3.124	4.868
2PKA	2.05	KALLIKREIN A	49-A	ASN	48-A	ND2	3.153	5.078
6CPA	2.00	CARBOXYPEPTIDASE A	60	ASN	188	ND2	3.169	5.539
1TNF	2.60	TUMOR NECR FACTOR A	20-B	ASN	19-B	ND2	3.208	5.137
4P2P	2.40	PHOSPHOLIPASE A2	68	ASN	67	ND2	3.231	5.183
2ER6	2.00	ENDOTHIAPEPSIN	224-E	ASN	300-E	ND2	3.284	5.925
1CSE	1.20	SUBTILISIN CARLSBERG	225-E	ASN	123-E	ND2	3.317	5.324
8CAT	2.50	CATALASE	171-B	ASN	170-B	ND2	3.323	4.935
3PSG	1.65	PEPSINOGEN	33	ASN	32	ND2	3.451	5.299
1FD2	1.90	FERREDOXIN	50	CYS	49	SG	3.431	5.216
3RN3	1.45	RIBONUCLEASE A	117	CYS	58	SG	3.469	6.242
1TIE	2.50	TRYPSIN INHIBITOR	40	CYS	39	SG	3.477	4.455

3ER3	2.00	ENDOTHIAPEPSIN	277-E	CYS	283-E	SG	3.489	5.862
2SEC	1.80	SUBTILISIN CARLSBERG	44-I	GLN	45-I	NE2	3.162	4.833
2MCG	2.00	IG BENCE JONES PROT.	46-2	GLN	40-2	NE2	3.282	4.045
1MBC	1.50	MYOGLOBIN	100	GLN	152	NE2	3.454	4.513
2CNA	2.00	CONCANAVALIN A	222	SER	225	OG	2.785	2.863
3RP2	1.90	RAT MAST CELL PROT 2	126-A	SER	125-A	OG	2.854	5.432
2ER6	2.00	ENDOTHIAPEPSIN	133-E	SER	132-E	OG	2.971	5.409
5RUB	1.70	RUBISCO	297-B	SER	296-B	OG	2.974	5.230
2PKA	2.05	KALLIKREIN A	111-Y	SER	110-Y	OG	2.988	3.164
2HFL	2.54	IGG1 FAB FRAGMENT	8-L	SER	7-L	OG	3.015	5.375
3PAL	2.40	PARVALBUMIN	72	SER	71	OG	3.018	5.160
3BLM	2.00	BETA LACTAMASE	174	SER	173	OG	3.019	5.246
3ICD	2.50	ISOCITR. DEHYDROGENASE	140	SER	139	OG	3.047	5.416
5MBA	1.90	MYOGLOBIN	58	SER	57	OG	3.077	5.059
2LZM	1.70	LYSOZYME	37	SER	36	OG	3.079	5.417
1TIM	2.50	TRIOSE PHOSPHATE ISOM.	80-A	SER	79-A	OG	3.087	5.264
8CAT	2.50	CATALASE	346-A	SER	345-A	OG	3.108	5.448
1BBP	2.00	BILIN BINDING PROT.	157-A	SER	156-A	OG	3.111	5.084
2LTN	1.70	LECTIN	13-A	SER	12-A	OG	3.126	5.375
1GP1	2.00	GLUTATHIONE PEROXIDASE	150-A	SER	149-A	OG	3.132	5.268
1PFK	2.40	PHOSPHOFRUCTOKINASE	256-A	SER	255-A	OG	3.195	3.584
1TIE	2.50	TRYPSIN INHIBITOR	47	SER	46	OG	3.232	5.142
3RNT	1.80	RIBONUCLEASE T1	73	SER	72	OG	3.279	5.576
1PBX	2.50	HEMOGLOBIN	54-A	SER	53-A	OG	3.280	5.398
2RNT	1.80	RIBONUCLEASE T1	55	SER	54	OG	3.299	5.240
1PCY	1.60	PLASTOCYANIN	86	SER	85	OG	3.304	6.893
4PEP	1.80	PEPSIN	271	SER	270	OG	3.338	5.091
2MCG	2.00	IG BENCE JONES PROT.	158-2	SER	157-2	OG	3.352	4.570
1CPC	1.66	C-PHYCOCYANIN	126-A	SER	125-A	OG	3.370	5.422
2LBP	2.40	LEUCINE BINDING PROT.	99	SER	98	OG	3.374	5.708
3CNA	2.40	CONCANAVALIN A	202	SER	201	OG	3.386	4.682
1FDL	2.50	LYSOZYME	8-L	SER	7-L	OG	3.389	5.427
2ER7	1.60	ENDOTHIAPEPSIN	149-E	SER	148-E	OG	3.395	5.290
2YHX	2.10	HEXOKINASE B	270	SER	269	OG	3.419	4.333
3ICB	2.30	CALCIUM-BINDING PROT.	3	SER	2	OG	3.432	5.370
1TIE	2.50	TRYPSIN INHIBITOR	87	SER	86	OG	3.451	5.662
1CHG	2.50	CHYMOTRYPSINOGEN A	8	SER	26	OG	3.460	5.207
1FCB	2.40	FLAVOCYTOCHROME B2	235-B	SER	234-B	OG	3.481	4.807
1BP2	1.70	PHOSPHOLIPASE A2	18	SER	16	OG	3.492	5.588
6TIM	2.20	TRIOSE PHOSPHATE ISOM.	81-B	SER	79-B	OG	3.499	5.411
1PHH	2.30	P-OH-BENZOATE HYDROXYLASE	36	THR	35	OG1	2.655	5.026
8CAT	2.50	CATALASE	150-A	THR	149-A	OG1	2.734	5.272
3P2P	2.10	PHOSPHOLIPASE A2	37-B	THR	36-B	OG1	2.998	4.136
2MCG	2.00	IG BENCE JONES PROT.	186-1	THR	185-1	OG1	3.008	4.627
8ABP	1.49	L-ARABINOSE BINDING PROT	66	THR	65	OG1	3.049	5.377
8I1B	2.40	INTERLEUKIN-1 BETA	78	THR	77	OG1	3.081	3.601
2HAD	1.90	HALOALKANE DEHALOGENASE	182	THR	181	OG1	3.126	5.885
1CPC	1.66	C-PHYCOCYANIN	4-A	THR	3-A	OG1	3.128	4.795
3FAB	2.00	IGG1 FAB' FRAGMENT	166-L	THR	165-L	OG1	3.163	3.577
1CPC	1.66	C-PHYCOCYANIN	125-L	THR	124-L	OG1	3.172	4.223
2TMN	1.60	THERMOLYSIN	277-E	THR	276-E	OG1	3.261	5.529
2PKA	2.05	KALLIKREIN A	219-Y	THR	218-Y	OG1	3.267	5.052
4GPB	2.30	GLYC. PHOSPHORYLASE B	488	THR	487	OG1	3.323	5.560
3RP2	1.90	RAT MAST CELL PROT 2	116-A	THR	115-A	OG1	3.333	4.842
1THB	1.50	HEMOGLOBIN	51-B	THR	50-B	OG1	3.348	5.063
1PPD	2.00	PAPAIN	15	THR	14	OG1	3.381	3.515
4TIM	2.40	TRIOSE PHOSPHATE ISOM.	180-A	THR	179-A	OG1	3.384	4.937
1TEC	2.20	THERMITASE	3-E	THR	2-E	OG1	3.406	3.948

2TEC	1.98	EGLIN C	3-E	THR	2-E	OG1	3.427	3.785
8ADH	2.40	ALCOHOL DEHYDROGENASE	60	THR	59	OG1	3.436	4.780
1CHO	1.80	CHYMOTRYPSIN ALPHA	152-E	THR	151-E	OG1	3.441	4.089
1THB	1.50	HEMOGLOBIN	119-A	THR	118-A	OG1	3.464	5.169
1TIM	2.50	TRIOSE PHOSPHATE ISOM.	178-A	THR	177-A	OG1	3.464	5.094
1CSC	1.70	CITRATE SYNTHASE	89	THR	233	OG1	3.466	5.359
2TEC	1.98	EGLIN C	205-E	THR	184-E	OG1	3.469	5.704
4XIA	2.30	XYLOSE ISOM.	6-B	THR	5-B	OG1	3.481	5.319
6APR	2.50	RHIZOPUSPEPSIN	139-E	THR	138-E	OG1	3.495	5.248

J. Mol. Biol. (1993) **230**, 543–574

Backbone-dependent Rotamer Library for Proteins

Application to Side-chain Prediction

Roland L. Dunbrack Jr and Martin Karplus

Department of Chemistry
Harvard University
Cambridge, MA 02138, U.S.A.

(Received 24 July 1992; accepted 26 October 1992)

A backbone-dependent rotamer library for amino acid side-chains is developed and used for constructing protein side-chain conformations from the main-chain co-ordinates. The rotamer library is obtained from 132 protein chains in the Brookhaven Protein Database. A grid of 20° by 20° blocks for the main-chain angles ϕ,ψ is used in the rotamer library. Significant correlations are found between side-chain dihedral angle probabilities and backbone ϕ,ψ values. These probabilities are used to place the side-chains on the known backbone in test applications for six proteins for which high-resolution crystal structures are available. A minimization scheme is used to reorient side-chains that conflict with the backbone or other side-chains after the initial placement. The initial placement yields 59% of both χ_1 and χ_2 values in the correct position (to within 40°) for thermolysin to 81% for crambin. After refinement the values range from 61% (lysozyme) to 89% (crambin). It is evident from the results that a single protein does not adequately test a prediction scheme.

The computation time required by the method scales linearly with the number of side-chains. An initial prediction from the library takes only a few seconds of computer time, while the iterative refinement takes on the order of hours. The method is automated and can easily be applied to aid experimental side-chain determinations and homology modeling. The high degree of correlation between backbone and side-chain conformations may introduce a simplification in the protein folding process by reducing the available conformational space.

Keywords: proteins; side-chains; rotamers; prediction; conformation

1. Introduction

An understanding of the conformations of side-chains is required for the analysis of protein folding and for the prediction of protein tertiary structure. Prediction methods can also be used in the structure determination of proteins from X-ray crystallography and nuclear magnetic resonance spectroscopy by providing a procedure for the initial placement of side-chains. They form part of any scheme to predict the structure of a protein from data for homologous proteins. Early work based on structural surveys (Janin *et al.*, 1978; Bhat *et al.*, 1979) and energy calculations (Gelin & Karplus, 1975, 1979), indicated that the side-chain dihedral angles in proteins generally corresponded to the potential energy minima of the isolated amino acid. In fact, as crystal structures have improved, a decreasing number of side-chains have been observed to deviate significantly from one of the isolated amino acid minima (Bhat *et al.*, 1979; James & Sielecki, 1983; Ponder & Richards, 1987). While some of the narrowing of the distributions may be caused by rotamer preferences introduced in modern refinement programs such as PROLSQ (Konnert & Hendrickson, 1980), the weighting factors are usually quite weak and are unlikely to dominate the experimental data in high-resolution structures.

Ponder & Richards (1987) determined the distributions of side-chain dihedral angle $\{\chi_1,\chi_2\}$ pairs for the amino acid residues from a set of ten proteins whose X-ray structures had been determined at a resolution of 2 Å or better (1 Å = 0·1 nm). They found that most side-chains are limited to a small number of the many possible $\{\chi_1,\chi_2\}$ minima. For example, while the leucinyl residue has nine possible $\{\chi_1,\chi_2\}$ conformers, two of these (g^+t and tg^-) account for 88% of the leucinyl residues in the survey. With a database of 61 protein structures, McGregor *et al.* (1987) found that certain side-chains exhibit rotamer preferences that depend on the main-chain secondary structure. For example, Trp has 75% of its χ_1 values near 180° in α-helices, while 62% of the χ_1 values are near −60° in β-sheets.

With an extended database (132 polypeptide

0022-2836/93/060543-32 $08.00/0

chains in 126 crystal structures at a resolution of 2·0 Å or better), it is possible to make a more detailed analysis of the relation between the backbone dihedral angles ϕ and ψ of an amino acid and the side-chain dihedral angle distributions. By examining all side-chain dihedral angles for all amino acids, we have found that there is a significant correlation between the backbone ϕ,ψ values and the side-chain dihedral angles, which goes beyond a correlation with secondary structure. Blocks corresponding to a 20° by 20° grid in ϕ and ψ yield meaningful probabilities for the χ values (χ_1,χ_2 . . .) of most of the amino acids. In some cases the database is not sufficient to determine the ϕ,ψ dependent probabilities. We shall show elsewhere that energy calculations for isolated dipeptides generally are in accord with the observed preferences. In this paper, we describe the results obtained for the side-chain dihedral angle distributions of the amino acids and demonstrate that such a "backbone-dependent rotamer library" is very useful in providing starting positions for predicting side-chain conformations of proteins.

A variety of methods have been suggested for determining side-chain conformations. The type of method that is appropriate depends, in part, on the complexity of the problem to be solved. For single-site mutations, a detailed energy function search of the conformational space available to the mutant side-chain (Shih *et al.*, 1985) can be made to determine its position. Also, free energy simulations can be used to introduce mutant side-chains (Tidor & Karplus, 1991). Good overall results are expected, since it has been shown (Gelin & Karplus, 1979) that potential energy functions of the molecular mechanics type are adequate for representing the interactions of buried side-chains. For surface side-chains, it was found that solvent and interactions with neighboring proteins in the crystal must be included. In contrast to their behavior in a crystal environment, surface side-chains in solution are likely not to have a unique orientation. Nuclear magnetic resonance studies of protein structures (Wüthrich, 1989) indicate that such flexibility is often present. The most detailed procedure for studying surface side-chains is to do free-energy mapping of the (χ_1,χ_2 . . .) angle distribution in the presence of an explicit model for the solvent (Straatsma & McCammon, 1992; Kuczera *et al.*, unpublished results). Also, additional energy terms can be introduced in molecular mechanics programs to approximate dielectric effects, the hydrophobic effect, and solvent structure around ionic and polar functional groups (Pettitt & Karplus, 1985; Schiffer *et al.*, 1992; Wesson & Eisenberg, 1992). A method such as CONGEN (Bruccoleri & Karplus, 1987) searches the conformational space to build the backbone and side-chains for limited regions of proteins (e.g. the hypervariable loops of antibodies). Lee & Subbiah (1991) have used a computationally intensive, simulated annealing approach and a van der Waals repulsive potential to predict the side-chain positions in proteins, given the backbone co-ordinates. Holm & Sander (1991) used backbone segments from a structural database to build full backbone co-ordinates from C^α co-ordinates, and then utilized the database of Tuffery *et al.* (1991) and simulated annealing to place side-chains. Several groups have used backbone co-ordinates to determine initial side-chain placements. Kabsch *et al.* (1990) and Wendoloski & Salemme (1992) searched the database for each side-chain to find a local backbone fold (plus and minus 1 or more amino acid residues) similar to the fold of the protein to be modeled. The side-chain was then placed according to the best such fragment or the most commonly found rotamer. Reid & Thornton (1989) built full backbone co-ordinates of flavodoxin from C^α co-ordinates with a method similar to that of Holm & Sander (1991), but they used the secondary-structure dependent rotamer library of McGregor *et al.* (1987) to predict side-chain positions. When clashes were observed, other common rotamer positions were tested and energy minimized. Desmet *et al.* (1992) have suggested that side-chain placement can be simplified based on the idea that side-chain rotamers can be excluded by pairwise searches and used the method for predicting the side-chains conformations from the known backbone structure starting with the Ponder & Richards (1987) rotamers.

The method described here for predicting side-chain conformations is most closely related to that proposed by Summers & Karplus (1989). In that approach, which was developed as part of a homology modeling scheme (Summers & Karplus, 1990), the side-chains are placed in accord with the known χ angles of the residues in a protein homologous to that being modeled. When steric clashes were observed in the initial placement, side-chain conformations were altered by use of a rigid rotation energy search of the conformational space of individual side-chains. A number of rules were formulated to determine which residue of a pair of clashing side-chains should be altered, depending on the amino acid type, its accessibility, whether or not it is identical to a template side-chain, its participation in hydrogen bonds in the template protein, etc. Residues or side-chain atoms for which there was no information in the template protein were added one at a time and placed according to rigid rotation energy search. The method was rather successful (92% for χ_1, 81% for χ_2) in building the side-chains of the C-terminal lobe of rhizopuspepsin on its backbone from the side-chain positions of the homologous C-terminal lobe of penicillopepsin (39% sequence identity).

The procedure used in this paper is designed to predict all of the side-chains from a knowledge of the backbone co-ordinates. Thus, it is concerned with the same problem as that studied by Lee & Subbiah (1991) and by Desmet *et al.* (1992). Because most of the calculations in the present method deal with one side-chain at a time, the time required scales linearly with the size of the system. The method is faster and more accurate than those of

Lee & Subbiah (1991) and Desmet *et al.* (1992). Also, it can be run on most workstations, an advantage over the approach of Lee & Subbiah (1991), which requires a large memory and is not suitable for bigger proteins such as thermolysin (316 residues). Side-chains can be built on known protein backbone co-ordinates, those optimized from a homologous protein template (Sali *et al.*, 1990), or those determined from some predictive scheme (e.g. starting with C^{α} co-ordinates). The essential new element in the method is that the side-chains are placed simultaneously with the aid of the backbone-dependent rotamer library. As we demonstrate, this provides considerably more information than averaged rotamer libraries (e.g. that of Ponder & Richards, 1987) and so yields a much improved starting set of side-chain positions. If the structure of a homologous protein is known, information about the side-chains of the target structure can be incorporated from the template. Once the initial placement has been made, the optimization procedure follows the philosophy of Summers & Karplus (1989), though some of the methodological details are significantly different. One consequence of these differences is that automation of the method is more straightforward. This is important because it is difficult not to be biased if human decision-making is required, particularly in test applications to known structures. Further, since there are many applications of the method, the less human labor involved in performing a prediction the better.

In the next section of this paper, we present the procedure used to calculate the backbone-dependent rotamer library, and then describe the scheme for setting up the initial side-chain positions and refining them to a final prediction. We also present various ways for evaluating the results of the side-chain predictions since no single criterion is adequate. The following section describes the results. Details of the backbone-dependent rotamer library are given. Full side-chain predictions for six proteins from the known backbone are presented. The proteins chosen for study are thermolysin (PDB code 3tln), ribonuclease A (7rsa), bovine pancreatic trypsin inhibitor (5pti), lysozyme (1lz1), crambin (1crn), and the C-terminal domain of rhizopuspepsin (2apr). Several of these proteins have been used to test other prediction methods. In addition, we apply the method to the penicillopepsin to rhizopuspepsin homology modeling problem, so as to be able to compare the present results with the approach of Summers & Karplus (1989). In the final section, we discuss the potential of the method and implications of the results for protein folding.

2. Methods

(a) *The ϕ,ψ rotamer library*

The library was calculated from the structures of 132 protein chains in 126 structures in the Brookhaven Protein Database refined at a resolution better than or equal to 2·0 Å. These proteins are listed in Table 1. Included in these 126 structures are 17 preliminary PDB files available by ftp from Brookhaven (at the Internet address: pdb.pdb.bnl.gov), which have allowed us to extend significantly the database from which the library is calculated. Several groups of homologous proteins are included in the list of structures. While proteins that are identical or nearly identical in sequence have not been included, homologous proteins have been included to increase the size of the database. The structures that are used have been chosen on the basis of several criteria: resolution; date of deposit in the database, in that later structures are likely to be better; and the absence of non-protein ligands that might alter side-chain positions in unpredictable ways. For the prediction of the six proteins described below, the rotamer libraries were determined after removing the protein and its homologues from the list. Thus, in effect, six separate rotamer libraries were calculated. Since the libraries are very similar, only the library calculated with all the proteins listed in Table 1 is described in Results. The backbone ϕ and ψ values were divided into $20° \times 20°$ blocks ($-180°$ to $-160°$, $-160°$ to $-140°$, etc. for ϕ and ψ), and the rotamer library was calculated for each $20° \times 20°$ block. Because of the small block size and steric constraints on the backbone, some regions of the ϕ,ψ map are underpopulated or even empty. Tests with coarser or variable grids confirm the present choice. Rotamer populations for each χ_i ($i = 1,2,3,4$) were calculated using the angular ranges listed in Table 2. For all side-chains (except Ala, Pro and Gly), the χ_1 values correspond to the rotamers of a tetrahedral carbon atom. They were divided into bins of $-120°$ to $0°$ (g^+ conformer), $0°$ to $120°$ (g^- conformer), and $120°$ to $240°$ (t conformer). The same limits were used for the dihedral angle χ_2 of all amino acids that have a χ_2, except for proline, the aromatics, asparagine, and aspartic acid. For proline, χ_1 was placed into 2 bins; $\chi_1 < 0°$ and $\chi_1 > 0°$ corresponding to the 2 proline conformations, C^{γ}-*exo* and C^{γ}-*endo*, respectively. The angle χ_2 of proline was treated analogously. The χ_2 values of phenylalanine, tyrosine and histidine were divided into bins of $0°$ to $60°$, $60°$ to $120°$ and $120°$ to $180°$, even though the expected value is near $\pm 90°$. These values were used to determine whether there were any significant populations more than $30°$ from the usual χ_2 value near $90°$. In well-populated areas of the map, there were no statistically significant deviations from $90°$. If χ_2 was less than $0°$, a χ_2 value of $\chi_2 + 180°$ was used. This is exact for Phe and Tyr, and generally true of His, since most crystal structures do not clearly distinguish whether a given His has a value of χ_2 or $\chi_2 + 180°$. Similarly, for Asp and Asn, χ_2 and $\chi_2 + 180°$ were treated as equivalent, and the limits used were $-90°$ to $-30°$ (g^+ conformer), $-30°$ to $30°$ (t conformer), and $30°$ to $90°$ (g^- conformer). Trp χ_2 was treated as either $0° < \chi_2 < 180°$ or $-180° < \chi_2 < 0°$. For the amino acids with flexible χ_3 and χ_4 dihedral angles (Lys, Arg, Glu, Gln), analogous ranges were used; i.e. the same limits as described for χ_1 were employed, except for χ_3 of Glu and Gln, where the limits described for Asp and Asn χ_2 were used.

(b) *Prediction method*

To make clear the procedure used in generating the side-chain positions, the steps involved are listed in Fig. 1. Explanatory comments on the various steps are given in what follows.

(i) *Construction of initial model*

(i)(a) *Backbone co-ordinates*

One is starting with a model of the backbone, which is either derived from the Cartesian co-ordinates of a target

Table 1

List of Protein Databank files used in backbone-dependent rotamer library

Name	Date	Code-Chain	Resolution (Å)
Protease inh. dom. of Alzheimer's amyloid	SEP90	1AAP-A	1·5
Actinoxanthin	DEC82	1ACX	2·0
Adenylate kinase isoenzyme-3	JAN90	1AK3-A	1·9
Alpha-lactalbumin	AUG89	1ALC	1·7
Aldolase A	MAY91	1ALD	2·0
Bilin binding protein	SEP90	1BBP-A	2·0
Carbonic anhydrase	FEB89	1CA2	2·0
Cytochrome c	MAR83	1CCR	1·5
Superoxide dismutase (co substituted)	FEB92	P1COB-A	2·0
Cholesterol oxidase	FEB91	1COX	1·8
Crambin	APR81	1CRN	1·5
Citrate synthase-L-malate	MAY90	1CSC	1·7
Subtilisin Carlsberg complex eglin-c	JUN88	1CSE-E	1·2
Subtilisin Carlsberg complex eglin-c	JUN88	1CSE-I	1·2
L7/L12 50 S ribosomal protein	SEP86	1CTF	1·7
Defensin	JAN91	P1DFN-A	1·9
Hemoglobin (erythrocruorin, deoxy)	MAR79	1ECD	1·4
FK506 binding protein complex	MAY91	1FKF	1·7
Gamma-II crystallin	AUG85	1GCR	1·6
Holo-D-glyceraldehyde-3-phos. dehydrogenase	JUN87	1GD1-O	1·8
Guanylate kinase	DEC91	P1GKY	2·0
Glycolate oxidase	JUN89	1GOX	2·0
Glutathione peroxidase	JUN85	1GP1-A	2·0
Oxidized high potential iron protein	APR75	1HIP	2·0
Human neutrophil elastase	APR89	1HNE-E	1·84
Alpha-amylase inhibitor HOE-467 A	JAN89	1HOE	2·0
Intestinal fatty acid binding protein	DEC90	1IFB	1·96
Lysozyme (mutant)	MAY91	1L58	1·65
Leghemoglobin (deoxy)	APR82	1LH4	2·0
Lambda repressor-operator complex	NOV91	P1LMB-A	1·8
Myoglobin (deoxy, pH 8·4)	AUG81	1MBD	1·4
Mesentericopeptidase	APR91	P1MEE-A	2·0
Oncomodulin	APR90	1OMD	1·85
Ovalbumin (egg albumin)	NOV90	P1OVA-A	1·9
Pseudoazurin (oxidized CU++ at pH 6·8)	JUN88	1PAZ	1·55
Human plasminogen Kringle 4	JUL91	P1PK4	1·9
Avian pancreatic polypeptide	JAN81	1PPT	1·37
434 repressor (amino-terminal domain)	DEC88	1R69	2·0
Retinol binding protein	APR90	1RBP	2·0
Rubredoxin	MAR88	1RDG	1·4
Bence-Jones immunoglobulin REI variable	MAR76	1REI-A	2·0
Barnase (G specific endonuclease)	MAR91	P1RNB	1·9
Selenomethionyl ribonuclease H	JUL90	1RNH	2·0
ROP: Col E1 repressor of primer	APR91	P1ROP-A	1·7
Ribonuclease SA	DEC90	P1SAR-A	1·8
Trypsin	APR88	1SGT	1·7
Scorpion neurotoxin (variant 3)	DEC82	1SN3	1·8
Staphylococcal nuclease	JUL89	1SNC	1·65
Trypsinogen	SEP79	1TGN	1·65
Hemoglobin (T state, partially oxygenated)	JAN90	1THB-A	1·5
Hemoglobin (T state, partially oxygenated)	JAN90	1THB-B	1·5
Tonin	JUN87	1TON	1·8
Ubiquitin	JAN87	1UBQ	1·8
Uteroglobin (oxidized)	APR89	1UTG	1·34
Iso-2-cytochrome c (reduced state)	OCT91	P1YEA	1·9
B-2036 composite cytochrome c (reduced)	OCT91	P1YEB	1·9
Triose phosphate isomerase	JAN90	1YPI-A	1·9
Cytochrome B562 (oxidized)	JAN90	256B-A	1·4
Actinidin (sulfhydryl proteinase)	NOV79	2ACT	1·7
Alpha-lytic protease	MAR85	2ALP	1·7
Acid proteinase (rhizopuspepsin)	MAR87	2APR	1·8
Azurin (oxidized)	OCT86	2AZA-A	1·8
Cytochrome c (prime)	AUG85	2CCY-A	1·67
Cytochrome c-3	NOV83	2CDV	1·8
Chymotrypsinogen A	JAN87	2CGA-A	1·8
Chymotrypsin inhibitor 2	SEP88	2CI2-I	2·0
Concanavalin A	APR75	2CNA	2·0
Cytochrome P450cam (camphor monooxygenase)	APR87	2CPP	1·63
Cytochrome c peroxidase	AUG85	2CYP	1·7
Endothia aspartic proteinase	NOV90	2ER7-E	1·6

Table 1 *(continued)*

Name	Date	Code-Chain	Resolution (Å)
Immunoglobulin FAB	APR89	2FB4-H	1·9
Immunoglobulin FAB	APR89	2FB4-L	1·9
Flavodoxin	FEB92	2PCR	1·8
D-galactose/D-glucose binding protein	FEB89	2GBP	1·9
Hemerythrin (met)	OCT90	2HMQ-A	1·66
Hemoglobin V (cyano, met)	AUG85	2LHB	2·0
Pea lectin	JUN90	2LTN-A	1·7
Pea lectin	JUN90	2LTN-B	1·7
Myohemerythrin	APR87	2MHR	1·7
Melittin	OCT90	2MLT-A	2·0
Prealbumin (human plasma)	SEP77	2PAB-A	1·8
Proteinase K	NOV87	2PRK	1·5
Lys 25-ribonuclease T1	JUL88	2RNT	1·8
Rous sarcoma virus protease	OCT89	2RSP-A	2·0
Sarcoplasmic calcium binding protein	AUG91	P2SCP-A	2·0
Cu, Zn superoxide dismutase	MAR80	2SOD-B	2·0
Thermitase complex with eglin	OCT90	2TEC-E	1·98
Thermolysin complex	JUN87	2TMN-E	1·6
Thioredoxin	MAR90	2TRX-A	1·68
Thymidylate synthase complex	JUL91	2TSC-A	1·97
Trp repressor (orthorhombic form)	DEC87	2WRP-R	1·65
GCN4 leucine zipper	JUL91	P2ZTA-A	1·8
Acid proteinase (penicillopepsin)	NOV90	3APP	1·8
Cytochrome B5 (oxidized)	JAN90	3B5C	1·5
Bacteriochlorophyll-A protein	JUN87	3BCL	1·9
Beta-lactamase	DEC90	3BLM	2·0
Cytochrome c-2 (reduced)	NOV83	3C2C	1·68
Chloramphenicol acetyltransferase A	JUL90	3CLA	1·75
Erabutoxin B	JAN88	2EBX	1·4
Native elastase	SEP87	3EST	1·65
Basic fibroblast growth factor	JAN92	3FGF	1·6
Glutathione reductase	FEB88	3GRS	1·54
Rat mast cell protease II	SEP84	3RP2-A	1·9
Proteinase A	MAY90	3SGA-E	1·8
Proteinase B from streptomyces griseus	JAN83	3SGB-E	1·8
Proteinase B from streptomyces griseus	JAN83	3SGB-I	1·8
Cytochrome c-551 (reduced)	JUL81	451C	1·6
Prophospholipase A-2	SEP90	4BP2	1·6
Calcium-binding parvalbumin	OCT89	4CPV	1·5
Enolase	NOV90	P4ENL	1·9
Ferredoxin	JUN88	4FD1	1·9
Interleukin-1 beta	MAR90	4I1B	2·0
Bovine calbindin D9K (minor A form)	AUG91	P4ICB	1·6
Pepsin	DEC89	4PEP	1·8
Beta trypsin, diisoprophylphosphoryl	APR88	4PTP	1·34
Carboxypeptidase A-alpha (Cox)	MAY82	5CPA	1·54
HIV-1 protease complex	APR90	5HVP-A	2·0
C-H-RAS P21 protein (amino acids 1-166)	APR90	5P21	1·35
Parvalbumin (alpha lineage)	SEP91	P5PAL	1·5
Trypsin inhibitor (crystal form II)	OCT84	5PTI	1·0
Rubisco (ribulose-1,5-bisphosphate)	MAY90	5RUB-A	1·7
Troponin-C	MAY88	5TNC	2·0
M-4 apo-lactate dehydrogenase	NOV87	6LDH	2·0
D-xylose isomerase	SEP90	6XIA	1·65
Plastocyanin	SEP89	7PCY	1·8
Ribonuclease A (phosphate-free)	JUN88	7RSA	1·26
L-arabinose-binding protein (mutant)	APR91	8ABP	1·49
Dihydrofolate reductase	MAY89	8DFR	1·7
Insulin	OCT91	9INS-A	1·7
Insulin	OCT91	9INS-B	1·7
Papain (Cys-25 oxidized)	MAR86	9PAP	1·65
Wheat germ agglutinin (isolectin 2)	APR90	9WGA-A	1·8

Name is derived from COMPND records in the PDB files; Date is from the HEADER records; Resolution is from the REMARK records. The code in the Protein Databank Code is prefixed by P if the file is a preliminary entry, available by anonymous ftp from the Brookhaven National Labs (pdb.pdb.bnl.gov). The chain used from each file is appended to the code; if there is no chain indicated, then the single chain in the file is used.

Table 2
Limits for rotamer library χ angles

A. *Ser, Thr, Cys, Val, Phe, His, Tyr*

	χ_1 limits
1	0° → 120°
2	120° → 240°
3	−120° → 0°

B. *Lys, Arg, Met, Gln, Glu, Ile, Leu*

	χ_1 limits	χ_2 limits
1	0° → 120°	0° → 120°
2	0° → 120°	120° → 240°
3	0° → 120°	−120° → 0°
4	120° → 240°	0° → 120°
5	120° → 240°	120° → 240°
6	120° → 240°	−120° → 0°
7	−120° → 0°	0° → 120°
8	−120° → 0°	120° → 240°
9	−120° → 0°	−120° → 0°

C. *Trp*

	χ_1 limits	χ_2 limits
1	0° → 120°	0° → 180°
3	0° → 120°	−180° → 0°
4	120° → 240°	0° → 180°
6	120° → 240°	−180° → 0°
7	−120° → 0°	0° → 180°
9	−120° → 0°	−180° → 0°

D. *Asp, Asn*

	χ_1 limits	χ_2 limits
1	0° → 120°	−90° → −30°
2	0° → 120°	−30° → 30°
3	0° → 120°	30° → 90°
4	120° → 240°	−90° → −30°
5	120° → 240°	−30° → 30°
6	120° → 240°	30° → 90°
7	−120° → 0°	−90° → −30°
8	−120° → 0°	−30° → 30°
9	−120° → 0°	30° → 90°

E. *Pro*

	χ_1 limits	χ_2 limits
1	0° → 60°	−60° → 0°
3	−60° → 0°	0° → 60°

χ_1 and χ_2 ranges are given for each defined rotamer for the amino acid side-chains. The numbers in the left-hand column are used in Fig. 2 to illustrate the preferred rotamers in different positions on the ϕ,ψ map. The limits for χ_3 and χ_4 are described in the text.

Backbone coordinates from target or homologous template protein

Sidechain (sc) placement from library and/or template protein

Disulfide minimization

Hydrogen atom minimization -> **Structure 0**

van der Waals clashes (sc's with backbone)

Sidechain minimizations for sc's which clash with backbone

Sidechain placement

Disulfide minimization

Hydrogen atom minimization -> ***Structure* 1**

van der Waals clashes (all atoms, except Val, Ile, Thr sc's)

Sidechain minimizations for sc's (except Val,Ile,Thr) which clash

Sidechain placement

Disulfide minimization

Hydrogen atom minimization -> **Structure 2**

van der Waals clashes (all atoms)

Sidechain minimizations for all sc's which clash with other atoms

Sidechain placement

Disulfide minimization

Hydrogen atom minimization -> **Structure 3**

Repeat until all clashes are resolved -> ***Structure* 4,5,6,...,N**

Figure 1. Outline of the method. Steps in the procedure for placing side-chains (sc) from the library and for resolving van der Waals conflicts between the side-chains and the backbone and other side-chains.

structure (e.g. a preliminary X-ray or nuclear magnetic resonance determined backbone) or from Cartesian co-ordinates from the experimental structure of a template, such as a homologous protein. If the template and target are of different lengths, some portion of the backbone must be added or deleted. There are a variety of methods for doing this, which are based either on database searches and template fitting (Summers & Karplus, 1990) or an energy function-based conformational search (e.g. Bruccoleri & Karplus, 1987) or a combination of the two.

(i)(b) *Side-chain placement*

Information about the initial placement of the side-chains either comes from the rotamer library alone or from the homologous template protein in combination with the rotamer library. The possible choices for starting co-ordinates are listed in Table 3. When the side-chain information comes from the rotamer library, the information is necessarily in the form of internal co-ordinates (bond lengths, bond angles and dihedral angles). In this case, bond lengths and angles from CHARMM minimized structures (Brooks *et al.*, 1983) are used for the side-chain in the tetrapeptide Acetyl-Ala-Xxx-Ala-$NHCH_3$; these have been calculated for all amino acids and are now used in the CHARMM program residue topology file. Since we are using the all-hydrogen atom parameter set (MacKerell *et al.*, unpublished results), both heavy atom and hydrogen atom bond lengths and angles were determined by the tetrapeptide minimizations just described. In previous work (Summers & Karplus, 1989), the polar hydrogen set was used, and bond length and angle information from CHARMM parameters without minimization were employed. The minimized structures provide a more accurate reflection of likely side-chain structures. Alternatively, one could use averaged bond lengths and angles from a database.

The initial side-chain dihedral angles for a given amino acid are determined from the backbone-dependent rotamer library by the following procedure. The most likely value of χ_1 for the 20° by 20° block corresponding to the backbone ϕ and ψ values for that residue is used; for that value of χ_1, the most common value of χ_2 is used. This is usually the same as picking the most common $\{\chi_1,\chi_2\}$ conformation for the side-chain, corresponding to a given ϕ,ψ, from columns 10 to 18 of Table 4 (see the legend to Table 4 for an explanation of the columns), but in some cases it is different. For example, consider the case in which g^- and g^+ have populations of 40% and 60%, respectively, for χ_1 (columns 7 and 9), but χ_2 is divided evenly between 2 conformations for $\chi_1 = g^+$, (say, g^- and

Table 3
Input data

Name of method	Backbone coor.	Side-chain dihedrals	Side-chain Bond lgths and ang.
targ/lib	Target	Library	Minimized tetramer
temp/lib	Template	Library	Minimized tetramer
targ/temp	Target	Template	Minimized tetramer
		+library	Minimized tetramer
temp/temp	Template	Template:	
		Identical sc	Template Cartesian co-ordinates
		Non-identical sc	Minimized tetramer
		+library	Minimized tetramer

Backbone co-ordinate information can come either from a homologous template protein or from the target protein whose side-chain conformations are to be predicted. Side-chain (sc) dihedral information comes either from the template or from the library either in the form of Cartesian co-ordinates or internal co-ordinates. Bond lengths (lgths) and angles (ang.) come either from the tetramer Ace-Ala-Xxx-Ala-$NHCH_3$, minimized for each possible side-chain (in the form of internal co-ordinates), or from the Cartesian co-ordinates of the template source protein.

t (columns 16 and 17) and there is only 1 conformation for $\chi_1 = g^-$ (say, t, (column 11))). The probabilities for the 3 conformations are 30% (g^+,g^-: column 16), 30% (g^+,t: column 17), and 40% (g^-,t: column 11). If one uses the most common conformation (g^-,t) for χ_1 and χ_2, one chooses the less common value of χ_1. It is better to use one of the conformations of χ_2 corresponding to $\chi_1 = g^+$, the more common rotamer, since if χ_1 is wrong then the value of χ_2 is not really meaningful.

If the number of side-chains in a particular block of the ϕ,ψ map is smaller than 4, the most common χ angle values for the side-chain obtained from a backbone-independent rotamer library is chosen. (The statistics for rotamer preferences independent of the backbone are listed in Table 5. These are discussed in Results.) For all side-chains, except Ser, Thr, Val and Pro, this sets χ_1 equal to $-60°$. For Ser and Thr, the most common χ_1 value is $+60°$, and for Val it is 180°. For proline, the C^γ-*endo* structure for the ring is chosen, with $\chi_1 = +28°$ since this is the average value for χ_1 in the C^γ-*endo* conformation. The most common χ_2 values are 180° (as they are for χ_3 and χ_4) except for aromatic χ_2 terms, which are 90° for Tyr, Phe, His and Trp. For Asn, the preferred χ_1,χ_2 conformation is $-60°,-60°$. These preferred conformations match the preferences calculated from a much smaller database by Ponder & Richards (1987). The only exception is for Met, where Ponder & Richards (1987) list the $-60°,-60°$ conformation as preferred from a sample of 16 residues. The present library contains 399 methionine residues, and the $-60°,180°$ conformation is preferred; the probabilities are 34% for $-60°,180°$ *versus* 22% for $-60°,-60°$.

If the structure of a homologous protein is known, it can be used to determine some of the information about the side-chain positions in the target protein. The form of this information depends on whether the target or template backbone is used. In method temp/temp (Table 3), where both the template backbone and side-chains are used in the initial structure, the Cartesian co-ordinates for side-chains that are identical in the template and the target can be used. For non-identical side-chains for which there is information in the template, the dihedral angles are transferred from the template according to the rules of Summers & Karplus (1989), while the bond lengths and angles come from the tetrapeptide minimizations. For most side-chain types, the dihedral angles are transferred directly, unless the transfer is to or from an aromatic residue or from Val to Thr or Ile. In the latter case, χ_1 is set to $\chi_1 - 120°$ of the template, because of the IUPAC definition of χ_1 of Val relative to Thr and Ile (Kendrew *et al.*, 1970). If the target side-chain is aromatic and the template side-chain is not, or *vice versa*, then the target side-chain is placed according to the library. Where there is no information in the template (e.g. Gly, Ala or Pro) or insufficient information (e.g. Ser → Arg) the additional dihedral angles are chosen from the backbone-dependent rotamer library. If the target backbone is used (method targ/temp in Table 3), however, as by Summers & Karplus (1989), then the template side-chain information must be in the form of internal co-ordinates, even for identical side-chains. For all side-chains, bond lengths and angles are obtained from the tetrapeptide minimizations. For identical side-chains, dihedral angles from the templates are used directly; for non-identical side-chains, dihedral angles are transferred as described above. For target side-chains without sufficient information in the template, the library is used.

Finally, the CHARMM residue topology file is used to set up the remaining co-ordinates that are undefined. This includes the Ala side-chains, the backbone hydrogen atoms, and Gly H^α. If there are known (or suspected) disulfide bonds, then these are set up within CHARMM, and the H^γ atoms are deleted. The cysteine S^γ atoms have already been placed according to the library or the template protein structure, and the bond between them is established in this step. They are adjusted further by minimization (see below).

At this point, a full set of Cartesian co-ordinates, including hydrogen atoms, can be generated from the information obtained as described above and summarized in Table 3.

(i)(c) *Disulfide bond minimization*

Cysteinyl residues involved in disulfide bonds are minimized for 100 ABNR steps (Brooks *et al.*, 1983) with the rest of the protein atoms held fixed. This yields the correct S–S bond distance and eliminates bad contacts with other protein atoms.

(i)(d) *Hydrogen atom minimization*

The positions of the hydrogen atoms in the model structure are minimized for 100 steps with the CHARMM program while all the heavy atoms in the protein are fixed. The resulting structure is the initial model (Fig. 1, Structure 0).

(ii) *Refinement of model*

Given the initial model, a series of steps is taken to refine the side-chain conformations. The main-chain coordinates are kept fixed throughout. A CHARMM calculation (Brooks *et al.*, 1983) is done to determine all side-chain atoms that have positive van der Waals interactions with any backbone atom or other side-chain atoms. These side-chains are reoriented by an iterative procedure, which first treats clashes with the backbone and subsequently those with other side-chains.

(ii)(a) *Side-chain minimizations (side-chain/backbone clashes)*

Any side-chain that clashes with the backbone and where the energy is above a certain threshold (see below) is examined to find if there are alternative conformations that do not clash with the backbone. Since side-chains that overlap the backbone are most likely to be in the wrong conformation, these side-chains are tested for alternative minima before side-chain–side-chain clashes are resolved. The search for alternative conformations is made by setting $\chi_1,\chi_2,\ldots$ equal to all possible combinations of values at the center of the intervals used for the rotamer library; e.g. for all side-chains except proline, χ_1 is set equal to 60°, 180°, −60° in turn in all possible combinations (3 conformations for side-chains with χ_1 only, 9 conformations for side-chains with χ_1 and χ_2 only, etc.). Aromatic χ_2 terms are set to 0°, 45°, 90°, 135°, etc. to cover the full conformation space. Minimizations are then performed for the given side-chain with all other protein atoms held fixed. Each clashing side-chain is minimized for 100 conjugate gradient steps against the same model (Fig. 1, Structure 0). Minimizations are performed for side-chains when an atom of that side-chain has a van der Waals interaction with an atom of the backbone exceeding the limits (Summers & Karplus, 1989):

Side-chain atom type	Side-chain or backbone atom type	Energy
C, N, O or S	With C or S	>5 kcal/mol
O or N	With O or N	>9 kcal/mol
C, N, O or S	With H	>10 kcal/mol
H	With H	>20 kcal/mol

The O, N/O, N limits are higher than heavy-atom interactions with carbon or sulfur, since these atoms can form hydrogen bond donor/acceptor pairs where the van der Waals repulsions between the heavy atoms can reach nearly 9 kcal/mol (1 cal = 4·184 J), because of the favorable electrostatic contributions in the full potential. The hydrogen atom limits are taken higher still because they can be expected to exhibit greater conformational flexibility.

After all minimizations have been performed for side-chains where there exist clashes with the backbone, the side-chains are simultaneously moved to the lowest energy conformation found for each one. The disulfide bonds and hydrogen atoms are then minimized with the rest of the protein atoms held fixed (see subsections (i)(c) and (i)(d), above). The resulting structure is a new model (Fig. 1, Structure 1).

(ii)(b) *Side-chain minimizations (side-chain–side-chain clashes except Ile, Thr, Val)*

Step (ii)(a) is repeated, except this time clashes between all atoms are included, including those between side-chains. Any residue that involves clashes according to the energetic cutoffs listed in step (ii)(a) is minimized according to the scheme just described, with the exception of Ile, Thr and Val. These are predicted with a high degree of accuracy from the library and it is best not to move them at this stage, since it is likely that the other side-chain involved in the clash is in an incorrect position. The resulting structure is a new model (Fig. 1, Structure 2).

(ii)(c) *Repeated side-chain minimizations (all clashes)*

Step (ii)(b) is repeated as many times as necessary to remove all clashes. If atoms in Ile, Thr or Val clash with any other atoms in the protein, they are moved at this stage according to the usual minimization scheme. The structures resulting from these rounds of reorientation and minimization are referred to as Structure 3, 4, etc. in Fig. 1. If the refinement steps do not remove all the clashes, a simultaneous minimization of the residues involved could be performed. This problem did not arise for any of the proteins studied and the converged model obtained here (Structure N where $N \leq 4$ for the 6 proteins) is the final structure.

(c) *Assessing the results*

There are a number of criteria that can be used to determine the "correctness" of the side-chain orientations in model-building schemes. They involve Cartesian root-mean-square deviations (r.m.s.d.†) of atoms and dihedral angle deviations. As in the work of Summers & Karplus (1989) and Wendoloski & Salemme (1992), we employ a dihedral angle criterion and consider a deviation of less than or equal to ±40° correct, based on the supposition that the predicted and experimental values correspond to the same minimum. r.m.s.d. values by themselves are unsatisfactory because they can lead to misleading results. Small side-chains can have dihedral angles far from the experimental values and still have low r.m.s.d. values. Large side-chains can also have quite small r.m.s.d. values and yet be in a different conformation from the crystal structure. It might be argued that such a structure is "correct", since the side-chain fills essentially the same volume. In low-resolution structures, this could be true, since experimental errors in dihedral angles can be large (e.g. for Val). If, however, the dihedral angles are accurately known from high-resolution structures, it is important to test whether a predictive method is able to determine the dihedral angles. Since we are using high-resolution structures to test the prediction scheme, we emphasize dihedral angle differences, though we also consider r.m.s.d. values, particularly to compare with the results of others.

When citing dihedral angle statistics, there are 2 ways of counting whether a certain χ_2 (or χ_3 or χ_4) is correct, depending on whether the deviation in χ_1 (or χ_2 or χ_3) from the experimental structure is considered. Lee & Subbiah (1991) report χ_2 angle statistics that do not depend on the accuracy of χ_1. Wendoloski & Salemme (1992), by contrast, report χ_{1+2} statistics; i.e. the percentage of residues that have both χ_1 and χ_2 correct (to within 40°). This information is useful, since if χ_1 is far wrong,

† Abbreviations used: r.m.s.d., root-mean-square deviation(s); BPTI, bovine pancreatic trypsin inhibitor.

the Cartesian positions of χ_2 atoms are likely to deviate significantly from their positions in the experimental structure, even if χ_2 is "correct". We report statistics for χ_1 for all side-chains (except Ala and Gly), χ_2 for all side-chains (except Ala, Gly, Ser, Thr, Val and protonated Cys) regardless of whether χ_1 is correct or not, and χ_{1+2} for all side-chains (except Ala, Gly, Ser, Thr and Val, but including cysteinyl residues involved in disulfide bonds, where χ_2 is the dihedral angle determined by atoms C^α, C^β and S^γ of a given cysteinyl residue and S^γ of the other involved in the disulfide bond).

Also, we report r.m.s.d. for each amino acid type determined for the 6 proteins whose side-chain positions have been predicted in order to compare our results with those of Lee & Subbiah (1991). We do not consider r.m.s.d. calculated for all the side-chains of a particular protein, since the results depend on the relative number of large *versus* small side-chains in the sequence.

Statistics are calculated for buried and surface residues separately. Surface residues are defined here as side-chains that have an exposure that is more than 10% of the possible value. Buried residues, conversely, are defined as those with an exposure that is 10% or less of the possible value. The possible exposure is calculated as the surface area determined with a 1·6 Å spherical probe of the side-chain in question in the peptide Acetyl-Xxx-$NHCH_3$, with the backbone dihedral angle ϕ equal to $-60°$, and ψ equal to 140°. The peptide was minimized for 100 ABNR steps using the program CHARMM (Brooks *et al.*, 1983). From the resulting co-ordinates, the total accessible surface area of the side-chain was calculated for all atoms in the side-chain, excluding C^β and H^β atoms.

(d) *Automation of method*

The method is fully automated and has been used on a Convex C220, a Sun Sparcstation, an IBM RS 6000 and a SGI 340. It consists of the backbone-dependent rotamer library, a small number of Unix scripts, 2 FORTRAN programs, and the program CHARMM (version 22). CHARMM is first used to convert the Brookhaven Protein Data Bank (PDB) co-ordinates to CHARMM format. This is followed by a script, which finds and processes the ϕ and ψ values for the protein, and another than produces a file with the sequence of the protein in CHARMM format. If a homologous protein is used to help place the side-chains, the internal co-ordinates in CHARMM format are also calculated for this protein, and the χ angles are processed. A FORTRAN program is then used, in accord with subsection (b)(i)(b), above, to generate a CHARMM script that determines the initial positions of the side-chains, based on the sequence, the backbone dihedral angles, the backbone-dependent rotamer library, and the side-chain positions of a homologous protein (if one is being used). Once the disulfide and hydrogen atom minimizations have been performed, the van der Waals overlaps are calculated. A second FORTRAN program processes the overlaps, and following the rules of subsection (b)(ii)(a), sets up the CHARMM commands to search the alternative side-chain minima. The internal co-ordinates for the new minima are written out by the CHARMM program and used to build a new structure. The procedure continues (subsections (b)(ii)(b) and (b)(ii)(c), above) until all the clashes have been removed. The routines are quite flexible, and a variety of inputs can be used. In some cases (e.g. homology modeling), only a certain number of side-chains need to be modeled into a known structure. The starting structure simply has these side-chains deleted, and the routines build these side-chains. Once the PDB or CHARMM backbone co-ordinates (and any side-chain co-ordinates that are to be used) are processed, the entire procedure can be performed by running a single command file.

(e) *Computer time*

The initial placement of side-chains from the library takes only a few seconds of central processing unit time on a single processor of an SGI 340. The iterative minimizations to refine the structure can take from 6 h (crambin) to 24 h (thermolysin) on a single processor of an SGI 340, depending on the size of the protein.

3. Results

We first describe the backbone-dependent rotamer library and then present the results of applying it with the refinement methodology to the prediction of the side-chain conformations to a set of six proteins of known structure.

(a) *The backbone-dependent rotamer library*

In Table 4, the total number of each side-chain appears, and the actual and relative populations of the various rotamers are listed according to side-chain type. These results form a backbone-independent rotamer library that can be compared to that of Ponder & Richards (1987). They are essentially the same, except for the statistics for methionine, as already mentioned. In Table 5, which is constructed from the backbone-dependent rotamer library, we list the rotamer populations for values of ϕ and ψ for which there are more than ten examples of a particular side-chain type. One should note the large variation in populations of particular rotamers as a function of ϕ and ψ, and the identity of the side-chain. The variation is not limited to the differences between α-helices or β-sheets, but other regions of the ϕ,ψ map exhibit particular preferences as well. As an example, many side-chains prefer $\chi_1 = 180°$ in canonical α-helices ($\phi = -47°$, $\psi = -57°$), but in nearby regions of the Ramachandran map (more negative values of ϕ, and more positive values of ψ), $\chi_1 = -60°$ is much more common. This is true for the aromatic residues, Leu, the longer side-chains (Arg, Glu, Gln, Lys and Met), Cys and Val. The variation in the most probable value can also be compared with the average value in Table 4.

While many amino acids in specific ϕ,ψ ranges prefer one rotamer over all others, in some cases two or more rotamers have nearly equal populations. In the latter case, removing one protein (and hence 1 or more side-chains from the data set) may switch the balance between the two. This happens for ribonuclease, where adding 7rsa to the database changes the predictions of six side-chains for the better. This can happen even when there are many side-chains in a given ϕ,ψ block. For example, both Met29 and Met30 in 7rsa are in the same block. Without them, their χ_2 percentages are g^-, t, g^+

Table 4

Backbone-independent rotamer library

Res.	Number in Database		No. χ_1	%χ_1	Rotamer (Table 2)
Cys	434	$\chi_1 = 60 \pm 60$	54	12·4	1
		$\chi_1 = 180 \pm 60$	109	25·1	2
		$\chi_1 = -60 \pm 60$	267	61·5	3
Ser	1717	$\chi_1 = 60 \pm 60$	739	43·0	1
		$\chi_1 = 180 \pm 60$	424	24·7	2
		$\chi_1 = -60 \pm 60$	540	31·5	3
Thr	1460	$\chi_1 = 60 \pm 60$	673	46·1	1
		$\chi_1 = 180 \pm 60$	126	8·6	2
		$\chi_1 = -60 \pm 60$	657	45·0	3
Val	1683	$\chi_1 = 60 \pm 60$	142	8·4	1
		$\chi_1 = 180 \pm 60$	1176	69·9	2
		$\chi_1 = -60 \pm 60$	362	21·5	3

Res.	Number in Database		No. χ_1	%χ_1	No. χ_2	%χ_2	No. χ_2	%χ_2	No. χ_2	%χ_2	Rotamer (Table 2)
					$\chi_2 = 30 \pm 30$				$\chi_2 = -30 \pm 30$		
Pro	988	$\chi_1 = 30 \pm 30$	547	55·4	39	3·9	0	0·0	508	51·4	1
		$\chi_1 = 30 \pm 30$	441	44·6	433	43·8	0	0·0	8	0·4	3
					$\chi_2 = 30 \pm 30$		$\chi_2 = 90 \pm 30$		$\chi_2 = 150 \pm 30$		
Phe	889	$\chi_1 = 60 \pm 60$	120	13·5	1	0·1	118	13·3	1	0·1	1
		$\chi_1 = 180 \pm 60$	319	35·8	42	4·7	273	30·7	4	0·4	2
		$\chi_1 = -60 \pm 60$	450	50·6	35	3·9	342	38·4	73	8·2	3
					$\chi_2 = 30 \pm 30$		$\chi_2 = 90 \pm 30$		$\chi_2 = 150 \pm 30$		
His	488	$\chi_1 = 60 \pm 60$	54	11·0	4	0·8	47	9·6	3	0·6	1
		$\chi_1 = 180 \pm 60$	164	33·5	35	7·2	108	22·1	21	4·3	2
		$\chi_1 = -60 \pm 60$	270	55·2	23	4·7	181	37·0	6	13·5	3
					$\chi_2 = 30 \pm 30$		$\chi_2 = 90 \pm 30$		$\chi_2 = 150 \pm 30$		
Tyr	856	$\chi_1 = 60 \pm 60$	102	11·9	5	0·6	97	11·3	0	0·0	1
		$\chi_1 = 180 \pm 60$	293	34·1	49	5·7	241	28·1	3	0·3	2
		$\chi_1 = -60 \pm 60$	461	53·7	19	2·2	359	41·8	83	9·7	3
					$\chi_2 = 90 \pm 90$				$\chi_2 = -90 \pm 90$		
Trp	325	$\chi_1 = 60 \pm 60$	51	15·6	17	5·2			34	10·4	1, 3
		$\chi_1 = 180 \pm 60$	105	32·2	62	19·0			40	12·3	4, 6
		$\chi_1 = -60 \pm 60$	169	51·8	126	38·7			41	12·6	7, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Leu	1739	$\chi_1 = 60 \pm 60$	35	2·0	20	1·1	13	0·7	2	0·1	1, 2, 3
		$\chi_1 = 180 \pm 60$	573	32·9	475	27·3	80	4·6	14	0·8	4, 5, 6
		$\chi_1 = -60 \pm 60$	1131	64·9	135	7·7	949	54·5	44	2·5	7, **8**, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Ile	1176	$\chi_1 = 60 \pm 60$	176	15·0	17	1·4	154	13·1	5	0·4	1, 2, 3
		$\chi_1 = 180 \pm 60$	133	11·3	35	3·0	92	7·8	5	0·4	4, 5, 6
		$\chi_1 = -60 \pm 60$	867	73·7	35	3·0	668	56·8	162	13·9	7, **8**, 9
					$\chi_2 = -60 \pm 60$		$\chi_2 = 0 \pm 30$		$\chi_2 = 60 \pm 30$		
Asp	1342	$\chi_1 = 60 \pm 60$	244	18·1	50	3·7	156	11·6	38	2·8	1, 2, 3
		$\chi_1 = 180 \pm 60$	405	30·0	61	4·5	238	17·7	104	7·7	4, 5, 6
		$\chi_1 = -60 \pm 60$	693	51·4	240	17·8	404	30·0	47	3·5	7, **8**, 9
					$\chi_2 = -60 \pm 60$		$\chi_2 = 0 \pm 30$		$\chi_2 = 60 \pm 30$		
Asn	1048	$\chi_1 = 60 \pm 60$	180	17·1	37	3·5	93	8·8	50	4·7	1, 2, 3
		$\chi_1 = 180 \pm 60$	300	28·5	72	6·8	103	9·8	124	11·8	4, 5, 6
		$\chi_1 = -60 \pm 60$	568	53·9	312	29·6	197	18·7	56	5·3	7, **8**, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Met	399	$\chi_1 = 60 \pm 60$	34	8·5	4	1·0	27	6·7	3	0·7	1, 2, 3
		$\chi_1 = 180 \pm 60$	130	32·3	38	9·5	83	20·6	9	2·2	4, 5, 6
		$\chi_1 = -60 \pm 60$	235	58·5	6	1·5	138	34·3	90	22·4	7, 8, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Glu	1169	$\chi_1 = 60 \pm 60$	129	10·8	7	0·6	83	7·0	37	3·1	1, 2, 3
		$\chi_1 = 180 \pm 60$	383	32·2	78	6·6	288	24·2	17	1·4	4, 5, 6
		$\chi_1 = -60 \pm 60$	657	55·2	98	8·2	389	32·7	162	13·6	7, **8**, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Gln	808	$\chi_1 = 60 \pm 60$	68	8·3	6	0·7	52	6·3	9	1·1	1, 2, 3
		$\chi_1 = 180 \pm 60$	270	33·0	80	9·8	168	20·5	17	2·1	4, 5, 6
		$\chi_1 = -60 \pm 60$	470	57·4	29	3·5	315	38·5	126	15·4	7, **8**, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Arg	807	$\chi_1 = 60 \pm 60$	76	9·3	9	1·1	64	7·8	2	0·2	1, 2, 3
		$\chi_1 = 180 \pm 60$	253	30·9	43	5·2	199	24·3	9	1·1	4, 5, 6
		$\chi_1 = -60 \pm 60$	478	58·3	21	2·6	368	44·9	88	10·7	7, **8**, 9

Table 4 *(continued)*

Res.	Number in Database		No. χ_1	%χ_1	No. χ_2	%χ_2	No. χ_2	%χ_2	No. χ_2	%χ_2	Rotamer (Table 2)
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Lys	1402	$\chi_1 = 60 \pm 60$	121	8·5	10	0·7	102	7·1	8	0·6	1, 2, 3
		$\chi_1 = 180 \pm 60$	477	33·4	90	6·3	360	25·2	27	1·9	4, 5, 6
		$\chi_1 = -60 \pm 60$	804	56·3	46	3·2	572	40·1	176	12·3	7, **8**, 9

Rotamer populations summed over the entire database.

For each amino acid, the total number of residues in the database is given (Number in Database), as well as a breakdown according to the χ_1 and χ_2 limits shown (all angles in degrees). χ_1 populations are broken down under the columns labeled No. χ_1 and %χ_1 for the total number and percentage of the side-chains of the given type in the database with χ_1 in the range denoted in the third column of each row. The χ_1 total add up to 100%. The remaining figures in the Table give the total number and percentages of particular χ_1/χ_2 combinations, for values of χ_1 and χ_2 denoted in the given row and column for each amino acid type. These χ_1/χ_2 percentage figures add up to 100%. The numbers in the last column refer to the conformation numbers listed in Table 2 and represented in the ϕ,ψ maps of Fig. 2. The numbers in bold type represent the most probable conformation for each amino acid type.

equal to 0, 40, 38, leading to a prediction of $t(180°)$; with them in the library, the percentages are 0, 39, 41, leading to the correct prediction for both of them (g^+ or $-60°$). This happens even though this ϕ,ψ block has 52 Met side-chains without 7rsa. In spite of such limitations, because the backbone selects different rotamers in different parts of the map, the predictive value of the backbone-dependent rotamer library is significantly higher than that of the average map. This will be discussed later in comparing predictions of the library in Table 4 (backbone-independent rotamer library) and the library in Table 5 (backbone-dependent rotamer library).

Figure 2 shows graphically the distribution of χ_1 and χ_2 values for the side-chains on Ramachandran (ϕ,ψ) plots. The numbers in Figure 2 refer to the numbered rotamer definitions in Table 2 with the most probable rotamer indicated. Residues with only χ_1 are represented by the numbers 1, 2 and 3 corresponding to χ_1 equal to 60°, 180° and −60°, respectively. Most other side-chains are represented by numbers 1 through 9 corresponding to three conformers for $\chi_1 = 60°$ ($\chi_2 = 60$, 180, $-60° \rightarrow$ numbers 1, 2, 3), $\chi_1 = 180°$ ($\chi_2 = 60$, 180, $-60° \rightarrow$ numbers 4, 5, 6), and $\chi_1 = -60°$ ($\chi_2 = 60$, 180, $-60° \rightarrow$ numbers 7, 8, 9). Aromatics have fewer possible conformations, and are listed in Table 2.

Figure 2 makes clear certain features of the relation between backbone and side-chain conformations that are useful for understanding protein structures. The amino acids can be grouped into a number of different kinds that exhibit similar behavior across the Ramachandran maps: (1) side-chains branched at C^β (Val, Ile, Thr); (2) side-chains branched at C^γ except Asp and Asn (aromatics, Leu); (3) Asp and Asn; (4) chains unbranched through C^δ (Arg, Lys, Met, Glu, Gln); (5) Ser and Cys; and (6) Pro.

The first group, side-chains possessing two γ heavy atoms, have steric requirements not found in other side-chains. Because of the definition of χ_1 of Val, conformations 1, 2, 3 of Val are equivalent to conformations of 2, 3, 1, respectively, of Thr and 4-6, 7-9, 1-3 of Ile. In this first group, the preferred conformations are strongly dependent on ψ and only weakly on ϕ. Values of $-30°$ and lower require a χ_1 of $-60°$ for Ile and Thr (equivalent to 180° for Val) to avoid clashes between the γ side-chain atoms and the backbone N of the succeeding residue; values of ψ from $-30°$ to $+40°$ yield mostly $\chi_1 = -60°$ ($+60°$ for Val), and β-sheet regions split at $\psi = 140°$ with $\chi_1 = +60°$ ($-60°$ for Val) below 140° and $\chi_1 = +60°$ ($-60°$ for Val) above 140°.

Side-chains with two δ heavy atoms (aromatics and Leu) are more complex in their behavior. In the α-helix region ($\phi,\psi = -57°, -47°$), these side-chains uniformly have $\chi_1 = 180°$. In nearby regions involving slightly unwound or distorted helices and turn conformations (type I with ϕ,ψ equal to $-60°$, $-30°$, type II′ with ϕ,ψ equal to $-80°$, 0° and type III with ϕ,ψ equal to $-60°$, $-30°$) $\chi_1 = -60°$ is strongly preferred. In the upper half of the Ramachandran map, χ_1 seems to vary more with ϕ than with ψ. At $\phi > -80°$ (e.g. type II turns), $\chi_1 = 180°$ (numbered 4, 5, 6 depending on χ_2) is common. In the middle region where most β-sheet conformations are found ($-140° < \phi < -180°$), $\chi_1 = -60°$ is common, and in the upper far left region ($\phi < -140°, \psi > 140°$) $\chi_1 = +60°$ occurs. Leucine has two predominant conformations, χ_1,χ_2 of $-60°,180°$ (numbered 8 in Fig. 2) and χ_1,χ_2 of $180°,60°$ (numbered 4 in Fig. 2). Near $\phi,\psi = 180°$, conformations with $\chi_1 = 60°$ are found. (Note: the Protein Data Bank uses the opposite orientation of $C^{\delta 1}$ and $C^{\delta 2}$ for leucine than IUPAC or CHARMM; the map uses the PDB definition.)

Residues Asn and Asp tend to have $\chi_1 = -60°$ (numbered 7, 8, 9) in α-helices, rather than $\chi_1 = 180°$. The distribution in the top half of the ϕ,ψ maps is dominated by ψ, with $\chi_1 = 180°$ conformations common below $\psi = 140°$. From $\psi = 140°$ to 160°, $\chi_1 = -60°$ is most common; above 160° (through 220°, or $-160°$, $\chi_1 = +60°$ is found. Since some positions are underpopulated (as shown by the numbers in italics in Figure 2), it is possible that some of the variation is caused by limitations in the data.

The longer side-chains, Met, Arg, Lys, Glu and Gln, all exhibit similar behavior; that is, the χ_1,χ_2 values are 180,180° in α-helices, +60,180° in the far upper left of the Ramachandran maps, some

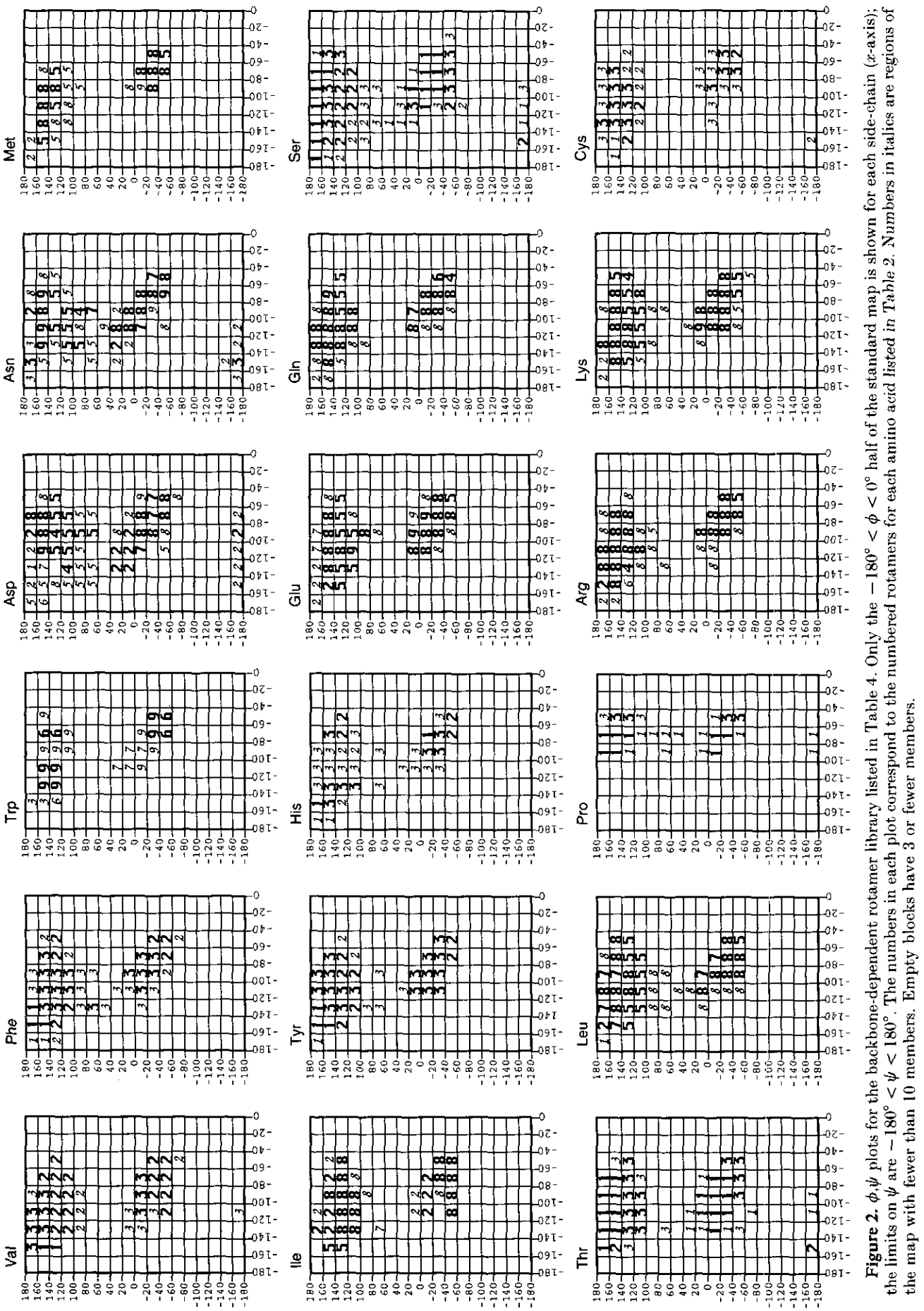


Figure 2. ϕ,ψ plots for the backbone-dependent rotamer library listed in Table 4. Only the $-180° < \phi < 0°$ half of the standard map is shown for each side-chain (x-axis); the limits on ψ are $-180° < \psi < 180°$. The numbers in each plot correspond to the numbered rotamers for each amino acid listed in Table 2. Numbers in italics are regions of the map with fewer than 10 members. Empty blocks have 3 or fewer members.

Table 5
Backbone-dependent rotamer library

1	2	3	4	5	6	7	8	9
C	10	−160	−140	120	140	0	70	30
C	15	−140	−120	120	140	0	27	73
C	16	−140	−120	140	160	0	6	94
C	16	−140	−120	160	180	25	0	75
C	14	−120	−100	100	120	0	86	14
C	14	−120	−100	120	140	0	43	57
C	16	−120	−100	140	160	6	6	88
C	19	−100	−80	−20	0	11	0	89
C	23	−100	−80	120	140	4	26	70
C	15	−100	−80	140	160	27	0	73
C	27	−80	−60	−60	−40	0	37	63
C	35	−80	−60	−40	−20	3	26	71
C	13	−80	−60	140	160	0	0	100
C	19	−60	−40	−60	−40	11	47	37
C	12	−60	−40	−20	−20	33	17	50
S	18	−180	−160	160	180	89	11	0
S	12	−160	−140	−180	−160	33	67	0
S	23	−160	−140	120	140	9	74	17
S	58	−160	−140	140	160	40	48	12
S	52	−160	−140	160	180	90	6	4
S	41	−140	−120	120	140	12	46	37
S	75	−140	−120	140	160	35	27	39
S	33	−140	−120	160	180	67	15	18
S	10	−120	−100	−60	−40	10	80	10
S	17	−120	−100	−20	0	71	6	18
S	21	−120	−100	0	20	43	14	43
S	17	−120	−100	100	120	6	53	41
S	43	−120	−100	120	140	2	51	47
S	29	−120	−100	140	160	31	28	38
S	21	−120	−100	160	180	62	10	29
S	16	−100	−80	−60	−40	25	31	44
S	32	−100	−80	−40	−20	47	16	38
S	67	−100	−80	−20	0	67	3	28
S	36	−100	−80	0	20	64	6	31
S	15	−100	−80	100	120	20	47	33
S	35	−100	−80	120	140	9	54	37
S	39	−100	−80	140	160	38	23	38
S	35	−100	−80	160	180	91	0	6
S	116	−80	−60	−60	−40	16	37	47
S	182	−80	−60	−40	−20	60	6	33
S	114	−80	−60	−20	0	82	2	16
S	10	−80	−60	100	120	0	60	40
S	38	−80	−60	120	140	5	61	34
S	75	−80	−60	140	160	47	27	25
S	30	−80	−60	160	180	87	0	10
S	63	−60	−40	−60	−40	6	38	54
S	78	−60	−40	−40	−20	53	12	36
S	15	−60	−40	−20	0	73	0	27
S	22	−60	−40	120	140	9	41	50
S	23	−60	−40	140	160	13	13	74
T	13	−160	−140	−180	−160	23	77	0
T	33	−160	−140	140	160	33	52	15
T	19	−160	−140	160	180	53	47	0
T	10	−140	−120	−180	−160	50	50	0
T	15	−140	−120	−20	0	73	20	7
T	19	−140	−120	100	120	11	0	89
T	67	−140	−120	120	140	7	6	87
T	76	−140	−120	140	160	62	11	28
T	62	−140	−120	160	180	89	8	3
T	15	−120	−100	−40	−20	87	0	13
T	42	−120	−100	−20	0	95	2	2
T	34	−120	−100	0	20	97	0	3
T	32	−120	−100	100	120	6	0	94
T	84	−120	−100	120	140	5	11	85
T	49	−120	−100	140	160	63	6	31
T	29	−120	−100	160	180	93	3	3
T	14	−100	−80	−60	−40	21	0	79
T	30	−100	−80	−40	−20	67	0	33
T	55	−100	−80	−20	0	95	0	5
T	12	−100	−80	0	20	100	0	0
T	15	−100	−80	100	120	0	7	93

Table 5 *(continued)*

T	46	−100	−80	120	140	0	4	93
T	27	−100	−80	140	160	67	11	19
T	32	−100	−80	160	180	94	3	3
T	112	−80	−60	−60	−40	4	1	95
T	114	−80	−60	−40	−20	54	6	39
T	41	−80	−60	−20	0	93	2	5
T	45	−80	−60	120	140	2	7	91
T	31	−80	−60	140	160	74	10	16
T	26	−80	−60	160	180	92	4	4
T	77	−60	−40	−60	−40	6	3	91
T	27	−60	−40	−40	−20	37	7	56
T	21	−60	−40	120	140	5	5	90
V	20	−160	−140	120	140	50	45	5
V	31	−160	−140	140	160	48	10	39
V	12	−160	−140	160	180	17	0	83
V	50	−140	−120	100	120	4	94	2
V	146	−140	−120	120	140	8	86	5
V	99	−140	−120	140	160	12	35	53
V	50	−140	−120	160	180	0	4	96
V	11	−120	−100	−60	−40	0	82	18
V	20	−120	−100	−20	0	5	15	80
V	71	−120	−100	100	120	0	97	3
V	181	−120	−100	120	140	7	88	4
V	49	−120	−100	140	160	14	43	43
V	12	−120	−100	160	180	0	0	100
V	13	−100	−80	−60	−40	8	92	0
V	15	−100	−80	−40	−20	0	53	47
V	13	−100	−80	−20	0	23	15	62
V	43	−100	−80	100	120	7	93	0
V	80	−100	−80	120	140	6	88	6
V	29	−100	−80	140	160	14	41	45
V	207	−80	−60	−60	−40	2	97	0
V	131	−80	−60	−40	−20	11	60	28
V	19	−80	−60	−20	0	32	21	47
V	15	−80	−60	100	120	0	93	7
V	62	−80	−60	120	140	2	94	5
V	27	−80	−60	140	160	0	56	44
V	109	−60	−40	−60	−40	2	92	6
V	39	−60	−40	−40	−20	28	54	18
V	16	−60	−40	120	140	6	75	19
P	18	−100	−80	−20	0	89	0	11
P	25	−100	−80	140	160	92	0	8
P	18	−100	−80	160	180	100	0	0
P	110	−80	−60	−40	−20	55	0	45
P	64	−80	−60	−20	0	86	0	14
P	70	−80	−60	120	140	63	0	37
P	194	−80	−60	140	160	56	0	44
P	66	−80	−60	160	180	82	0	18
P	49	−60	−40	−60	−40	22	0	78
P	124	−60	−40	−40	−20	33	0	67
P	84	−60	−40	120	140	25	0	75
P	67	−60	−40	140	160	22	0	78

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
F	10	−160	−140	120	140	30	66	10	0	30	0	0	60	0	0	10	0
F	36	−160	−140	140	160	69	6	25	0	69	0	0	6	0	0	25	0
F	22	−160	−140	160	180	95	5	0	0	95	0	5	0	0	0	0	0
F	10	−140	−120	60	80	0	10	90	0	0	0	10	0	0	0	80	10
F	12	−140	−120	100	120	0	67	33	0	0	0	25	33	8	0	33	0
F	37	−140	−120	120	140	8	41	49	3	3	3	0	38	3	3	41	5
F	48	−140	−120	140	160	31	4	65	0	31	0	0	4	0	0	65	0
F	16	−140	−120	160	180	56	0	44	0	56	0	0	0	0	0	38	6
F	10	−120	−100	−20	0	10	10	80	0	10	0	10	0	0	0	50	30
F	17	−120	−100	100	120	0	18	82	0	0	0	6	12	0	0	71	12
F	39	−120	−100	120	140	0	18	82	0	0	0	3	15	0	3	74	5
F	31	−120	−100	140	160	13	3	84	0	13	0	0	3	0	10	68	6
F	12	−100	−80	−40	−20	8	0	92	0	8	0	0	0	0	8	58	25
F	22	−100	−80	−20	0	23	9	68	0	23	0	0	9	0	5	45	18
F	12	−100	−80	0	20	0	0	100	0	0	0	0	0	0	0	83	17
F	13	−100	−80	100	120	0	23	77	0	0	0	8	15	0	8	69	0
F	36	−100	−80	120	140	0	47	53	0	0	0	6	42	0	3	44	6
F	28	−100	−80	140	160	0	11	89	0	0	0	0	11	0	7	64	18
F	89	−80	−60	−60	−40	0	79	21	0	0	0	12	66	0	2	13	6

Table 5 *(continued)*

F	65	−80	−60	−40	−20	3	29	68	0	3	0	2	28	0	8	42	18
F	12	−80	−60	−20	0	42	0	58	0	42	0	0	0	0	8	33	17
F	22	−80	−60	120	140	0	73	27	0	0	0	5	68	0	5	9	14
F	20	−80	−60	140	160	0	20	80	0	0	0	0	20	0	10	55	15
F	102	−60	−40	−60	−40	2	79	19	0	2	0	8	70	2	4	6	9
F	15	−60	−40	−40	−20	7	53	40	0	7	0	0	53	0	20	13	7
F	10	−60	−40	120	140	0	90	10	0	0	0	20	70	0	0	10	0
H	10	−160	−140	140	160	10	40	50	0	10	0	0	40	0	0	50	0
H	13	−160	−140	160	180	69	0	31	8	62	0	0	0	0	0	31	0
H	12	−140	−120	100	120	0	33	67	0	0	0	8	8	17	8	50	8
H	12	−140	−120	120	140	0	50	50	0	0	0	8	42	0	0	50	0
H	21	−140	−120	140	160	5	14	81	0	5	0	5	10	0	0	76	5
H	14	−100	−80	−40	−20	0	14	86	0	0	0	7	7	0	7	43	36
H	15	−100	−80	−20	0	20	0	80	7	13	0	0	0	0	0	53	27
H	54	−80	−60	−60	−40	2	52	44	0	2	0	11	35	6	11	20	13
H	39	−80	−60	−40	−20	8	26	67	0	8	0	3	21	3	8	46	13
H	14	−80	−60	−20	0	79	7	14	0	79	0	0	7	0	0	0	14
H	22	−80	−60	120	140	0	73	27	0	0	0	14	50	9	0	27	0
H	13	−80	−60	140	160	15	38	46	0	15	0	8	31	0	0	15	31
H	32	−60	−40	−60	−40	3	81	16	0	3	0	22	53	6	3	6	6
H	10	−60	−40	120	140	0	90	10	0	0	0	0	80	10	0	10	0
W	18	−140	−120	120	140	0	11	89	0	0	0	11	0	0	78	0	11
W	16	−140	−120	140	160	31	0	69	0	0	31	0	0	0	56	0	13
W	14	−120	−100	120	140	0	36	64	0	0	0	14	0	21	57	0	7
W	10	−120	−100	140	160	20	10	70	10	0	10	10	0	0	30	0	40
W	37	−80	−60	−60	−40	0	78	22	0	0	0	51	0	27	19	0	3
W	27	−80	−60	−40	−20	15	33	48	4	0	11	11	0	22	33	0	15
W	12	−80	−60	120	140	0	67	33	0	0	0	25	0	42	33	0	0
W	12	−80	−60	140	160	8	50	42	0	0	8	50	0	0	33	0	8
W	29	−60	−40	−60	−40	3	69	28	0	0	3	45	0	21	17	0	10
W	14	−60	−40	−40	−20	36	21	43	0	0	36	14	0	7	36	0	7
Y	13	−160	−140	120	140	8	77	15	0	8	0	23	54	0	0	15	0
Y	24	−160	−140	140	160	58	8	33	4	54	0	0	8	0	0	33	0
Y	27	−160	−140	160	180	93	4	4	0	93	0	0	4	0	0	4	0
Y	13	−140	−120	100	120	0	54	46	0	0	0	0	54	0	0	38	8
Y	47	−140	−120	120	140	4	28	68	0	4	0	2	26	0	4	62	2
Y	56	−140	−120	140	160	11	7	82	0	11	0	2	5	0	0	80	2
Y	22	−140	−120	160	180	59	0	41	0	59	0	0	0	0	5	36	0
Y	10	−120	−100	−40	−20	0	0	100	0	0	0	0	0	0	0	100	0
Y	14	−120	−100	−20	0	7	7	86	0	7	0	0	7	0	0	64	21
Y	23	−120	−100	0	20	0	0	100	0	0	0	0	0	0	0	74	26
Y	13	−120	−100	100	120	0	46	54	0	0	0	8	38	0	0	54	0
Y	34	−120	−100	120	140	0	21	76	0	0	0	3	18	0	0	71	6
Y	41	−120	−100	140	160	2	5	93	0	2	0	2	2	0	2	88	2
Y	11	−120	−100	160	180	9	0	91	0	9	0	0	0	0	0	82	9
Y	14	−100	−80	−40	−20	7	21	71	0	7	0	7	14	0	0	50	21
Y	24	−100	−80	−20	0	17	0	79	13	4	0	0	0	0	4	58	17
Y	10	−100	−80	0	20	10	0	90	0	10	0	0	0	0	0	70	20
Y	15	−100	−80	100	120	0	87	13	0	0	0	13	67	7	0	13	0
Y	28	−100	−80	120	140	0	64	36	0	0	0	21	43	0	0	29	7
Y	19	−100	−80	140	160	0	11	89	0	0	0	0	11	0	0	53	37
Y	11	−100	−80	160	180	18	9	73	0	18	0	0	9	0	0	64	9
Y	63	−80	−60	−60	−40	0	84	16	0	0	0	10	75	0	2	6	8
Y	52	−80	−60	−40	−20	10	37	54	0	10	0	6	31	0	12	19	23
Y	12	−80	−60	−20	0	25	8	67	0	25	0	0	8	0	8	42	17
Y	40	−80	−60	120	140	0	73	28	0	0	0	10	63	0	3	13	13
Y	21	−80	−60	140	160	14	19	67	0	14	0	0	19	0	5	48	14
Y	68	−60	−40	−60	−40	1	84	13	0	1	0	10	74	0	1	6	6
Y	14	−60	−40	−40	−20	36	29	36	0	36	0	14	14	0	0	7	29
L	11	−160	−140	120	140	9	91	0	0	9	0	64	27	0	0	0	0
L	15	−160	−140	140	160	27	33	40	20	7	0	27	7	0	7	33	0
L	14	−160	−140	160	180	64	21	14	29	29	7	14	0	7	0	14	0
L	23	−140	−120	100	120	0	83	17	0	0	0	52	26	0	0	17	0
L	36	−140	−120	120	140	0	56	44	0	0	0	42	14	0	8	36	0
L	52	−140	−120	140	160	6	4	90	0	6	0	2	2	0	19	63	8
L	10	−140	−120	160	180	20	0	80	20	0	0	0	0	0	10	70	0
L	15	−120	−100	0	20	0	0	100	0	0	0	0	0	0	7	93	0
L	59	−120	−100	100	120	0	54	46	0	0	0	41	10	2	12	34	0
L	111	−120	−100	120	140	0	50	50	0	0	0	41	8	1	9	38	3
L	56	−120	−100	140	160	0	5	95	0	0	0	4	2	0	14	79	2
L	12	−120	−100	160	180	0	0	100	0	0	0	0	0	0	25	75	0

Table 5 *(continued)*

L	16	−100	−80	−60	−40	0	31	69	0	0	0	25	0	6	6	63	0
L	24	−100	−80	−40	−20	0	17	83	0	0	0	17	0	0	8	71	4
L	39	−100	−80	−20	0	0	0	100	0	0	0	0	0	0	8	87	5
L	15	−100	−80	0	20	7	0	93	7	0	0	0	0	0	7	73	7
L	36	−100	−80	100	120	0	53	47	0	0	0	44	6	3	6	42	0
L	76	−100	−80	120	140	0	50	50	0	0	0	46	4	0	3	46	1
L	60	−100	−80	140	160	2	3	95	2	0	0	2	2	0	8	83	2
L	15	−100	−80	160	180	7	0	93	7	0	0	0	0	0	7	80	7
L	264	−80	−60	−60	−40	0	44	55	0	0	0	40	3	1	5	48	2
L	246	−80	−60	−40	−20	0	21	79	0	0	0	17	2	1	7	67	4
L	34	−80	−60	−20	0	3	9	88	3	0	0	9	0	0	15	71	3
L	15	−80	−60	100	120	0	80	20	0	0	0	80	0	0	0	20	0
L	70	−80	−60	120	140	0	59	41	0	0	0	44	11	3	9	31	1
L	62	−80	−60	140	160	0	13	87	0	0	0	11	2	0	10	74	3
L	135	−60	−40	−60	−40	0	56	44	0	0	0	46	9	1	4	39	1
L	51	−60	−40	−40	−20	0	25	75	0	0	0	24	2	0	8	61	6
L	16	−60	−40	120	140	0	69	31	0	0	0	56	13	0	0	31	0
L	10	−60	−40	140	160	10	0	90	0	10	0	0	0	0	20	70	0
I	14	−160	−140	120	140	0	57	43	0	0	0	7	50	0	0	36	7
I	24	−160	−140	140	160	21	75	4	0	21	0	8	63	4	0	4	0
I	31	−140	−120	100	120	0	3	97	0	0	0	0	3	0	6	74	16
I	104	−140	−120	120	140	3	13	83	0	3	0	1	12	1	5	67	11
I	47	−140	−120	140	160	45	26	30	0	45	0	4	21	0	0	30	0
I	28	−140	−120	160	180	93	7	0	14	79	0	0	7	0	0	0	0
I	10	−120	−100	−60	−40	0	10	90	0	0	0	0	10	0	0	80	10
I	13	−120	−100	−20	0	92	8	0	0	92	0	8	0	0	0	0	0
I	57	−120	−100	100	120	0	0	100	0	0	0	0	0	0	5	74	21
I	123	−120	−100	120	140	4	4	92	1	3	0	0	4	0	6	66	20
I	39	−120	−100	140	160	49	13	38	5	44	0	8	5	0	0	23	13
I	13	−100	−80	−60	−40	0	8	92	0	0	0	0	0	0	8	54	31
I	15	−100	−80	−20	0	73	13	13	0	73	0	7	0	7	7	7	0
I	43	−100	−80	100	120	0	2	98	0	0	0	0	2	0	2	67	28
I	68	−100	−80	120	140	6	3	91	3	3	0	0	3	0	0	66	25
I	11	−100	−80	140	160	27	9	64	0	27	0	0	9	0	0	45	18
I	165	−80	−60	−60	−40	0	4	96	0	0	0	2	1	1	1	84	12
I	85	−80	−60	−40	−20	13	21	66	2	11	0	13	7	1	2	47	16
I	14	−80	−60	−20	0	50	36	14	7	43	0	7	29	0	0	14	0
I	34	−80	−60	120	140	0	6	94	0	0	0	0	6	0	15	59	21
I	12	−80	−60	140	160	50	25	25	0	42	8	8	17	0	8	8	8
I	100	−60	−40	−60	−40	0	4	96	0	0	0	3	1	0	2	82	11
I	22	−60	−40	−40	−20	5	36	59	0	5	0	9	27	0	0	41	18
I	10	−60	−40	120	140	0	20	80	0	0	0	0	20	0	10	50	20
D	13	−140	−120	0	20	85	0	15	8	77	0	0	0	0	15	0	0
D	11	−140	−120	20	40	82	18	0	9	64	9	0	9	9	0	0	0
D	18	−140	−120	100	120	6	89	6	6	0	0	17	67	6	6	0	0
D	10	−120	−100	−20	0	50	0	50	30	20	0	0	0	0	40	10	0
D	26	−120	−100	0	20	58	12	31	0	58	0	0	4	8	19	12	0
D	11	−120	−100	20	40	82	0	18	18	64	0	0	0	0	9	0	9
D	43	−120	−100	100	120	0	81	19	0	0	0	12	65	5	5	9	5
D	15	−120	−100	120	140	0	53	47	0	0	0	27	13	13	7	33	7
D	15	−120	−100	140	160	7	0	93	0	0	7	0	0	0	60	20	13
D	13	−100	−80	−180	−160	92	8	0	0	85	8	8	0	0	0	0	0
D	27	−100	−80	−40	−20	0	7	93	0	0	0	0	4	4	37	52	4
D	46	−100	−80	−20	0	30	2	65	11	17	2	0	0	2	35	30	0
D	53	−100	−80	0	20	64	6	30	9	49	6	0	4	2	17	13	0
D	15	−100	−80	60	80	0	87	13	0	0	0	0	80	7	7	7	0
D	25	−100	−80	80	100	0	80	20	0	0	0	8	68	4	4	16	0
D	47	−100	−80	100	120	0	89	11	0	0	0	15	68	4	0	9	2
D	20	−100	−80	120	140	5	60	30	5	0	0	10	35	15	20	10	0
D	30	−100	−80	140	160	0	7	93	0	0	0	0	7	0	53	37	0
D	18	−100	−80	160	180	78	0	22	6	50	22	0	0	0	11	11	0
D	111	−80	−60	−60	−40	7	12	80	4	4	0	0	0	12	14	60	5
D	168	−80	−60	−40	−20	7	11	82	5	2	1	2	1	8	26	52	4
D	46	−80	−60	−20	0	26	11	63	4	17	4	2	0	7	17	41	4
D	21	−80	−60	100	120	0	86	14	0	0	0	10	76	0	0	10	5
D	42	−80	−60	120	140	0	69	31	0	0	0	21	31	17	12	17	2
D	33	−80	−60	140	160	6	21	70	3	3	0	3	15	3	27	36	6
D	12	−80	−60	160	180	50	0	50	17	17	17	0	0	0	0	50	0
D	98	−60	−40	−60	−40	7	14	78	2	5	0	0	4	10	15	58	4
D	43	−60	−40	−40	−20	7	12	79	7	0	0	0	0	12	5	70	5
D	11	−60	−40	120	140	0	64	36	0	0	0	9	36	18	0	27	9
D	13	40	60	20	40	0	15	85	0	0	0	8	0	8	54	31	0

Table 5 *(continued)*

D	20	40	60	40	60	5	20	75	0	0	5	0	5	15	35	35	5
D	10	60	80	20	40	0	10	90	0	0	0	0	10	0	40	50	0
N	10	−160	−140	−180	−160	100	0	0	20	30	50	0	0	0	0	0	0
N	18	−160	−140	160	180	72	17	11	11	22	39	11	6	0	6	6	0
N	14	−140	−120	20	40	57	14	29	7	36	14	7	0	7	21	7	0
N	13	−140	−120	80	100	0	92	8	0	0	0	8	46	38	8	0	0
N	12	−140	−120	100	120	0	92	8	0	0	0	25	42	25	0	0	8
N	16	−140	−120	140	160	19	0	81	0	6	13	0	0	0	56	13	6
N	10	−120	−100	−20	0	30	0	70	20	10	0	0	0	0	60	10	0
N	35	−120	−100	0	20	43	6	51	17	23	3	3	3	0	43	3	3
N	14	−120	−100	20	40	43	0	57	0	36	7	0	0	0	36	21	0
N	27	−120	−100	100	120	0	74	26	0	0	0	37	30	7	22	4	0
N	13	−120	−100	120	140	0	54	46	0	0	0	54	0	0	38	0	8
N	23	−120	−100	140	160	4	9	87	0	0	4	0	0	9	52	30	4
N	30	−100	−80	−20	0	13	7	80	7	7	0	0	3	3	50	27	3
N	50	−100	−80	0	20	40	8	50	6	34	0	0	2	6	22	22	6
N	16	−100	−80	60	80	19	38	44	13	6	0	0	31	6	31	13	0
N	14	−100	−80	80	100	7	71	21	7	0	0	7	50	14	7	7	7
N	30	−100	−80	100	120	0	73	27	0	0	0	20	53	0	13	10	3
N	20	−100	−80	120	140	10	35	55	5	0	5	25	5	5	30	15	10
N	16	−100	−80	140	160	13	6	81	6	6	0	0	0	6	56	19	6
N	10	−100	−80	160	180	60	30	10	10	30	20	0	20	10	0	10	0
N	73	−80	−60	−60	−40	1	23	74	0	0	1	4	0	19	32	38	4
N	96	−80	−60	−40	−20	3	13	84	1	2	0	1	0	11	34	44	6
N	39	−80	−60	−20	0	26	15	59	8	10	8	0	3	13	28	28	3
N	25	−80	−60	120	140	0	56	44	0	0	0	12	20	24	20	12	12
N	14	−80	−60	140	160	7	36	57	0	0	7	7	0	29	43	14	0
N	45	−60	−40	−60	−40	0	20	78	0	0	0	4	0	16	13	58	7
N	13	−60	−40	−40	−20	8	23	69	8	0	0	8	0	15	31	31	8
N	22	40	60	20	40	0	23	77	0	0	0	0	9	14	68	5	5
N	47	40	60	40	60	0	36	64	0	0	0	2	13	21	36	21	4
N	18	60	80	0	20	0	22	78	0	0	0	0	6	11	50	11	17
N	38	60	80	20	40	0	24	74	0	0	0	0	13	11	50	8	16
M	16	−160	−140	140	160	38	50	13	6	31	0	19	25	6	0	13	0
M	26	−140	−120	140	160	27	19	54	4	23	0	0	15	4	0	50	4
M	14	−120	−100	120	140	0	64	36	0	0	0	14	43	7	7	7	21
M	14	−120	−100	140	160	0	14	86	0	0	0	0	14	0	0	50	36
M	11	−100	−80	−40	−20	0	9	91	0	0	0	0	9	0	45	18	27
M	15	−100	−80	120	140	0	47	53	0	0	0	13	33	0	0	27	27
M	10	−100	−80	140	160	0	0	100	0	0	0	0	0	0	0	60	40
M	51	−80	−60	−60	−40	0	31	65	0	0	0	20	10	2	0	39	25
M	54	−80	−60	−40	−20	4	13	81	0	4	0	4	7	2	0	39	41
M	11	−80	−60	−20	0	27	18	55	9	9	9	9	9	0	0	27	27
M	10	−80	−60	120	140	0	70	30	0	0	0	10	50	10	0	10	20
M	36	−60	−40	−60	−40	0	58	42	0	0	0	14	42	3	0	36	6
M	17	−60	−40	−40	−20	12	35	53	0	12	0	18	18	0	0	18	35
E	15	−160	−140	120	140	0	87	13	0	0	0	7	80	0	7	7	0
E	11	−160	−140	140	160	55	36	9	0	45	0	0	36	0	0	9	0
E	16	−140	−120	100	120	0	75	25	0	0	0	19	56	0	0	25	0
E	16	−140	−120	120	140	0	56	44	0	0	0	25	31	0	0	38	6
E	32	−140	−120	140	160	16	16	69	0	16	0	0	16	0	0	47	19
E	13	−120	−100	−20	0	8	8	85	0	8	0	0	8	0	8	62	15
E	16	−120	−100	0	20	13	6	81	0	6	6	0	6	0	6	63	13
E	17	−120	−100	100	120	6	41	53	0	0	6	6	35	0	0	35	18
E	36	−120	−100	120	140	6	44	50	0	6	0	0	44	0	6	31	14
E	22	−120	−100	140	160	18	0	77	0	18	0	0	0	0	0	64	9
E	16	−100	−80	−40	−20	6	13	81	0	6	0	0	13	0	13	56	13
E	29	−100	−80	−20	0	10	0	86	3	3	0	0	0	0	3	41	41
E	11	−100	−80	0	20	9	0	91	0	0	9	0	0	0	18	45	27
E	10	−100	−80	80	100	20	40	40	0	10	10	20	20	0	10	20	10
E	18	−100	−80	100	120	0	67	33	0	0	0	11	56	0	0	28	6
E	27	−100	−80	120	140	4	59	37	0	4	0	7	48	4	4	22	11
E	30	−100	−80	140	160	17	20	63	3	10	3	10	10	0	3	37	23
E	200	−80	−60	−60	−40	1	44	54	0	1	1	11	31	2	9	34	12
E	229	−80	−60	−40	−20	12	22	65	0	10	2	5	16	1	11	39	14
E	51	−80	−60	−20	0	41	8	49	0	10	31	4	4	0	12	24	14
E	23	−80	−60	120	140	9	57	35	0	4	4	4	48	4	0	17	17
E	21	−80	−60	140	160	10	19	67	0	5	5	0	19	0	29	19	19
E	122	−60	−40	−60	−40	4	50	43	0	3	1	7	40	2	8	26	7
E	54	−60	−40	−40	−20	15	33	46	2	7	6	9	24	0	15	22	9
E	11	−60	−40	120	140	0	55	45	0	0	0	0	36	18	18	27	0

Table 5 *(continued)*

Q	15	−160	−140	140	160	40	20	40	0	33	7	0	20	0	0	27	13
Q	20	−140	−120	120	140	0	45	55	0	0	0	10	35	0	0	50	5
Q	20	−140	−120	140	160	15	0	80	0	15	0	0	0	0	0	60	20
Q	14	−140	−120	160	180	14	7	79	0	14	0	0	7	0	7	29	43
Q	15	−120	−100	0	20	7	7	87	0	7	0	7	0	0	0	67	20
Q	12	−120	−100	100	120	0	50	50	0	0	0	8	42	0	0	42	8
Q	35	−120	−100	120	140	3	46	51	0	3	0	9	37	0	3	34	14
Q	22	−120	−100	140	160	0	14	86	0	0	0	0	14	0	0	59	27
Q	10	−120	−100	160	180	10	10	80	0	10	0	0	0	10	0	40	40
Q	17	−100	−80	−40	−20	12	35	53	6	6	0	6	24	0	0	35	18
Q	22	−100	−80	−20	0	18	0	82	0	9	9	0	0	0	5	41	36
Q	17	−100	−80	0	20	6	0	94	6	0	0	0	0	0	0	59	35
Q	12	−100	−80	100	120	0	42	58	0	0	0	17	25	0	17	8	33
Q	29	−100	−80	120	140	3	55	41	0	3	0	17	34	0	3	34	3
Q	22	−100	−80	140	160	14	27	55	5	9	0	0	18	9	5	41	9
Q	101	−80	−60	−60	−40	0	46	52	0	0	0	13	27	4	4	41	8
Q	109	−80	−60	−40	−20	5	32	62	0	4	1	14	17	1	5	49	9
Q	30	−80	−60	−20	0	20	0	80	0	20	0	0	0	0	17	37	27
Q	26	−80	−60	120	140	12	62	27	0	8	4	35	27	0	0	19	8
Q	20	−80	−60	140	160	10	25	65	0	0	5	10	15	0	10	40	15
Q	60	−60	−40	−60	−40	2	65	33	2	0	0	25	32	8	3	18	12
Q	40	−60	−40	−40	−20	5	48	45	0	0	5	8	38	3	5	35	5
Q	11	−60	−40	120	140	9	82	9	0	9	0	27	55	0	9	0	0
Q	12	40	60	40	60	0	8	92	0	0	0	0	8	0	0	58	33
R	20	−160	−140	140	160	25	15	60	0	25	0	0	15	0	0	35	25
R	17	−160	−140	160	180	53	18	29	12	41	0	6	12	0	0	18	12
R	21	−140	−120	120	140	5	52	33	5	0	0	10	43	0	0	24	10
R	35	−140	−120	140	160	11	23	63	0	11	0	0	23	0	0	40	23
R	11	−140	−120	160	180	36	0	64	0	27	9	0	0	0	0	36	27
R	13	−120	−100	100	120	0	23	77	0	0	0	8	15	0	0	69	8
R	23	−120	−100	120	140	0	48	52	0	0	0	13	35	0	0	52	0
R	24	−120	−100	140	160	4	17	79	0	4	0	0	17	0	4	58	17
R	15	−120	−100	160	180	13	0	87	7	7	0	0	0	0	7	73	7
R	22	−100	−80	−40	−20	5	14	73	0	5	0	9	5	0	0	55	18
R	25	−100	−80	−20	0	8	0	92	0	8	0	0	0	0	8	76	8
R	11	−100	−80	0	20	9	9	82	0	9	0	9	0	0	0	64	18
R	27	−100	−80	120	140	4	48	48	0	4	0	0	48	0	7	37	4
R	22	−100	−80	140	160	9	14	77	5	5	0	0	14	0	0	41	36
R	103	−80	−60	−60	−40	1	49	50	0	1	0	10	38	0	2	41	8
R	107	−80	−60	40	−20	9	26	63	2	7	0	7	17	3	1	54	7
R	23	−80	−60	−20	0	26	4	70	4	17	0	0	4	0	0	57	13
R	27	−80	−60	120	140	0	41	56	0	0	0	7	33	0	4	48	4
R	22	−80	−60	140	160	9	18	73	0	9	0	0	18	0	0	68	5
R	72	−60	−40	−60	−40	0	64	36	0	0	0	8	53	3	3	31	3
R	27	−60	−40	−40	−20	22	37	41	4	19	0	19	19	0	0	30	11
K	17	−160	−140	120	140	0	82	18	0	0	0	35	41	6	0	18	0
K	28	−160	−140	140	160	29	32	36	0	29	0	4	29	0	11	21	4
K	10	−140	−120	100	120	20	60	20	0	10	10	20	40	0	0	20	0
K	39	−140	−120	120	140	0	46	51	0	0	0	8	38	0	3	41	5
K	45	−140	−120	140	160	16	13	71	0	16	0	2	11	0	4	49	18
K	18	−140	−120	160	180	17	0	83	0	17	0	0	0	0	0	67	17
K	10	−120	−100	−40	−20	10	20	70	0	10	0	0	20	0	0	40	30
K	16	−120	−100	−20	0	0	13	88	0	0	0	6	6	0	6	50	31
K	21	−120	−100	0	20	5	0	95	0	0	5	0	0	0	0	71	24
K	16	−120	−100	100	120	6	56	38	0	6	0	13	38	6	0	25	13
K	40	−120	−100	120	140	3	33	65	0	3	0	10	18	5	3	53	10
K	33	−120	−100	140	160	6	15	76	0	6	0	3	12	0	9	48	18
K	21	−120	−100	160	180	10	5	86	5	0	5	0	5	0	5	62	19
K	29	−100	−80	−40	−20	7	24	69	0	7	0	10	14	0	10	38	17
K	48	−100	−80	−20	0	13	6	81	0	6	6	0	4	2	6	52	23
K	19	−100	−80	0	20	16	0	84	5	11	0	0	0	0	11	42	26
K	27	−100	−80	100	120	4	70	26	0	4	0	11	56	4	0	15	11
K	33	−100	−80	120	140	0	52	45	0	0	0	12	39	0	3	33	9
K	38	−100	−80	140	160	3	3	92	0	3	0	0	3	0	5	61	26
K	176	−80	−60	−60	−40	4	52	41	1	3	0	10	38	4	2	30	9
K	188	−80	−60	−40	−20	10	27	60	2	7	1	5	21	1	2	44	13
K	54	−80	−60	−20	0	19	7	74	0	19	0	0	7	0	2	48	22
K	10	−80	−60	100	120	0	40	50	0	0	0	10	30	0	0	50	0
K	48	−80	−60	120	140	8	54	38	0	6	2	13	40	2	0	27	10
K	28	−80	−60	140	160	4	11	86	0	4	0	0	11	0	11	61	11
K	134	−60	−40	−60	−40	2	57	40	0	2	0	7	46	4	3	31	4
K	74	−60	−40	−40	−20	9	34	51	1	8	0	8	23	3	1	41	9
K	23	−60	−40	120	140	4	74	22	4	0	0	4	61	9	0	22	0

Table 5 *(continued)*

K	10	−60	−40	140	160	10	50	40	0	10	0	10	40	0	0	40	0
K	11	40	60	40	60	0	0	91	0	0	0	0	0	0	0	82	9

Column 1, residue (1-letter code); column 2, number of side-chains in ϕ,ψ range in structure database; column 3, lower ϕ limit; column 4, upper ϕ limit; column 5, lower ψ limit; column 6, upper ψ limit (all angles in degrees); columns 7 to 9, χ_1 populations of 60°, 180°, −60° rotamers (total = 100%); columns 10 to 18, χ_1,χ_2 rotamer populations according to definitions of 1 to 9 for applicable residues as defined in Table 2 (total = 100%).

180,180° values near $\phi,\psi = -60$, 120° (except Arg) and near $\phi,\psi = -140°, -120°$, and $\chi_1,\chi_2 = -60,180°$ (number 8) nearly everywhere else.

Serine is similar to Thr in most regions of the map. However, there is a large difference in the $\psi = 80$ to 140° region, where $\chi_1 = 180°$ is common for Ser while Thr prefers −60° to avoid contact with the backbone carbonyl oxygen atom. Serine prefers +60° in much of the map, as does Thr, to make hydrogen bonds to the backbone. Since Cys cannot do this, this conformation is largely absent and only the $\chi_1 = 180$ and −60° are found commonly through most of the map with $\chi_1 = 180°$ for $\phi =$ 100° to 120° and in α-helices, and $\chi_1 = -60°$ in most of the rest of the map. Free cysteinyl residues and disulfide-bonded cysteinyl residues were not distinguished in calculating the library.

Finally proline, as noted by Cung *et al.* (1987), exhibits the C^γ-*exo* conformation for $\phi > -60°$ and the C^γ-*endo* conformation for $\phi < -60°$.

(b) *Prediction of side-chain conformations in proteins from the known backbone co-ordinates*

We applied the targ/lib method (see Table 3) to six proteins in the Brookhaven Protein Database by using the backbone co-ordinates from the X-ray structures and initially building the side-chains from the ϕ,ψ rotamer library. These proteins are rhizopuspepsin (C-terminal domain: PDB code 2apr), lysozyme (1lz1), crambin (1crn), bovine pancreatic trypsin inhibitor (5pti), ribonuclease A (7rsa), and thermolysin (3tln). All of these structures were used in the library except 1lz1 (2·0 Å resolution) and 3tln (1·6 Å resolution), which are represented by a highly homologous structure. The relevant information is listed in Table 1. In each case, a library of the form of Table 5 was recalculated with the protein to be predicted and its homologs removed.

As an example, we give detailed results for the small protein crambin in Table 6, for the structures numbered 0, 1 and N in Figure 1 (i.e. from the library alone, after backbone/side-chain clashes have been resolved, and after all side-chain/side-chain clashes are resolved; $N = 2$ in this case). Table 6 lists the experimental χ angles as well as the χ angles predicted from the backbone-dependent rotamer library. All of the initial χ angles are either 60, 180, −60, 0, 90, or −90° except for those of proline and cysteine. Cysteine co-ordinates are minimized in determining each of the structures (see Methods, section (b)(i)(c)), and proline χ_1 is set equal to 28 or −28°, which along with the backbone co-ordinates determines χ_2 of proline. We note that, as in the library, residues such as Asn and His are deemed correct in χ_2 if χ_2 or $\chi_2+180°$ is correct within 40° of the crystal structure. This takes account of the fact already mentioned, that it is usually not possible to distinguish the two positions in the X-ray structure.

In predicting the side-chain orientations for crambin, the backbone-dependent library does well (see Structure 0 in Table 6). Only Thr1, Arg10, Tyr29 and Asn46 are moved in the first set of side-chain minimizations to take care of clashes with the backbone. In the series of minimizations to remove side-chain/side-chain clashes, only Phe13 and Arg17 are moved. Phe13 remains in a good conformation and Arg17 is moved from an incorrect to a correct conformation. Of the four residues that are in incorrect conformations in the final structure, three were never minimized (Leu18, Ile25 and Asp43), and one (Tyr29) was minimized in the first round, but remained in a conformation with the incorrect χ_1 (near −60°, instead of near 180°). Minimizing all of the side-chains at once (data not shown) was found not to improve the final pedictions for crambin or for the other proteins tested here. However, minimizing both the X-ray structure and the final predicted structure with the minimization protocol of Summers & Karplus (1989) (a series of Powell minimizations with decreasing harmonic force constraints on side-chain atom positions) produces generally lower average r.m.s.d. for all side-chain types. In many cases, the predicted and experimental angles are identical. This demonstrates that the predicted and the X-ray positions are in the same local minimum of the force field (see Table 9, 4th column).

The deviations from the X-ray structures for the residues in all six proteins for χ_1 and χ_2 are presented in the stacked histograms of Figure 3. The numbers and fractions of residues correct to within 40° are listed in Table 7 for structures 0, 1 and N ($N = 2$, 3 or 4 for all cases tested here). Also listed are the results predicted directly (without refinement) from the backbone-independent library of Table 4 for comparison with Structure 0 for each protein, predicted directly from the backbone-dependent library. The results are broken down into χ_1, χ_2 and χ_{1+2} predictions. Since the results vary significantly from protein to protein, it is clear that a prediction method cannot be assessed on the basis of tests on one or two proteins (e.g. Desmet *et al.*, 1992). Lysozyme gives the poorest result, probably because it has a high content of charged residues. As already mentioned, they are difficult to predict,

Table 6

Side-chain results for crambin from backbone co-ordinates only

Res. no.	χ	Type	Exp.	Structure 0 Pred (Dif) Cor?	Structure 1 Pred (Dif) Cor?	Structure $N(N=2)$ Pred (Dif) Cor?
1	1	Thr	−59	60 (119)n	−47 (13)y	−47 (13)y
2	1	Thr	56	60 (4)y	60 (4)y	60 (4)y
3	1	Cys	−51	−53 (−1)y	−51 (0)y	−51 (0)y
3	2	Cys	−75	−74 (1)y	−75 (0)y	−75 (0)y
4	1	Cys	−65	−48 (16)y	−66 (−2)y	−69 (−4)y
4	2	Cys	−83	−101 (−19)y	−82 (1)y	−79 (4)y
5	1	Pro	32	28 (−5)y	28 (−5)y	28 (−5)y
5	2	Pro	−41	−39 (3)y	−39 (3)y	−39 (3)y
6	1	Ser	69	60 (−9)y	60 (−9)y	60 (−9)y
7	1	Ile	−74	−60 (14)y	−60 (14y)	−60 (14)y
7	2	Ile	173	−180 (7)y	180 (7)y	180 (7)y
8	1	Val	160	−180 (20)y	180 (20)y	180 (20)y
10	1	Arg	177	−60 (123)n	180 (3)y	180 (3)y
10	2	Arg	64	−180 (116)n	65 (1)y	65 (1)y
10	3	Arg	67	−180 (113)n	72 (5)y	72 (5)y
10	4	Arg	177	−180 (3)y	175 (−2)y	175 (−2)y
11	1	Ser	−66	−60 (6)y	−60 (6)y	−60 (6)y
12	1	Asn	−70	−60 (10)y	−60 (10)y	−60 (10)y
12	2	Asn	−23	0 (23)y	0 (23)y	0 (23)y
13	1	Phe	−175	−180 (−5)y	−180 (−5)y	−167 (7)y
13	2	Phe	−90	90 (0)y	90 (0)y	70 (−20)y
14	1	Asn	−73	−60 (13)y	−60 (13)y	−60 (13)y
14	2	Asn	−24	0 (24)y	0 (24)y	0 (24)y
15	1	Val	171	−180 (9)y	180 (9)y	180 (9)y
16	1	Cys	180	178 (−1)y	178 (−1)y	178 (−1)y
16	2	Cys	−93	−89 (4)y	−90 (4)y	−90 (3)y
17	1	Arg	−67	−60 (7)y	−60 (7)y	−65 (2)y
17	2	Arg	−80	−180 (−100)n	180 (−100)n	−73 (7)y
17	3	Arg	−72	60 (132)n	60 (132)n	−78 (−6)y
17	4	Arg	157	−180 (23)y	−180 (23)y	128 (−30)y
18	1	Leu	−76	−180 (−104)n	−180 (−104)n	−180 (−104)n
18	2	Leu	−63	−180 (−117)n	−180 (−117)n	−180 (−117)n
19	1	Pro	13	28 (15)y	28 (15)y	28 (15)y
19	2	Pro	−14	−37 (−23)y	−37 (−23)y	−37 (−23)y
21	1	Thr	−45	−60 (−15)y	−60 (−15)y	−60 (−15)y
22	1	Pro	−24	−28 (−3)y	−28 (−3)y	−28 (−3)y
22	2	Pro	33	30 (−2)y	30 (−2)y	30 (−2)y
23	1	Glu	−72	−60 (12)y	−60 (12)y	−60 (12)y
23	2	Glu	−172	−180 (−8)y	−180 (−8)y	−180 (−8)y
23	3	Glu	−22	0 (22)y	0 (22)y	0 (22)y
25	1	Ile	−75	−60 (15)y	−60 (15)y	−60 (15)y
25	2	Ile	−72	−180 (−108)n	−180 (−108)n	−180 (−108)n
26	1	Cys	−65	−64 (1)y	−63 (2)y	−63 (3)y
26	2	Cys	−58	−59 (−1)y	−60 (−2)y	−60 (−2)y
26	3	Cys	−86	−86 (0)y	−85 (1)y	−84 (2)y
28	1	Thr	53	60 (7)y	60 (7)y	60 (7)y
29	1	Tyr	−178	−60 (118)n	−77 (101)n	−77 (101)n
29	2	Tyr	55	90 (35)y	7 (132)n	7 (132)n
30	1	Thr	61	60 (−1)y	60 (−1)y	60 (−1)y
32	1	Cys	−54	−57 (−3)y	−57 (−3)y	−56 (−2)y
32	2	Cys	−118	−106 (12)y	−113 (5)y	−116 (2)y
32	3	Cys	106	101 (−4)y	105 (−1)y	106 (0)y
33	1	Ile	65	60 (−5)y	60 (−5)y	60 (−5)y
33	2	Ile	171	−180 (9)y	−180 (9)y	−180 (9)y
34	1	Ile	−60	−60 (0)y	−60 (0)y	−60 (0)y
34	2	Ile	168	−180 (12)y	−180 (12)y	−180 (12)y
35	1	Ile	65	60 (−5)y	60 (−5)y	60 (−5)y
35	2	Ile	169	−180 (11)y	180 (11)y	180 (11)y
36	1	Pro	4	28 (23)y	28 (23)y	28 (23)y
36	2	Pro	−4	−36 (−32)y	−36 (−32)y	−36 (−32)y
39	1	Thr	−52	−60 (−8)y	−60 (−8)y	−60 (−8)y
40	1	Cys	−63	−65 (−2)y	−65 (−2)y	−65 (−2)y
40	2	Cys	−75	−72 (3)y	−72 (3)y	−73 (3)y
40	3	Cys	−79	−79 (0)y	−78 (1)y	−78 (1)y
41	1	Pro	28	28 (0)y	28 (0)y	28 (0)y
41	2	Pro	−35	−41 (−6)y	−41 (−6)y	−41 (−6)y
43	1	Asp	59	−60 (−119)n	−60 (−119)n	−60 (−119)n
43	2	Asp	−24	−60 (−36)y	−60 (−36)y	−60 (−36)y

Table 6 *(continued)*

Res. no.	χ	Type	Exp.	Structure 0 Pred (Dif) Cor?	Structure 1 Pred (Dif) Cor?	Structure $N(N=2)$ Pred (Dif) Cor?
44	1	Tyr	−74	−60 (14)y	−60 (14)y	−60 (14)y
44	2	Tyr	86	90 (4)y	90 (4)y	90 (4)y
46	1	Asn	−57	−60 (−3)y	−60 (−4)y	−60 (−4)y
46	2	Asn	113	−60 (7)y	−68 (−2)y	−68 (−1)y

χ angle predictions (in degrees) for rounds 0 (library), 1 (after backbone/side-chain conflicts have been resolved), and $N = 2$ (after all side-chain conflicts have been resolved). Differences between prediction (Pred) and the experimental structure (Exp.) are listed under column Dif. If the χ angle is correct to within 40°, then a y is entered in the Cor? column, otherwise, an n is listed.

since the conformations in the crystal structure often depend on the effects of solvent and interactions with other proteins (Gelin & Karplus, 1979).

To compare the present results with those of Lee & Subbiah (1991) we consider χ_1 and χ_2 independent of χ_1, since they did not report χ_{1+2}. For crambin, our results are 92% and 89% (for χ_1 and χ_2, respectively), compared with 70% and 60% for Lee & Subbiah (1991); for bovine pancreatic trypsin inhibitor (BPTI), 85% and 78% compared with 76% and 55%; for ribonuclease, 79% and 71% compared with 58% and 61%; and for lysozyme, 77% and 66% compared with 80% and 68%, which is the only protein where the results of Lee & Subbiah (1991) are better than those reported here. For the other two proteins (thermolysin and penicillopepsin, a homolog of rhizopuspepsin), Lee & Subbiah (1991) report only core residue predictions and these are compared below.

It is of interest also to compare the initial placement resulting from the backbone-dependent rotamer library (Structure 0) with the predictions that are obtained from the backbone-independent library of Table 4. As noted before, this library is essentially that of Ponder & Richards (1987), except for methionine χ_2. Desmet *et al.* (1992) have used the Ponder & Richards (1987) rotamers as the beginning of their prediction scheme, so these results reflect the starting structure in their method for each protein studied here. The refinement technique used by them is different from the present one, although the same general principle (removal of van der Waals clashes) is involved. Results for the side-chain placement with the backbone-independent library are listed in the last column of Table 7. Most of the difference between the backbone-independent and backbone-dependent results concern χ_1. The χ_1 predictions from the backbone-independent and backbone-dependent libraries are 52% and 67% for thermolysin, 65% and 80% for BPTI, 68% and 86% for crambin, 64% and 76% for lysozyme, 54% and 71% for the rhizopuspepsin C-terminal domain, and 56% and 72% for ribonuclease. The χ_2 prediction rates are all within 7% from the two libraries. The χ_{1+2} results reflect those for χ_1, and so differ significantly between the libraries. Since the refinement procedure works better when more of the side-chains are close to their experimental conformation, the better placement from the backbone-dependent library is very useful in obtaining the correct conformation for side-chains that are not correct in the initial model.

In Figures 4 and 5, stereo plots of the various side-chain predictions are compared with the X-ray values for BPTI and crambin, respectively. In (a) of each Figure, the X-ray structure is presented alone; in (b) the X-ray structure is compared with the results from the backbone-independent library; in (c) are shown the X-ray structure and the initial backbone-dependent library prediction; and in (d) the X-ray structure and the final refined prediction are compared; (e) shows the X-ray structure and final predicted structure that have both been minimized according to the method of Summers & Karplus (1989). In accord with the numerical results, the Figures illustrate that the predicted and X-ray side-chain positions for most residues are in the same energy minimum and that there is a general improvement in going from (b) to (e).

In Figure 6, the results for core and surface residues are compared with the results for all residues for the six proteins, where core residues are those defined as having less than 10% exposure, and surface residues have greater than 10% exposure (see Methods, section (c)). For five out of six of the proteins, buried residues are more accurately predicted than exposed residues. This is true of the results obtained directly from the rotamer library and even more so after the side-chain minimizations have been performed. BPTI is the sole exception. Since BPTI is quite small (58 residues), the difference between buried and accessible residue predictions corresponds to only two or three residues. Lee & Subbiah (1991) report results for the core region of thermolysin and penicillopepsin, and they predict χ_1 and χ_2 with fractions of 82% and 76% correct for thermolysin and 81% and 81% for penicillopepsin. For comparison, in the core region of thermolysin we predict 78% and 80% of the residues correctly, and in the core region of the C-terminal domain of rhizopuspepsin, a protein homologous to penicillopepsin, we predict 88% and 83% of the residues correctly.

In Tables 8 and 9, the results for the various types of side-chains are summarized for the six proteins. The library does well for hydrophobic

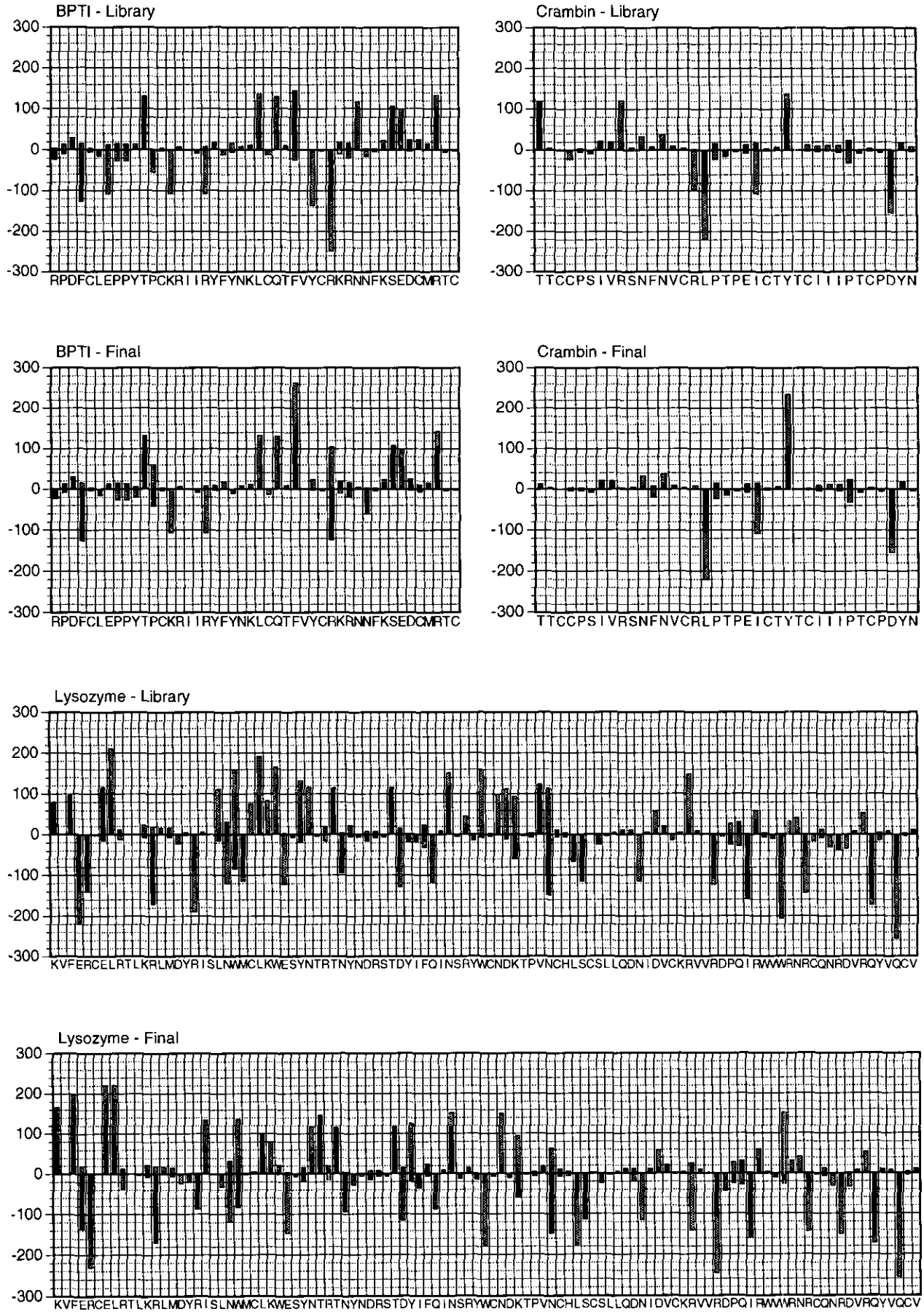

BPTI - Library
300
200
100
0
-100
-200
-300
RPDFCLEPPYTPCKRIIRYFYNKLCQTFVYCRKRNNFKSEDCMRTC
Crambin - Library
TTCCPSIVRSNFNVCRLPTPEICTYTCIIIPTCPDYN
BPTI - Final
Crambin - Final
Lysozyme - Library
KVFERCELRTLKRLMDYRISLNWMCLKWESYNTRTNYNDRSTDYIFQINSRYWCNDKTPVNCHLSCSLLQDNIDVCKRVVRDPQIRWWWRNRCQNRDVRQYVQCV
Lysozyme - Final

Fig. 3.

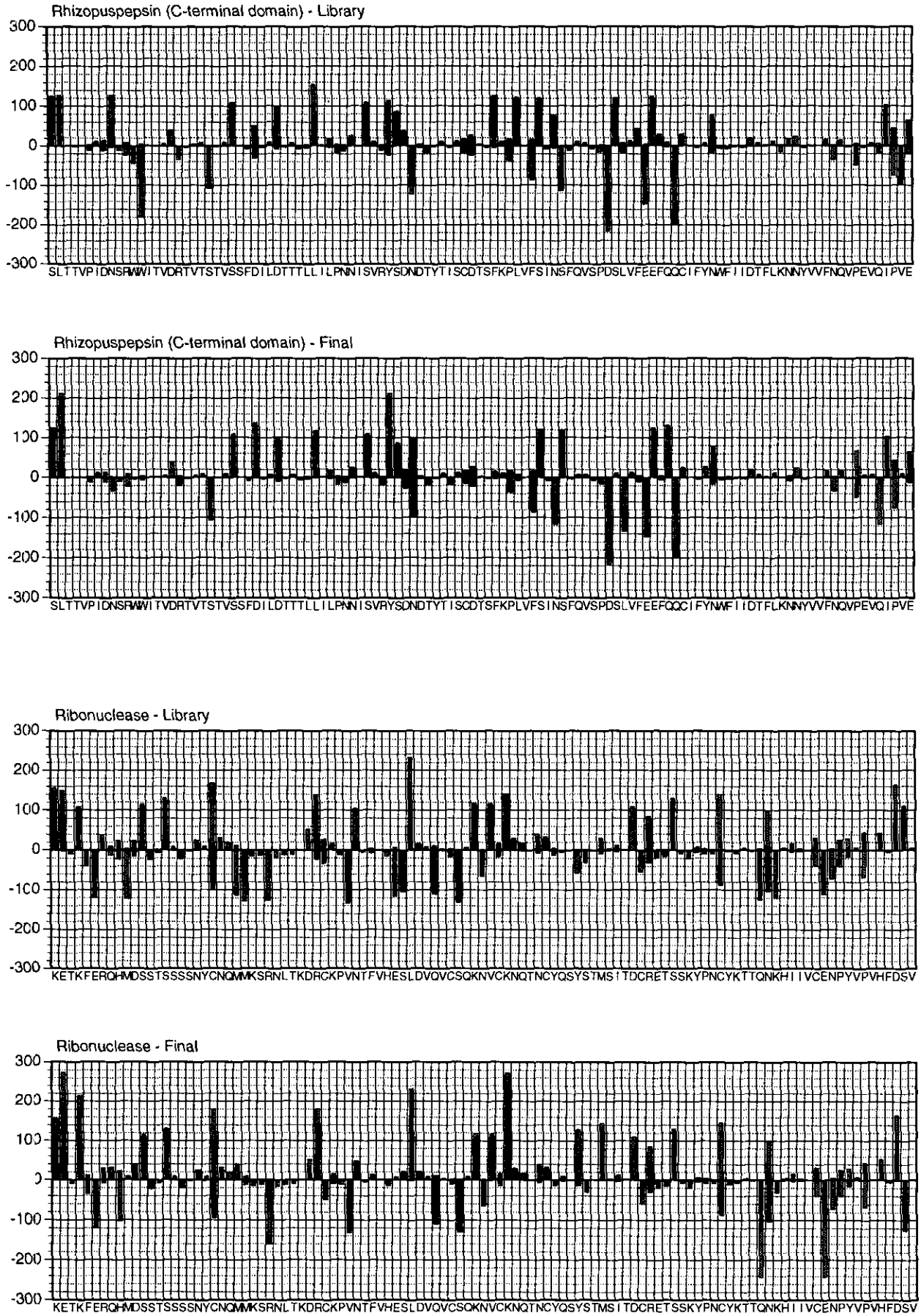

Rhizopuspepsin (C-terminal domain) - Library
300
200
100
0
-100
-200
-300
SLTTVPIDNSRWWITVDRTVTSTVSSFDILDTTTLLILPNNISVRYSDNDTYTISCDTSFKPLVFSINSFQVSPDSLVFEEFQQCIFYNWFIIDTFLKNNYVVFNQVPEVQIPVE
Rhizopuspepsin (C-terminal domain) - Final
300
200
100
0
-100
-200
-300
SLTTVPIDNSRWWITVDRTVTSTVSSFDILDTTTLLILPNNISVRYSDNDTYTISCDTSFKPLVFSINSFQVSPDSLVFEEFQQCIFYNWFIIDTFLKNNYVVFNQVPEVQIPVE
Ribonuclease - Library
300
200
100
0
-100
-200
-300
KETKFERQHMDSSTSSSSNYCNQMMKSRNLTKDRCKPVNTFVHESLDVQVCSQKNVCKNQTNCYQSYSTMSITDCRETSSKYPNCYKTTQNKHIIVCENPYVPVHFDSV
Ribonuclease - Final
300
200
100
0
-100
-200
-300
KETKFERQHMDSSTSSSSNYCNQMMKSRNLTKDRCKPVNTFVHESLDVQVCSQKNVCKNQTNCYQSYSTMSITDCRETSSKYPNCYKTTQNKHIIVCENPYVPVHFDSV

Fig. 3.

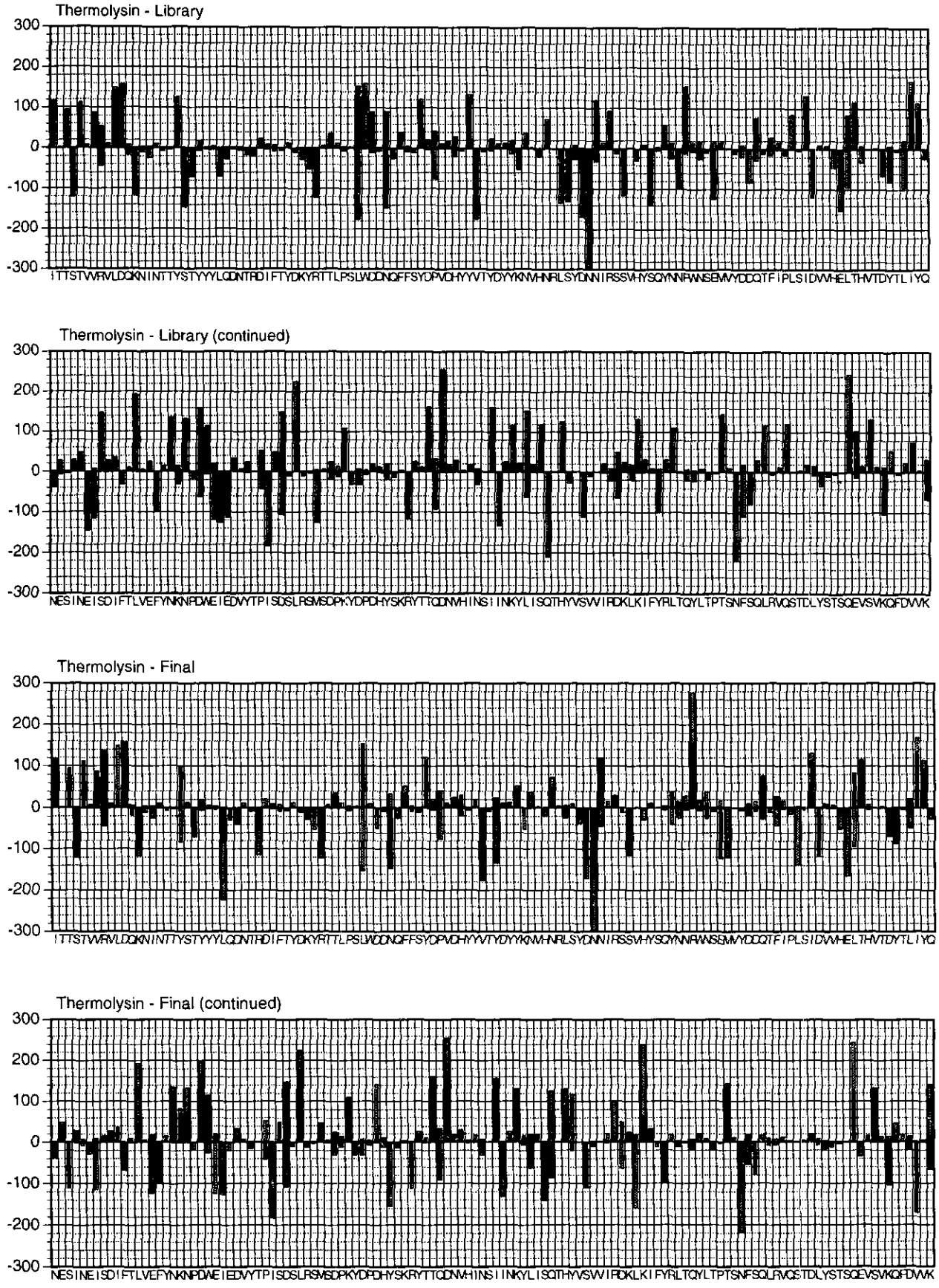


Figure 3. Results for χ_1 and χ_2 for 6 proteins calculated from backbone co-ordinates only. Results for the structure generated from the library and after the minimizations are completed as shown for each protein. The sequence and number of each residue is listed below the x-axis, and the bars represent deviations (in deg.) from the crystal structure for each side-chain. Deviations for χ_1 (filled bars) and χ_2 (hatched bars) are stacked, so that the deviation in χ_2 is given by the length of each hatched bar, rather than by the summed length of the filled and hatched bars.

Table 7
Predictions of side-chain conformations from backbone co-ordinates

	Structure 0 (library)			Structure 1			Structure *N*			Backbone-independent lib.
χ	No. cor.	No. inc.	Fraction correct	No. cor.	No. inc.	Fraction correct	No. cor.	No. inc.	Fraction correct	Fraction correct
A. *Thermolysin (N = 3)*										
1	169	83	0·67	180	72	0·71	187	65	0·74	0·52
2	131	48	0·73	131	48	0·73	130	49	0·73	0·69
1+2	149	103	0·59	156	96	0·62	161	91	0·64	0·43
B. *Bovine pancreatic trypsin inhibitor (N = 2)*										
1	37	9	0·80	41	5	0·89	39	7	0·85	0·65
2	32	9	0·78	33	8	0·80	32	9	0·78	0·71
1+2	29	17	0·63	34	12	0·74	33	13	0·72	0·52
C. *Crambin (N = 2)*										
1	32	5	0·86	34	3	0·92	34	3	0·92	0·68
2	23	4	0·85	23	4	0·85	24	3	0·89	0·81
1+2	30	7	0·81	32	5	0·86	33	4	0·89	0·59
D. *Lysozyme (N = 4)*										
1	80	25	0·76	81	24	0·77	81	24	0·77	0·64
2	57	28	0·67	59	26	0·69	56	29	0·66	0·67
1+2	63	42	0·60	67	38	0·64	64	41	0·61	0·48
E. *Rhizopuspepsin, C-terminal domain (N = 2)*										
1	82	33	0·71	89	26	0·77	94	21	0·82	0·54
2	65	11	0·86	67	9	0·88	62	14	0·82	0·83
1+2	78	37	0·68	87	28	0·76	88	27	0·77	0·50
F. *Ribonuclease A (N = 3)*										
1	79	30	0·72	78	31	0·72	86	23	0·79	0·56
2	55	20	0·73	54	21	0·72	53	22	0·71	0·71
1+2	68	41	0·62	70	39	0·64	76	33	0·70	0·42

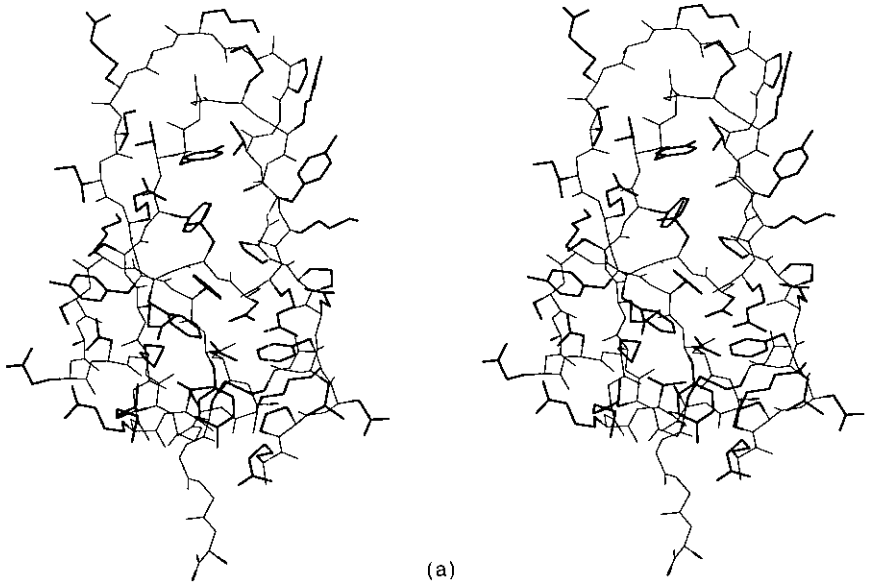
χ angle results are listed for χ_1,χ_2 and χ_{1+2}. The latter is defined as those side-chains with both χ_1 and χ_2 correct to within 40° of the experimental structure. Side-chains with only a single χ angle (e.g. Ser) are also included in χ_{1+2}. No. cor. refers to the number of residues correct (within 40°) for the given protein (for χ_1, χ_2 and χ_{1+2}), and No. inc. refers to the number incorrect. The fraction correct is equal to No. cor./(No. cor. + No. inc.), and is given for each structure named at the top of the Table. The structure numbers (0, 1, *N*) refer to the numbers in Fig. 1, with 0 being the backbone-dependent library prediction, 1 being the prediction after the first refinement, and *N* being the final structure after *N* refinement cycles. The value of *N* is given for each protein after its name. In the final column are predictions based on a backbone-independent library, according to the most prevalent conformations across the ϕ,ψ map (Table 4). With the exception of Met, these are the results that would be predicted by the rotamer library of Ponder & Richards (1987).

amino acids except for leucine χ_{1+2}. Leucine can exhibit very different χ angles and be nearly coincident in the atom positions. Lee & Subbiah (1991) note that if χ_1 is altered by 30° to 40° and χ_2 is changed by 150° to 140°. The C^δ atoms are nearly superimposable on the initial structure, while C^γ is shifted only slightly. Of the 19 leucinyl residues that are incorrectly placed in the final structures of the six proteins, inspection of the dihedral errors indicates that nine of them are likely to be misplaced because of the positional degeneracy of the two conformations. It is likely that in the X-ray structure it is not possible to distinguish one conformation from the other, so that the low prediction rate for leucine may be deceptive.

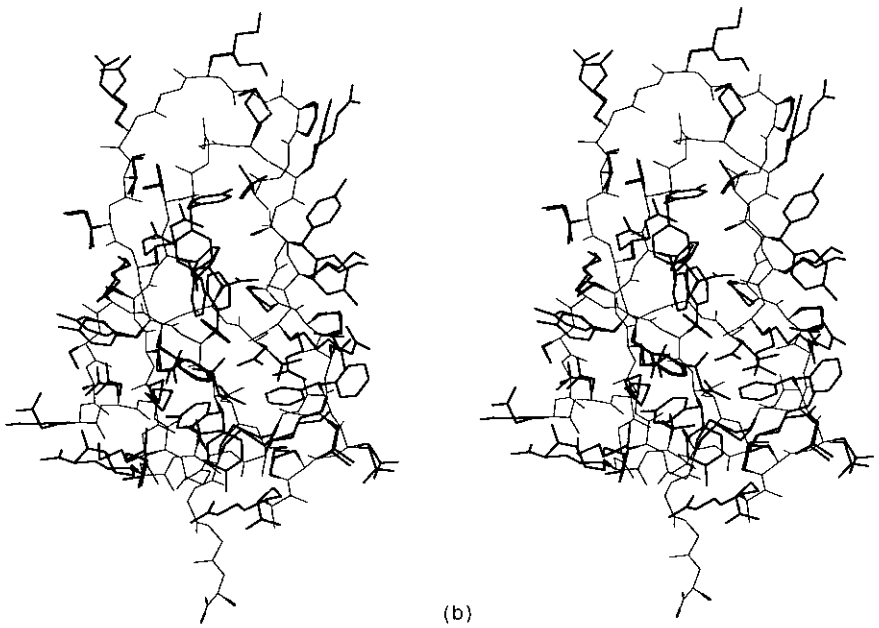
The library also performs well for aromatic amino acids, except Trp χ_2, which is greatly improved upon reorientation and minimization. Since one conformation of Trp is likely to clash with the backbone or side-chains of other residues, refinement often introduces the correct χ_2. The method does well for Cys, in part because all the Cys in the chosen proteins are involved in disulfide bridges. They minimize to correct conformations once the disulfide bond is established. Thr is predicted with much greater accuracy than Ser, because with two heavy atoms for Thr in the γ position, there is much less steric freedom in relation to the backbone. Since Ser is quite small and is able to form hydrogen bonds to the backbone, packing is often not the dominant interaction in determining its conformation.

The polar and charged amino acids are least well placed by the rotamer library, especially glutamic acid. While minimization improves most of them, only Gln reaches close to two-thirds of the values correct for χ_{1+2}. The χ_1 values are all in the 60 to 80% range in the final structures, but the χ_2 values are more variable in accuracy and therefore the χ_{1+2} results are poor. These amino acids depend on local hydrogen-bonding interactions with other side-chains and with solvent. Neither the library nor the potential energy function used in the refinement accounts for solvent effects in sufficient detail to predict their conformations well. Also, such side-chains may not have well-defined conformations, since they are often exposed and have high *B*-factors (Summers & Karplus, 1989).

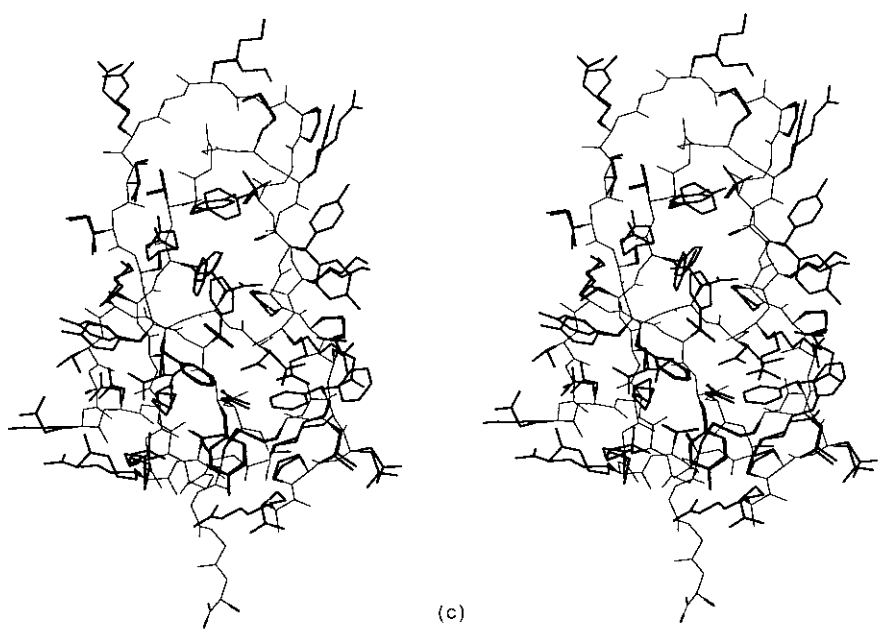
In Table 9, we list the r.m.s.d. for the side-chains obtained for the six predicted structures, and compare these with the results of Lee & Subbiah (1991). In most cases, our results compare favorably



(a)



(b)



(c)

Fig. 4.

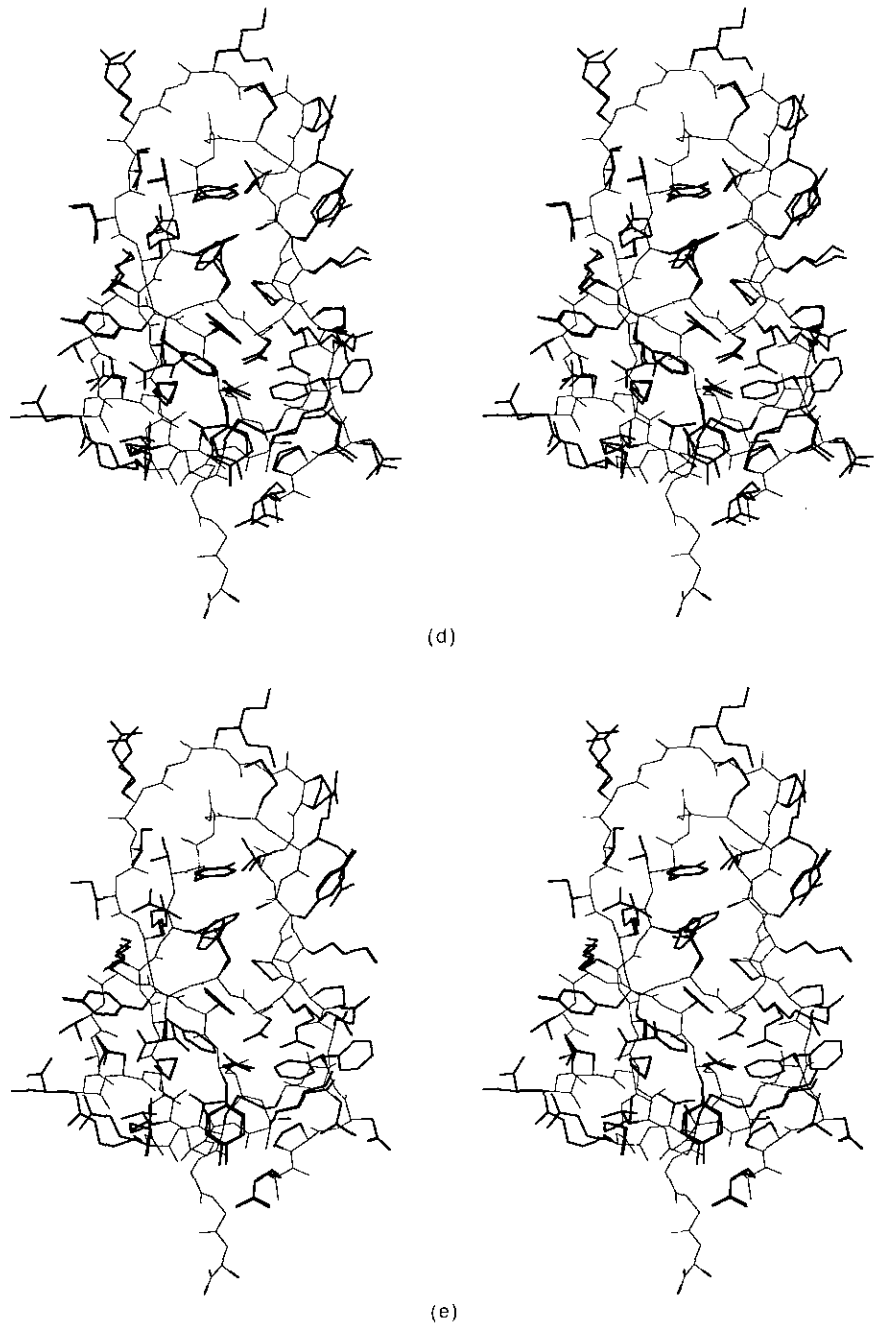
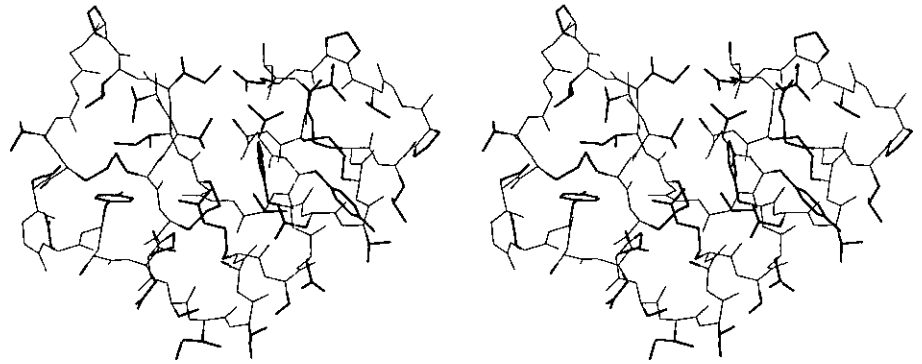


Figure 4. BPTI structures: (a) X-ray structure alone; (b) X-ray and backbone-independent library prediction; (c) X-ray and backbone-dependent library prediction; (d) X-ray and final predicted structure; (e) minimized X-ray and minimized final predicted structure (see the text).

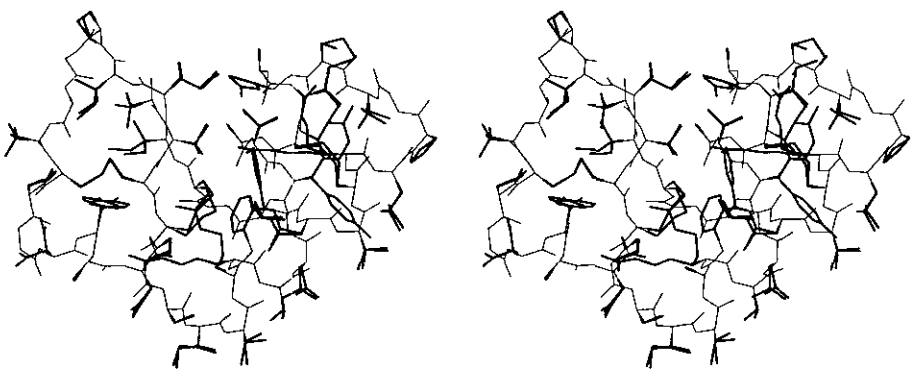
with those of Lee & Subbiah (1991). One should note the large variation in r.m.s.d. among the different kinds of side-chains. The bigger side-chains have larger r.m.s.d., even when they are well predicted, as in the case for phenylalanine. Averaging over all side-chains in a single protein gives results that depend heavily on the sequence, so we do not provide such averages here.

(c) *Method targ/temp applied to pen → rhi*

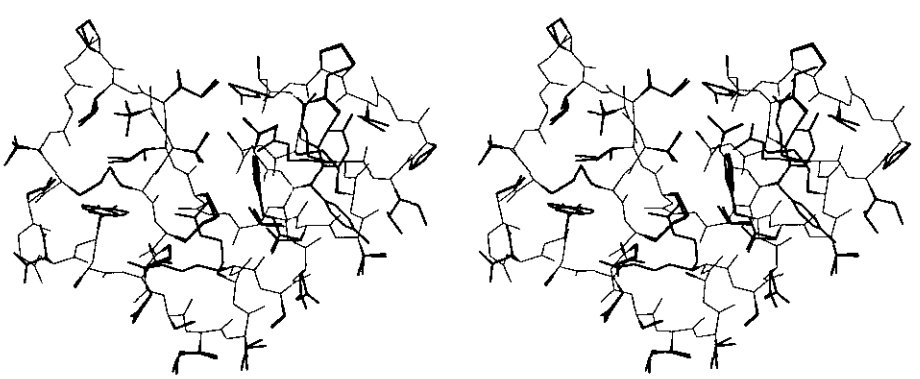
We applied the method described here to the problem studied by Summers & Karplus (1989). In that paper, information about side-chain conformations for the C-terminal domain of rhizopuspepsin was taken from the homologous protein, penicillopepsin (3 app) which has a 39% sequence identity with rhizopuspepsin. Cys, Pro and backbone coordinates were taken from the target X-ray structure (2 apr), and the other side-chains were modeled *via* their dihedral angles and rigid rotations as described in the Introduction. After the rigid rotations were completed (essentially equivalent to the final step in the present method), Summers & Karplus (1989) had predicted 86% of χ_1 and 75% of χ_2 correctly. With some additional checks and com-



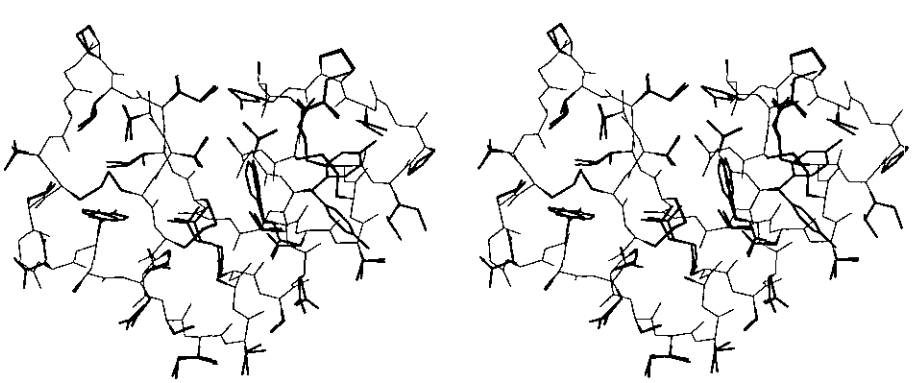
(a)



(b)

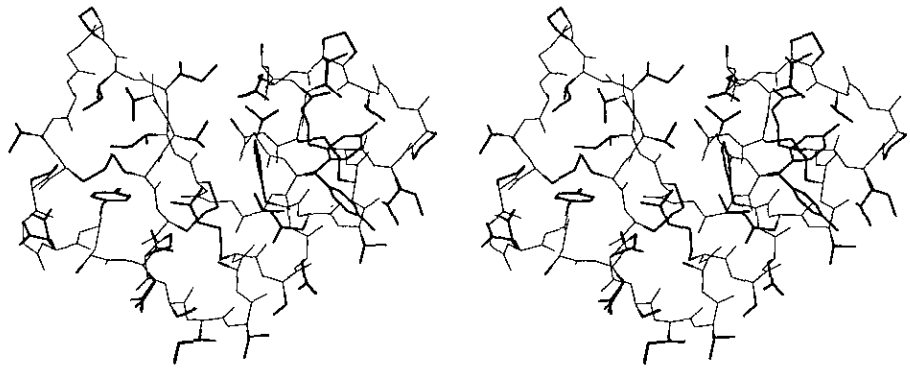


(c)



(d)

Fig. 5.



(e)

Figure 5. Crambin structures: (a) X-ray structure alone; (b) X-ray and backbone-independent library prediction; (c) X-ray and backbone-dependent library prediction; (d) X-ray and final predicted structure; (e) minimized X-ray and minimized final predicted structure (see the text).

parisons with the homologous protein, 92% and 81% accuracy was achieved for χ_1 and χ_2, respectively.

We performed the same calculation with Pro and Cys obtained from the target conformation, and built side-chains according to the homologous protein in combination with the rotamer library for the residues for which there was no information in the homolog (e.g. Gly → Asp). The final results are 88% of χ_1 and 80% of χ_2 correct, slightly worse than the 92% for χ_1 and 81% for χ_2 obtained by Summers & Karplus (1989). The method described here is simpler to apply than that of Summers & Karplus (1989), it is fully automated and does not require a homologous protein. However, the results obtained here without using the homologous protein are significantly worse for χ_1 (82%) and the same (80%) for χ_2. It is possible that some of the more complex refinement procedures used by Summers & Karplus (1989) could improve the present results.

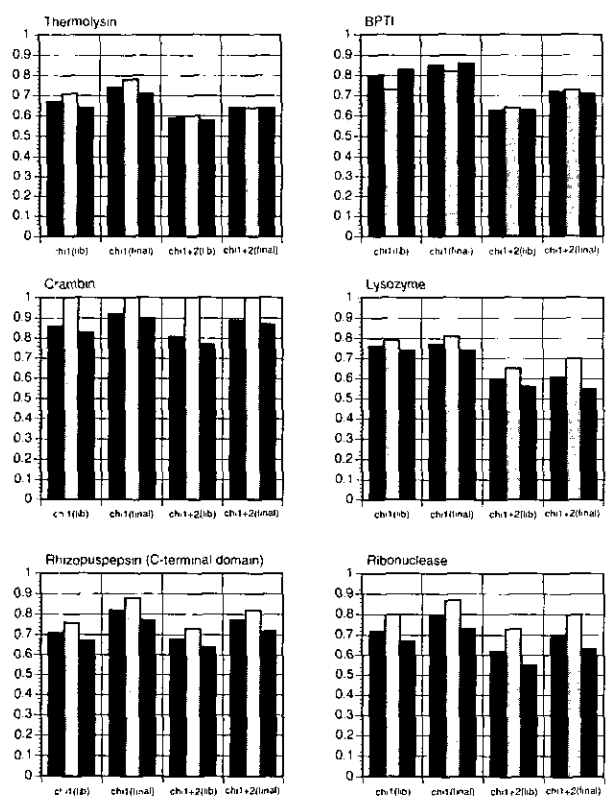


Figure 6. Fractions correct for all side-chains (black bars), buried side-chains (less than 10% exposure) (light grey bars), and surface side-chains (greater than 10% exposure) (dark grey bars) for the 6 proteins calculated from the backbones alone. Results for χ_1 alone and χ_{1+2} are shown from the library and after the minimizations are completed.

4. Discussion

An analysis has been made of the relation between the conformations of side-chains and the

Table 8

Results for six proteins for specific side-chain types

		Fractions correct					
Res.	No.	χ_1 Lib.	χ_1 Final	χ_2 Lib.	χ_2 Final	χ_{1+2} Lib.	χ_{1+2} Final
A. *Small, polar*							
Ser	63	0·57	0·65				
Thr	62	0·84	0·84				
Cys	30	0·90	0·93				
B. *Hydrophobic*							
Val	56	0·88	0·91				
Ile	44	0·84	0·86	0·86	0·84	0·73	0·73
Leu	37	0·59	0·68	0·49	0·54	0·49	0·49
Pro	29	0·86	0·79	0·83	0·76	0·83	0·76
Met	9	1·00	1·00	0·44	0·67	0·44	0·67
C. *Aromatic*							
Phe	30	0·67	0·83	0·90	0·77	0·57	0·70
Tyr	50	0·80	0·86	0·92	0·82	0·72	0·74
His	13	0·77	0·92	1·00	0·92	0·77	0·85
Trp	11	0·82	0·82	0·36	0·73	0·27	0·64
D. *Polar and charged*							
Asn	54	0·63	0·76	0·67	0·70	0·54	0·61
Asp	50	0·72	0·74	0·74	0·76	0·64	0·62
Gln	32	0·56	0·72	0·75	0·72	0·44	0·59
Glu	23	0·43	0·61	0·57	0·65	0·13	0·39
Arg	39	0·64	0·74	0·67	0·62	0·38	0·51
Lys	32	0·63	0·66	0·81	0·69	0·53	0·53

Table 9

Average root-mean-square deviation in Cartesian co-ordinates

	This work			Lee & Subbiah	
Res.	No.	r.m.s.d. X-ray: *N*	r.m.s.d. Min. X-ray: Min. *N*	No.	r.m.s.d. X-ray: Prediction
Cys	30	0·61	0·53	33	1·32
Ser	63	0·99	0·95	55	1·17
Thr	62	0·80	0·77	49	1·22
Val	56	0·63	0·60	74	0·97
Ile	44	0·91	0·88	55	0·89
Leu	37	1·51	1·45	63	1·00
Phe	30	1·88	1·72	41	1·29
Tyr	50	1·95	1·75	37	1·17
His	13	1·30	1·31	11	1·58
Trp	11	2·44	2·30	15	2·20
Pro	29	0·40	0·29	—	—
Asp	50	1·44	1·25	43	1·31
Asn	54	1·46	1·29	42	1·70
Glu	23	2·24	2·20	48	1·73
Gln	32	1·78	1·78	29	2·01
Met	9	1·09	1·03	17	1·27
Lys	32	2·16	2·14	46	2·72
Arg	39	2·88	2·86	35	3·40

Root-mean-square deviation in heavy-atom Cartesian co-ordinates is given for each amino acid type across the 6 proteins tested in this paper. For symmetric residues, r.m.s.d. for both χ_2 and $\chi_2+180°$ (or χ_3 and $\chi_3+180°$) were tested, and the lower value was used. In addition to the deviations of Structure *N* from the X-ray structure (column 3), the deviations of a CHARMM minimized Structure *N* from a CHARMM minimized X-ray structure are given in column 4. The minimization procedure used is the same as that in Summers & Karplus (1989), and consists of minimizing all of the side-chains simultaneously subject to gradually reduced harmonic constraints. For comparison, the results of Lee & Subbiah (1991) are given for the 9 proteins tested in that paper (column 6).

local backbone geometry in proteins for which high-resolution structures are available. The results show that the most probable side-chain dihedral angle values are affected by the ϕ and ψ angles of the local backbone. This relationship is of interest for protein folding and for structure prediction. Based on the results, a backbone-dependent library of side-chain rotamers has been developed from the available high-resolution structures. The portions of the library that are well populated tend to have specific side-chain conformational preferences. This result provides a basis for understanding the success of the side-chain placement studies that make use of backbone templates with similar ϕ,ψ angles (see Introduction).

The backbone-dependent rotamer library serves as the starting point for a prediction scheme of side-chain orientations from the backbone co-ordinates. After the initial placement, the side-chain positions are refined by reorientation and energy calculations to eliminate side-chain–backbone and side-chain–side-chain van der Waals repulsions. This iterative procedure, which scales approximately linearly with the size of the protein, leads to the results that 78% of χ_1, 74% of χ_2 and 69% of χ_{1+2} are correctly predicted for a set of six proteins ranging in size from crambin to thermolysin.

The results obtained here have a number of implications for studies of protein folding and structure. The library with or without side-chain minimization provides a starting point for building full protein models from crystallographic backbone co-ordinates that can be refined with the experimental structure factors. Also, the results demonstrate that model building from template protein backbones is feasible and may be sufficiently reliable to be used in drug design. The approach used here, which is an extension of the work of Summers & Karplus (1989), can serve as a starting point for such model building.

The mutual influence of backbone and side-chain conformations may have a role in protein folding since there is a reduction in the conformational space that must be searched in the actual folding process and in theoretical model studies. It was pointed out some time ago (Gelin & Karplus, 1975) that the side-chain conformations in a protein tend to correspond to minima that are selected from those that exist in the isolated dipeptide. This concept was embodied in the rotamer library, independent of the backbone conformation, that was proposed by Ponder & Richards (1987). The present results go further and indicate that the local backbone structure can play an important role in the selection process, though neighboring side-chains and tertiary contacts are also involved. The latter may be most important in the stabilization of given side-chain conformer rather than in its selection. In the limit, this implies that in the process of protein folding, the correct backbone and side-chain dihedral angles are introduced in a concerted fashion.

Both van der Waals exclusions between the backbone and side-chains and the tendency toward the g^-, t and g^+ conformations severely limit the conformation space that a side-chain can occupy, and reduce by many orders of magnitude the space that must be searched to find a structure with no repulsive overlaps of side-chains. This conclusion is supported by the recent work of Desmet *et al.* (1992).

Although the test application of the side-chain prediction scheme described here are quite successful, it should be noted that we have not examined the effect on the predictions of errors in the backbone positions. Also, there are a number of possibilities for improvements of the method. Polar and charged residues are least well predicted. For the protein interior, inclusion of hydrogen bonding and other electrostatic interactions may be useful. For surface side-chains, the prediction problems may be due in part to the fact that such side-chains have multiple conformations. However, lack of explicit solvent in the model is also a limitation, since it must affect the conformations of such hydrogen-bonding side-chains. It is possible to modify the potential energy function used for the iterative minimizations to mimic the effect of solvent for exposed residues. Wesson & Eisenberg

(1992) have recently modified the CHARMM potential to explicitly favor solvent accessibility for polar atoms (N and O) and favor solvent inaccessibility for non-polar atoms (C) by use of surface-area dependent corrections. A similar approach has been used by Schiffer *et al.* (1992) in a study of the alanine dipeptide. Another approach is to add terms of the RISM type (Pettitt & Karplus, 1985; Ramé *et al.*, 1990), where solute–solvent correlation functions are used to calculate the effect of the solvent on solute–solute interactions. Both of these solvent corrections are being incorporated into the prediction program to determine their effects on the accuracy of surface side-chain predictions. Finally, as Ponder & Richards (1987) have described, the mean positions of many rotamers do not lie exactly at 60°, 180° and −60°, and so slightly different orientations could be used in the placement. However, even without such improvements, the method proposed should be useful in a variety of applications.

This work was supported in part by a grant from the National Science Foundation and Polygen/Molecular Simulations, Inc. The calculations were performed on a Convex C220 and a Silicon Graphics SGI 340. We thank Hsiang-ai Yu for helpful discussions, and Roland Stote and Aaron Dinner for technical assistance.

Note: A copy of the full backbone-dependent rotamer library is available upon request. Write to R.L.D. or M.K or send electronic mail to dunbrack@tammy.harvard.edu.

References

Bhat, T. N., Sasisekheran, V. & Vijayan, M. (1979). An analysis of side-chain conformations in proteins. *Int. J. Pept. Protein Res.* **13**, 170–184.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.

Bruccoleri, R. E. & Karplus, M. (1987). Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers*, **26**, 137–168.

Cung, M. T., Vitoux, B. & Marraud, M. (1987). Flexibility of Pro-Pro sequences: IR and NMR experiments. *New J. Chem.* **11**, 503–510.

Desmet, J., DeMaeyer, M., Hazes, B. & Lasters, I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature (London)*, **356**, 539–542.

Gelin, B. R. & Karplus, M. (1975). Sidechain torsional potentials and motion of amino acids in proteins: bovine pancreatic trypsin inhibitor. *Proc. Nat. Acad. Sci., U.S.A.* **72**, 2002–2006.

Gelin, B. R. & Karplus, M. (1979). Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry*, **18**, 1256–1268.

Holm, L. & Sander, C. (1991). Atomic structure of the actin—DNase I complex. *J. Mol. Biol.* **218**, 183–194.

James, M. N. G. & Sielecki, A. R. (1983). Structure refinement of penicillopepsin at 1·8 Å resolution. *J. Mol. Biol.* **183**, 299–361.

Janin, J., Wodak, S., Levitt, M. & Maigret, B. (1978). Conformation of amino acid side-chains in proteins. *J. Mol. Biol.* **125**, 357–386.

Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. (1990). *Nature (London)*, **347**, 37–44.

Kendrew, J. *et al.* (1970). IUPAC-IUB commission on biochemical nomenclature. Abbreviations and symbols for the description of the conformation of polypeptide chains. *Biochemistry*, **9**, 3471–3479.

Konnert, J. H. & Hendrickson, W. A. (1980). A restrained-parameter thermal-factor refinement procedure. *Acta Crystallogr. sect. A*, **36**, 344–350.

Lee, C. & Subbiah, S. (1991). Prediction of protein side-chain conformation by packing optimization. *J. Mol. Biol.* **217**, 373–388.

McGregor, M. J., Islam, S. A. & Sternberg, M. J. E. (1987). Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* **198**, 295–310.

Pettitt, B. M. & Karplus, M. (1985). The potential of mean force surface for the alanine dipeptide in aqueous solution: a theoretical approach. *Chem. Phys. Letters*, **121**, 194–201.

Ponder, J. W. & Richards, F. M. (1987). Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775–791.

Ramé, G. L., Lau, W. F. & Pettitt, B. M. (1990). Flexibility of tripeptides in solution: free energy molecular mechanics. *Int. J. Pept. Protein Res.* **35**, 315–327.

Reid, L. S. & Thornton, J. M. (1989). Rebuilding flavodoxin from C_α coordinates: a test study. *Proteins: Struct. Funct. Genet.* **5**, 170–182.

Sali, A., Overington, J. P., Johnson, M. S. & Blundell, T. L. (1990). From comparisons of protein sequences and structures to protein modelling and design. *Trends Biochem. Sci.* **15**, 235–240.

Schiffer, C. A., Caldwell, J. W., Stroud, R. M. & Kollman, P. A. (1992). Inclusion of solvation free energy with molecular mechanics energy: alanyl dipeptide as a test case. *Protein Sci.* **1**, 396–400.

Shih, H. H.-L., Brady, J. & Karplus, M. (1985) Structure of proteins with single-site mutations: a minimum perturbation approach. *Proc. Nat. Acad. Sci., U.S.A.* **82**, 1697–1700.

Straatsma, T. P. & McCammon, J. A. (1992). Alchemical free energy simulation. *Annu. Rev. Phys. Chem.* **43**, 407–435.

Summers, N. L. & Karplus, M. (1989). Construction of side-chains in homology modelling. Application to the C-terminal lobe of rhizopuspepsin. *J. Mol. Biol.* **210**, 785–812.

Summers, N. L. & Karplus, M. (1990). Modeling of globular proteins: a distance-based data search procedure for the construction of insertion/deletion regions and Pro ↔ non-Pro mutations. *J. Mol. Biol.* **216**, 991–1016.

Tidor, B. & Karplus, M. (1991). Simulation analysis of the stability mutant R96H of T4 lysozyme. *Biochemistry*, **30**, 3217–3228.

Tuffery, P., Etchebest, C., Hazout, S. & Lavery, R. (1991). A new approach to the rapid determination of protein sidechain conformations. *J. Biomol. Struct. Dynam.* **8**, 1267–1289.

Wendoloski, J. J. & Salemme, F. R. (1992). PROBIT: a statistical approach to modeling proteins from partial coordinate data using substructure libraries. *J. Mol. Graph.* **10**, 124–127.

Wesson, L. & Eisenberg, D. (1992). Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1**, 227–235.
Wüthrich, K. (1989). The development of nuclear magnetic resonance spectroscopy as a technique for protein structure determination. *Acc. Chem. Res.* **22**, 36–44.

Edited by B. Honig

Chapter 4

Conformational Analysis of the Backbone-Independent and Backbone-Dependent Rotamer Preferences of Protein Sidechains

Abstract

An analysis of protein sidechain/sidechain and sidechain/backbone interactions is performed using simple stereochemical principles and the molecular mechanics program CHARMM. Experimental data such as infrared and Raman spectroscopy as well as *ab initio* calculations have shown that steric interactions between the terminal carbon atoms in such molecules as butane and pentane (interacting through one and two dihedral degrees of freedom respectively) determine the relative energies of their local minima. In particular, gauche energy minima (C-C-C-C dihedral (χ) ~ 60° or -60°) are approximately 0.8 kcal/mol higher than trans (χ = 180°). In pentane, the {g+,g+} and {g-,g-} minima are 1.4 kcal/mol higher in energy than trans,trans or approximately twice the single gauche energy. But when the dihedrals are of opposite sign (syn-pentane conformations), the energy is significantly higher, 3.2 kcal/mol, because of the close proximity of the terminal carbon atoms. Gauche and syn-pentane interactions between backbone N and C atoms and sidechain heavy atoms are found to determine backbone-conformation independent rotamer preferences in proteins. We tabulate these interactions and compare the results with the backbone-independent rotamer library.

Backbone-conformation dependent rotamer preferences are determined by interactions not only with backbone N and C of residue i, but also with backbone atoms C_{i-1}, O_i, and N_{i+1}. The positions these atoms relative to the sidechain atoms of residue i are dependent on the backbone dihedral angles ϕ and ψ. The appropriate dihedrals between these backbone atoms and sidechain atoms are calculated, and used to predict the effect of backbone conformation on rotamer preferences calculated independently of the backbone. CHARMM calculations are performed to calculate backbone-independent and backbone-dependent rotamer preferences. These are compared with the steric analysis and data from a backbone-conformation dependent rotamer library we have calculated from 184 protein chains in 175 protein structures in the Brookhaven Protein Databank.

In most cases, the calculations agree relatively well with the experimental distributions of rotamers. The exceptions are discussed.

I. Introduction

Rotamer libraries [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11] describe the probabilities of amino acid sidechain conformations in proteins. They have been used in structure determination [12] and structure prediction in combination with energy minimization schemes [11, 13, 14], information from homologous proteins [9], and Monte Carlo optimization and/or simulated annealing [15, 16]. Such rotamer libraries have been determined from available crystal structures in databanks such as the Brookhaven Protein Databank.

As a larger number of higher resolution structures have become available in recent years, distributions of χ angles have sharpened [5], and it has become possible to analyze rotamer preferences as a function of backbone conformation [3, 8, 9, 11]. Janin et al. [3] compiled the χ_1 rotamer populations for sidechains with a single γ heavy atom (Trp, Tyr, Phe, Met, Lys, Asp, Asn, Glu, Gln, Lys, Arg, His) summed together from 19 proteins for four regions of the ϕ,ψ map. They defined the regions as follows: β having $\phi < 0°$ and $30° < \psi < 210°$; α_R having $\phi < 0°$ and $-120° < \psi < 30°$; α_L having $\phi > 0°$ and $-60° < \psi < 90°$; and "other" as all other regions of the ϕ,ψ map not included in the first three regions. Because of the limited data then available, their results are summed over large areas of the ϕ,ψ map. As a result, they found only a weak correlation of the preferred χ_1 rotamer with secondary structure. The g+, t, and g- populations (χ_1 near +60°, 180°, and -60° respectively) were 10%, 40%, and 50% in the β-sheet region, 10%, 33%, and 57% in the α_R region, and 3%, 24%, and 73% in the α_L region. McGregor et al. [8] compiled statistics for the three χ_1 rotamers for 17 residues types (excluding Ala and Gly because they do not have χ_1 rotamers and Pro because the ring structure allows only two predominant structures) for six kinds of secondary structure: α-helix (center), α-helix (N-end), α-helix (C-end), β-sheet (center), β-sheet (ends), and non-α/non-β. Their database consisted of 61 proteins solved at a resolution of 2.0 Å or better. Their results for individual amino acids demonstrated correlations of backbone and sidechain

conformations for certain kinds of sidechains, especially those with two γ heavy atoms (Thr, Val, and Ile) and aromatic sidechains. They also presented average χ angles and standard deviations for each amino acid and secondary structure type. Sutcliffe et al. [9] used a set of 11 globin and immunoglobulin proteins to calculate the most preferred χ_1, χ_2 (and χ_3 and χ_4 where applicable) values for all the sidechains in α and β conformations. They found that Arg, Glu, His, Leu, Lys, Met, Thr and Trp had different predominant rotamers in the α and β regions. This information was then used to build other globin and immunoglobulin structures by homology modeling.

With an extended database of 132 protein chains with resolution better than or equal to 2.0 Å, we recently compiled a backbone-dependent rotamer library [11] (Chapter 3 of this thesis) for individual amino acids. The rotamer library gave the distribution for each occupied 20° by 20° region of the ϕ,ψ conformation space of the backbone. The variation of sidechain rotamer preferences in different 20° by 20° cells of the Ramachandran map was significant. χ_1 rotamer preferences showed detectable patterns as a function of ϕ and ψ for all sidechains. In the β-sheet region for example, a number of sidechains (e.g. Phe, Tyr, Met, Lys, Arg) were found to have preferences for g+ rotamers ($0° \le \chi_1 \le 120°$) over the t and g- rotamers ($120° \le \chi_1 \le 240°$ and $-120° \le \chi_1 \le 0°$ respectively) when ϕ was between -180° and -150° and ψ was between 150° and 180°. Through the rest of the β region, g- rotamers were preferred in most 20° by 20° cells of the ϕ,ψ map, except when $\phi < -140°$ and $\phi > -70°$ while $\psi < 140°$ where t rotamers were most populous. Many other correlations with ϕ and ψ values were observed. The library was shown to be useful as a tool for predicting sidechain conformations from known backbone coordinates. The results obtained were a significant improvement over backbone-independent rotamer libraries [7].

In this paper, we show how the well-established principles of conformational analysis can be applied to protein sidechains to rationalize their backbone-independent and backbone-dependent rotamer preferences. The field of "conformational analysis" in

organic chemistry has a long history [17], and generally refers to "the analysis of the physical and chemical properties of a compound in terms of the conformation (or conformations) of the pertinent ground states, transition states, and (in the case of spectra) excited states." [18, p. 1]. Early work in the field concentrated on the conformations of cyclohexane (reviewed by Hanack [19], pp. 13-21) and the difference in energy between the eclipsed and staggered forms of hydrocarbons such as ethane (reviewed by Pitzer [20]). The idea that organic molecules possess "conformations" which can interconvert at room temperature via rotation about single bonds and that these conformations determine their properties only gained widespread notice in the early 1950's from the pioneering work of D. H. R. Barton on cyclohexane derivatives and the steroid nucleus [21, 22, 23].

The conformational analysis of a molecule can encompass a number of techniques both experimental and theoretical [19, p. 24-26]. Experimental methods include x-ray crystallography, NMR, electron diffraction, and microwave, Raman, UV, and infrared spectroscopy. These techniques can be used to determine the conformations of a molecule, their relative energies, and the barriers to interconversion between them. Also, the chemical reactions of model compounds and the analysis of the products can be used to infer information about conformations and energy distribution. Theoretical methods run from simple enumeration of torsional, steric, and electrostatic interactions that one would expect from the covalent structure of the molecule to molecular mechanics and *ab initio* calculations of the structures and relative energies of potential energy minima and barriers between them on the molecule's rigid rotor or adiabatic potential energy surfaces.

The conformational analysis of protein sidechains presented in this work consists of three elements. The first is the presentation of the experimentally determined backbone-independent and backbone-dependent rotamer preferences from an updated survey (cf. Dunbrack and Karplus Ref. [11]) of x-ray structures from the Brookhaven Protein Databank. The new libraries are calculated from 184 protein chains determined to

2.0 Å resolution or better. Average χ values and root-mean-square deviations from these averages are presented for the backbone-independent rotamers of 18 sidechain types. The populations of all χ_1 rotamers for the 18 sidechain types are shown as a function of ϕ and ψ in 10° intervals for values of $\phi \leq -40°$ and $-180° \leq \psi \leq 180°$. For $\phi > -40°$ (where populations are very low), only the preferred rotamers are given. Certain patterns can be easily observed in the rotamer preferences of sidechains as a function of the backbone conformation dihedrals ϕ and ψ. We compare the abilities of the backbone-independent and backbone-dependent libraries to predict the conformations of sidechains in the Protein Databank, and show that backbone-conformation effects are significant.

The second element involves the enumeration of steric interactions of protein sidechains with the backbone and other sidechain atoms. These follow from the well-known electronic and steric effects on rotation about single carbon-carbon bonds. The energies and structures of the rotational isomers of alkanes have been studied since the early 1930's [24]. Until 1937, the methyl groups in ethane were thought to rotate freely without hindrance. To explain spectroscopic data and specific heats, Kemp and Pitzer [25] suggested that there must be a barrier to rotation in ethane to be 3.0 kcal/mol occurring when the hydrogen atoms were "eclipsed" (i.e. with a dihedral H-C-C-H=0°). The minima on the potential energy surface therefore occurred with the hydrogens "staggered" with each H-C-C-H dihedral equal to +60°,180° or -60°. Pitzer [26] used simple potential energy calculations and heat capacity and entropy data to predict the energies of the rotational isomers of butane and longer hydrocarbon chains. From steric hindrances about the central C-C bond of butane, he found that the three minimum energy conformations of butane were of different energy. The lowest energy conformation was the trans structure with a C-C-C-C (χ) of 180°. In addition there were two gauche structures with χ equal to +60° and -60°. To fit the experimental entropies of n-butane and n-heptane, Pitzer assigned an energy of 0.8 kcal/mol to the gauche minima relative to the global minimum trans conformation.

In order to explain the experimental thermodynamics of longer hydrocarbons, Pitzer realized that certain combinations of neighboring dihedrals in longer hydrocarbons were not allowed because of steric hindrance between carbon atom i and i+4 of the chain [26, 27]. When gauche dihedrals followed trans dihedrals (or vice versa) the energy of the gauche conformation of butane (0.8 kcal/mol) was used. Similarly, when two gauche dihedrals of the same sign followed one another, an energy approximately equal to twice the butane gauche energy was used. But when gauche dihedrals of opposite sign followed one another (i.e. g+ followed by g- or g- followed by g+), steric hindrances occur between the terminal carbon atoms which end up quite close in space (these interactions are sometimes referred to as "syn-pentane" interactions). Pitzer assigned these an infinite energy in a statistical mechanical analysis of saturated hydrocarbon chains, and found reasonable agreement with experimental entropies and heat capacities. Hoeve [28] and Abe et al. [29] have used the experimental variation of the end-to-end distance of hydrocarbon chains with temperature to estimate the energy of the {g+,g-} and {g-,g+} elements in hydrocarbon chains. Hoeve found that values between 2.3 and 2.6 kcal/mol for the difference in energy between {g+,g-} and {t,g±} were necessary to reproduce the experimental data. Abe et al. estimated the difference to be between 2.5 and 3.0 kcal/mol. Using semi-empirical energy expressions, they found pairs of {g+,g-} and {g-,g+} minima displaced significantly from the expected {+60°,-60°} dihedrals. The minima were located at $\{\chi_1,\chi_2\}$ values of $\{\pm 65°, \mp 100°\}$ and $\{\pm 100°, \mp 65°\}$.

Recent high level *ab initio* calculations [30] and experimental measurements [31, 32] have shown that the single gauche interaction in butane is about 0.9 kcal/mol higher than the global minimum trans conformation (180°). The *ab initio* energy of two consecutive gauche interactions of like sign in pentane was found to have an energy of 1.4 kcal/mol above the global minimum {t,t} conformation, while the syn-pentane {g+,g-} conformation had an energy of 3.3 kcal/mol above {t,t}.

Becker has described a counting-method ("Abzählverfahren") for calculating the steric energies of rotational isomers of branched and unbranched alkanes [33]. The method consisted of considering the 3n conformations of an alkane molecule, where n is the number of rotational degrees of freedom (excluding methyl rotations which do not produce different conformations) and enumerating the gauche and steric syn-pentane interactions present in each conformation. He ascribed an energy "a" of 0.5 kcal/mol to each gauche interaction and an additional 5a or 2.5 kcal/mol to each {g+,g-} interaction. The total energy for a {g+,g-} pair was therefore 6a or 3.0 kcal/mol over a t,t conformation. Becker estimated the {g+,g-} energy (relative to {t,g-} or {t,g+}) from differences in the heats of formation of 2,2,4,4 tetramethyl pentane from n-nonane and 2,2,4 trimethyl pentane from n-octane. The conformational partition function was then used to calculate the thermodynamic properties of each molecule from the energies of the various isomers estimated from the gauche and {g+,g-} steric interactions of each conformation. Enumerating the gauche and syn-pentane interactions and using the energy values of the individual interactions to calculate the sum for each molecule, Becker successfully reproduced the relative experimental reactivities of complex alkanes.

In this paper, we perform a similar analysis on the backbone-independent and backbone-dependent rotamers of protein sidechains. The method consists of enumerating gauche and syn-pentane interactions of sidechain heavy atoms with other sidechain atoms and atoms of the backbone. In the case of backbone-independent rotamers, only the backbone nitrogen and carbonyl carbon atoms are considered, since their positions relative to the sidechain atoms do not depend on the backbone conformation dihedrals ϕ and ψ. We show that in the case of non-polar sidechains enumerating these interactions works well in predicting which sidechain conformations are essentially forbidden in all backbone conformations because of large steric energies arising from {g+,g-} and {g-,g+} conformations.

In the more complicated case of the backbone-dependent rotamers, the positions of backbone heavy atoms which can form {g+,g-} or {g-,g+} steric hindrances with sidechain γ atoms must be considered. These backbone atoms include the carbonyl carbon preceding the ith residue, whose position is determined chiefly by the backbone dihedral ϕ, and the carbonyl oxygen of the ith residue and the backbone nitrogen of the succeeding residue whose positions are determined by the backbone dihedral ψ. By tabulating gauche and syn-pentane interactions between sidechain and backbone atoms in all possible backbone conformations and local energy minima conformations of the sidechain, we show that these simple steric interactions can be used to explain most of the features of the backbone-dependent experimental rotamer library.

The third element of our conformational analysis of sidechain rotamer preferences consists of using molecular mechanics calculations to estimate the energies of sidechain rotamers. A number of groups have used energy calculations to locate steric and electrostatic interactions amongst sidechain atoms and between sidechain atoms and backbone atoms that could be involved in determining sidechain conformational preferences. The earliest calculations by Ramachandran and coworkers [1, 34, 35] used hard sphere models to locate steric clashes, and compared the results with the three structures then available – myoglobin, lysozyme, and chymotrypsin. Subsequent models included van der Waals potentials and electrostatic interactions in rigid rotations of sidechains using experimental bond lengths and bond angles [3, 6, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46]. Later, adiabatic molecular mechanics calculations were performed with bond lengths, bond angles, and torsional angles allowed to vary to locate potential energy local minima and energy barriers for individual sidechains [47, 48, 49]. *Ab initio* and semi-empirical quantum mechanical calculations have also been used [50].

Theoretical work to explain backbone-independent rotamer preferences have used a model for each residue type by fixing ϕ and ψ of the dipeptide of the given residue [3] or by allowing them to relax to their energy minimum [38, 39]. Plots of the energy as a

function of χ_1 or χ_1 and χ_2 have been compared to the backbone-independent rotamer preferences, and steric interactions amongst sidechain atoms and between sidechain atoms and the residue backbone noted. A number of groups have performed calculations to explain the distribution of χ_1 rotamers across the Ramachandran map by calculating the energy of the dipeptides (N-acetyl,N′-methylamide) of each residue for each rotamer as a function of ϕ and ψ [3, 40, 41, 50]. Contour maps of the energy as a function of ϕ and ψ were plotted for each χ_1 rotamer (viz., 60°, 180°, and -60°), and compared with the experimental distributions by marking the crystallographic ϕ,ψ points on the appropriate graph given χ_1. These calculations emphasized the limitations that individual sidechains place on the backbone conformation, rather than the limitations the backbone places on the sidechain conformation. They were useful in demonstrating the steric interactions of sidechain and backbone atoms that help determine protein conformations.

We have used the program CHARMM to calculate the energies of sidechain rotamers in both the backbone-independent and backbone-dependent contexts. The former was accomplished by considering the atoms in a single residue fragment consisting of only N, Cα, Hα, C, and the residue sidechain. The latter was accomplished by performing calculations on the dipeptide of each residue type (i.e. N-acetyl-Xxx-$NHCH_3$). For all of these calculations, we have used the CHARMM22 potential energy function, which includes all atoms (polar and non-polar hydrogens) and which has been optimized to represent a variety of intramolecular and intermolecular interactions in proteins [51]. We calculated the energies of all the possible rotamers for the single residue fragment and for the dipeptide as a function of the backbone dihedrals ϕ and ψ in 10° intervals. These energies can be used to predict rotamer populations of the backbone-independent and backbone-dependent libraries. With a much larger database than was available in the 1970's and a more refined potential energy function, these comparisons are more meaningful than those in earlier studies. Also, dividing the database into small intervals of ϕ and ψ provides greater detail in backbone/sidechain interactions than using

rotamer libraries based only on three or four types of secondary structure [8]. The comparison is of interest for understanding the source of the conformational preferences, the accuracy of the potential energy function, and the possible effects of the protein environment. The last has been shown [46] to contribute significantly to stabilizing a given minimum and to perturbing the χ values from their "ideal" positions.

In the Methods section, we provide details on the data and methods used to tabulate the experimental backbone-independent and backbone-dependent rotamer libraries, the enumeration of steric interactions, and the CHARMM calculations of both backbone-independent and backbone-dependent rotamers. In the following section, we present the results of the enumeration analysis and CHARMM calculations and compare these results with the rotamer libraries. The library used here is larger than the one in our previous work [11], and has been presented in a more convenient and useful format (using the Macintosh spreadsheet program Excel) for comparison with the steric analysis and CHARMM calculations.

For most non-polar sidechains and most areas of the ϕ,ψ map the energy calculations and enumeration of steric interactions agree with the experimentally derived preferences. While the steric enumeration is able to predict which conformations are not allowed for polar sidechains, the energy calculations do a poor job in reproducing the experimental data. We analyze the discrepancies in terms of the effect of the protein environment and possible defects in the parameterization.

II. Methods

A. The backbone-independent and backbone-dependent rotamer libraries

The rotamer library used in this paper is an extended version of the one we used earlier in a sidechain conformation prediction scheme [11]. The present library has 52 new structures (March 1993) that were not available in May 1992, when the previous library was compiled. The total number of protein chains is now 184 structures from 175

entries in the Brookhaven Protein Database at a resolution better than or equal to 2.0 Å. The proteins used are listed in Table 1. As before, several groups of homologous proteins are included. The new library contains 27,683 sidechains. Rotamers were defined according to certain limits on χ angles, and these are listed in Table 2 for each amino acid type. As in the previous paper, we assume that the χ_2 orientation of the His sidechain (either approximately +90° or -90°) is not uniquely determined in most crystal structures. The same is also likely to be true for Asn χ_2 and Gln χ_3.

A complete rotamer library was calculated for ϕ,ψ values 10° apart (-180°,-170°,...0°,...,170°,180°). At each point, sidechains were included if they had values of ϕ and ψ within 10° of the ϕ,ψ values at that point. With these values, 20° by 20° blocks were constructed, as in the previous paper. For example, the experimental distribution of phenylalanines at ϕ=-60°, ψ=150° includes all sidechains with $-70° \leq \phi < -50°$ and $140° \leq \psi < 160°$. Thus, each sidechain is counted four times in such a procedure. The purpose is to use a large enough window to obtain reasonable statistics for each value of ϕ and ψ. Rotamer populations for each χ_1 (i = 1,2) were calculated using the angular ranges listed in Table 2. For all sidechains (except Ala, Pro, and Gly), the χ_1 values correspond to the rotamers of a tetrahedral carbon atom. Alanine and glycine do not have sidechain dihedral angles, and proline has two rotamer populations with χ_1 averages of +21° and -20° (see Table 3). The choices for the various limits are discussed in Dunbrack and Karplus [11].

B. Enumeration of steric interactions

The steric analysis of backbone-independent and backbone-dependent rotamers was performed by counting the number of gauche and syn-pentane interactions of sidechain heavy atoms with other sidechain atoms and the backbone. In this treatment, all heavy atoms were considered to be equivalent. Each sidechain structural class was analyzed separately for both backbone-independent and -dependent rotamers. For

backbone-independent rotamers, gauche interactions of γ heavy atoms with backbone N were listed when χ_1 (the dihedral N-Cα-Cβ-Xγ for X=C, O, or S) took on the values of +60° or -60°. These interactions can be seen from the Newman projections in Figure 1 for the three χ_1 rotamers of a sidechain with a single γ heavy atom. Valine has a methyl group at χ_1 and another at χ_1+120°, and so the additional gauche interactions of the second methyl group (when χ_1 = -60° and 180°) were also listed. Threonine and isoleucine have methyl groups at χ_1-120°, and the gauche interactions with backbone N were also listed for the atoms at these positions. Interactions with the backbone carbonyl carbon were treated in a similar fashion. In this case, the relevant dihedral for locating the gauche interactions is defined here as "χ_1^C" or C-Cα-Cβ-Xγ, which is equal to χ_1-120° due to the stereochemistry of L-amino acids. When this dihedral was equal to +60° or -60° a gauche interaction was listed for the χ_1 rotamer. The analysis was extended to the dihedrals C-Cα-Cβ-Cγ_2 of valine (at χ_1), isoleucine (at χ_1+120°), and threonine (at χ_1+120°).

The δ heavy atoms of longer sidechains can form gauche interactions with Cα and syn-pentane interactions (hereafter referred to as "p" interactions in the Tables and in the text) with the backbone N and C atoms. These occur whenever the two connecting dihedrals fall in {+60°,-60°} or {-60°,+60°} combinations. In the case of the backbone nitrogen, the relevant dihedrals are usually χ_1 (N-Cα-Cβ-Cγ) and χ_2 (Cα-Cβ-Cγ-Xδ). The p interactions of the sidechain with the carbonyl carbon C were listed when the dihedrals χ_1^C and χ_2 combined to form {+60°,-60°} or {-60°,+60°} pairs. Some sidechains have more than one δ heavy atom, and the appropriate displacements from χ_2 must be made (viz., Leu Cδ_2 is located at χ_2+120°; aromatic, Asp, and Asn δ_2 atoms are located at χ_2+180°).

The analysis of backbone-dependent rotamers was performed by compiling a list of syn-pentane interactions of sidechain γ heavy atoms with backbone C_{i-1}, O_i, and N_{i+1} atoms for residue i. The positions of these three atoms are dependent on the backbone

dihedrals ϕ and ψ (see Figure 14). The dihedrals connecting C_{i-1} and Xγ are "ϕ^{β}" or C_{i-1}-N-Cα-Cβ and χ_1. ϕ^{β} is approximately equal to ϕ-120°. The dihedrals connecting backbone O and Cγ of the same residue are "$\psi^{O\beta}$" or O-C-Cα-Cβ (equal to ψ-60°) and χ_1^{C} (C-Cα-Cβ-Xγ). Between N_{i+1} and Xγ, the dihedrals are "$\psi^{N\beta}$" or N_{i+1}-C-Cα-Cβ (equal to ψ+120°) and χ_1^{C}. In the cases of Thr, Ile, and Val the interactions of these backbone atoms with the Cδ_2 atoms have also been tabulated. To locate the p interactions between C_{i-1}, O_i, and N_{i+1} and the γ heavy atoms, all values of ϕ and ψ for which ϕ^{β}, $\psi^{O\beta}$, and $\psi^{N\beta}$ were equal to +60° or -60° were considered. These were combined with the values of χ_1 and χ_1^{C} for the γ heavy atoms to locate the rotamers which make syn-pentane interactions with certain backbone conformations. In reality, these interactions take place at dihedral values distorted from the usual +60° and -60°, and occur across a 20-40° range, but we have listed them when the appropriate dihedrals are ±60°.

C. CHARMM calculation of backbone-independent rotamer preferences

The molecular simulation program CHARMM [52] was used to calculate relative energies of the rotamers of amino acid sidechains independent of backbone conformation. This was accomplished by using a truncated molecule consisting of backbone N, Cα, Hα, and C atoms and the complete sidechain of each amino acid. The fragment corresponds to that used in the conformational analysis of backbone-independent rotamers (Section b above; see Fig. 1) Nitrogen was given a charge q (N) = -0.16e, while q (Cα) = 0.07e, q (Hα) = 0.09e, and q (C) = 0.0e. In another series of calculations, the backbone was truncated further, including only N, Cα, and Hα1, with backbone C replaced with a hydrogen atom, Hα2. Other calculations used backbone C, Cα, Hα1, with backbone N replaced with a hydrogen atom, Hα3. These calculations were used to investigate interactions of the γ heavy atoms with the backbone N and C atoms individually. In general, the sidechain charges sum to zero for neutral sidechains as does the backbone.

The backbone N of N-CH_2-CH_2-CH_3 was given a negative charge of -0.16e, which is the sum of the NH1 and HN charges in the CHARMM 22 force field [53]. The Cα charge was -0.02e, which is the sum of the full backbone Cα charge of 0.07e and -0.09e to compensate the additional Hα2 charge of 0.09e. The backbone C of C-CH_2-CH_2-CH_3 had a charge of 0.00e, representing the sum of the carbonyl carbon and oxygen charges in the full force field of +0.51e and -0.51e respectively. In this case, Cα had a charge of -0.18e to compensate the 0.09e on each hydrogen. The effect is to collapse the NH and C=O dipoles in the full CHARMM force field for the protein backbone [53] in an untruncated backbone into a net negative point charge on N and zero charge on C. All other parameters, including the sidechain charges, were identical to the CHARMM22 parameter set [51]. Thus, in contrast to the simple steric analysis that treats gauche and syn-pentane interactions with the N and C atoms equivalently, the CHARMM calculations take into account their differing sizes and relative partial charges. This can lead to differences in interaction energies of sidechain atoms with the backbone N and C atoms.

To calculate rotamer preferences for backbone-independent conformations, minimizations were performed starting from χ_1 values near likely potential minima (viz. χ_1 = 60°, 180°, and -60°). For sidechains with more than a χ_2 degree of freedom, minimizations were performed from all likely combinations of χ_1 and χ_2. Sidechains with a single γ heavy atom and a single δ heavy atom (Met, Glu, Gln, Arg, and Lys) were truncated to a 2-amino pentanoic acid sidechain or -CβH_2-CγH_2-CδH_3). Minimizations were performed for 1000 steps of the CHARMM conjugate gradient minimizer with a dielectric constant of 1.0. No constraints were applied, and all atoms were allowed to move in the minimizations. We also calculated local minima and saddlepoint energies for butane and pentane. The saddlepoint conformations were calculated by constraining the C-C-C-C dihedrals to 120° or 0° with force constants of 10,000 kcal/mol, and then minimizing.

Predicted rotamer distributions were calculated with the equation $p_i = 100\ (\exp(-\Delta E_i/RT))/\sum_j (\exp(-\Delta E_j/RT))$. We have taken the temperature T to be 300K, although it has been argued recently that a much higher T should be used [54]. The entropic contribution to free energy due to fluctuations in the χ dihedrals about their averages is small, which justifies using the energies of the local minima in the equation for p_i. To demonstrate that this is true, we have calculated the probabilities of the ith rotamer of 2-amino butanoic acid using the equation $p_i = 100\ (\sum_k \exp(-\Delta E_{ik}/RT))/\sum_j \sum_k (\exp(-\Delta E_{jk}/RT))$, where the sums over k are for values of each χ angle in 5° intervals from χ_1=0° to 120° for g+ rotamers, 120° to 240° for t rotamers, and -120° to 0° for g- rotamers. The predicted populations were then compared with backbone-independent data from the PDB-based library described above.

D. CHARMM calculations of backbone-dependent rotamer preferences

Backbone-dependent rotamer preferences were calculated with unconstrained χ_1 (and χ_2) dihedrals starting from near the likely minima (60°,180°,-60°) by fixing the ϕ and ψ dihedrals of the N-acetyl N′-methylamide of each amino acid with force constants of 10,000 kcal/mol, and minimizing with 1000 steps of the conjugate gradient minimizer. In some cases, the minimized χ_1 values were up to 35° from the starting conformations. In others, no local minimum was found and the final χ_1 value was over 100° from the original value. If the minimized dihedral angle value was more than 60° away from the initial value, a force constant of 100 kcal/mol was applied to the χ dihedrals at their initial values, and the minimization was repeated. This ensures that in the tables of rotamer energies and calculated probabilities, the conformations in each column correspond to the correct rotamers. Probabilities for the rotamers at each ϕ,ψ point were calculated using the same Boltzmann weighting as the backbone-independent probabilities described above. These were then compared to the backbone-dependent rotamer library presented in tabular form (Table 4) with the Macintosh spreadsheet program Excel 3.0.

III. Results

A. Protein Databank Rotamer Preferences

(i) *Backbone-independent rotamer library*

The backbone-independent rotamer library calculated with the present database is given in Table 3. We also list the average values and standard deviations for the χ angles of each conformation. We note that the results are essentially the same as our earlier library, and very similar to the library of Ponder and Richards [7], who used only 10 proteins. The only differences in the most common rotamers preferred are for methionine and threonine. Ponder and Richards' library had only 16 methionines, while ours has 617. The most common conformation for Met is the {g-,t} rotamer (-60°,180°) in our database (36%), and {t,t} in the Ponder and Richards database (19% in our database). For Thr, in the Ponder and Richards database, there is a slight plurality of g+ conformations (47.9%) over g- conformations (45.0%). In our database, the order is reversed, with g- leading g+ by 46.0% to 45.6%. Our database has 2125 threonines compared to Ponder and Richards' 169 threonines. In some other cases, the percentages for χ_1 rotamers differs significantly from the Ponder and Richards rotamers even though the relative order does not. For example, the isoleucine distribution amongst the {g-,t}, {g-,g-}, {g+,t}, and {t,t} rotamers is 45%, 18%, 16%, and 13% in the older database, and 59%, 13%, 13%, and 7% in the database described here. The {g-,t} rotamer of leucine represented 64% of 147 leucines in the older database and only 54% of 2613 leucines in our database. Generally the average dihedral angles for conformations that were well represented in the older database are similar to the average dihedrals here. In cases with very few examples in the Ponder and Richards data, our averages are much closer to the standard rotamer values of 60°, -60°, and 180°. The rms deviations in Table 3 are somewhat larger in our library than the Ponder and Richards library, generally by 3-4°.

(ii) *Backbone-dependent rotamer library*

The backbone-dependent libraries for each sidechain type are presented in Tables 4a and 4b. ϕ values from -180° to -40° are listed across the top of each page of Table 4a. ψ values from -180° to 180° are listed along the side. Sidechains with more than one degree of freedom have been presented only with their χ_1 populations. Tables 4a cover the most populated regions of the map, where $\phi \leq -40°$, one page for each sidechain. The following six pages (Table 4b) cover the preferred rotamers for $\phi \geq -30°$, three sidechains to a page.

On the $\phi \leq -40°$ tables, the conformation that has the most representatives for the given amino acid and ϕ,ψ range in the database is highlighted by bold type. Also, when the largest number of χ_1 rotamers for a given ϕ,ψ occurs for $\chi_1=+60°$, the ϕ,ψ element in the table is colored light gray surrounded by thick borders. When the largest number of χ_1 rotamers occurs for $\chi_1=180°$ or $\chi_1=-60°$, the ϕ,ψ element is colored in dark gray or uncolored respectively. We have listed the gross numbers, rather than the percentages in each table to make clear the sparseness of some regions of the map. We have compiled the results for χ_2, but have not depicted these in the tables in the interests of saving space. In applications of the backbone-dependent rotamer library to protein modeling, these are used in conjunction with χ_1. For most non-aromatic sidechains, the majority of χ_2 rotamers are near 180° across the ϕ,ψ map. One exception is leucine, which has either $\{\chi_1,\chi_2\}=\{-60°,180°\}$ or $\{180°,60°\}$ as the predominant conformation in different regions of the ϕ,ψ map. Aromatic amino acids have $\chi_2=90°$ or -90° as their predominant χ_2 rotamers throughout the maps.

In the six pages of Table 4b, the values of $\phi/10$ from $\phi=-30°$ to 180° are listed across the top for each sidechain (i.e. $\phi/10=-3$ to 18), three sidechains to a page. Down the side of each page, the values of ψ from -180° to +180° are given. At each gridpoint, either 1, 2, or 3 is printed according to whether the g+,t, or g- rotamer is the most common rotamer for that block of ϕ,ψ space. The populations throughout this region are

very sparse, in most cases only 1 or 2 sidechains per block. When the populations are greater than 5, the rotamer designation (1, 2, or 3) is printed in bold type.

(iii) *Ability of the backbone-independent and backbone-dependent rotamer libraries to predict sidechain conformation*

In Table 5, we have compiled statistics on how well the backbone-independent rotamer library in Table 3 and the backbone-dependent rotamer library in Table 4 are able to predict protein sidechain conformation without any optimization [11]. We used the library to predict χ_1 of the sidechains of all of the proteins in Table 1. In each case the protein to be predicted was removed from the database so as to reduce bias in the results. Homologous proteins were not removed as each prediction was made, but we expect that the effect would be small in a database of this size (184 proteins). "Correct" conformations were taken to be those within 40° of the crystal structure minimum [55, 56]. The canonical values for the g+, t, and g- rotamers (60°, 180°, and -60°) were used as the basis for comparison (except for proline, where 40°±40° and -40°±40° were the ranges used). We have provided statistics in terms of the total number of sidechains of each type ("tot") as well as those within 40° of one of the three local minima ("sum"). For several sidechains (e.g. Leu, Glu, Gln, Arg, and Lys) 6-7% of the conformations in the database are outside of "correct" limits for any of the three possible χ_1 rotamers. Given the magnitude of the additional potential energy of these conformations (> 4 kcal/mol), some of these sidechains may also be improperly placed in the crystal structures, or may have special tertiary contacts [46, 57] or be involved in unusual contacts with non-protein ligands. We have broken down the results into the sidechains which are correctly predicted by the *most* populated rotamer ("r1") for the backbone dihedrals ϕ,ψ, the *next most* populated rotamer ("r2"), and the *least* populated rotamer ("r3") in Table 4 (note that sum = r1+r2+r3). In this way, we can observe the distributions of sidechains amongst the rotamer classes and which sidechain conformations are highly sensitive to the

backbone and which are not. In the last four columns of Table 5, we have compared the results of the backbone-dependent library ("bb-dep") of Table 4 with the backbone-independent rotamer library ("bb-ind") of Table 3. The columns labeled "bbind/tot" and "bbind/sum" give the prediction results of the backbone-independent rotamer library as a percentage of all the sidechains (bbind/tot) or only of those sidechains with χ_1 within 40° of a standard rotamer value (bbind/sum). The next to last column contains the ratio of the backbone-dependent prediction with the backbone-independent prediction for each sidechain type (bb-dep/bb-ind). The last column contains the difference in the fraction of correctly predicted residues ((bb-dep – bb-ind)/tot).

The sidechains can be grouped into a number of structural classes that correlate with the ability of the backbone-dependent rotamer library to predict their conformation. These are (1) serine and cysteine; (2) methionine, glutamic acid, glutamine, arginine, and lysine; (3) phenylalanine, histidine, tryptophan, histidine, leucine, aspartic acid, and asparagine; (4) threonine, valine, and isoleucine; and (5) proline.

The first group, serine and cysteine are small hydrogen bonding sidechains with a single γ heavy atom. In addition, cysteine is involved in many disulfide bonds with other sidechains, such that its conformation is determined by interactions with sidechains distant in sequence. These sidechains are not very well predicted by the backbone-independent or dependent libraries. Serine is the most poorly predicted by the backbone-independent library (42%) and next to the bottom by the backbone-dependent library (57%). Also only 85% of serines are in one of the top two most common rotamers in each ϕ,ψ cell of Table 4. One third of the remaining 15% are not within 40° of any of the three rotamers, and two thirds are in the least populated rotamer conformations ("r3"). The backbone conformation does aid in the prediction of serine conformations by raising the predictive ability from 42% to 57%. Cysteine by contrast has a less dramatic jump from 59% to 67%.

The second group includes longer sidechains with only a single Cγ and single Cδ (Sδ in Met). These sidechains are not as well predicted by either library as are the other sidechains. The backbone-independent library results in 53-59% prediction rates, while the backbone-dependent library results in improvements of only 3-7%. The backbone-dependent values are 64% (Met), 56% (Glu), 61% (Gln), 60% (Arg), and 60% (Lys). These sidechains have significant populations of rotamers outside the rotameric limits of 60°±40°, 180°±40°, and -60°±40°. Excluding the sidechains outside these limits, the predictions (i.e. r1/sum) are 3-5% better than those from r1/tot.

Group 3 includes sidechains with one γ and two δ heavy atoms, including the aromatics, leucine, aspartic acid, and asparagine. Even though these residues are quite different from one another in terms of charge, polarity, and size, they are quite similar in terms of the libraries' predictive abilities. All seven have r1/tot predictions in the narrow range of 67-74%, and r2/tot predictions between 19 and 26%. Their backbone-independent predictions are all between 48% (Asp) and 54% (His) with the exception of Leu (60%), and all are improved significantly by consideration of the backbone conformation. Phe, Tyr, and Asp are affected by the backbone the most, with improvements of 21%, 22%, and 24% over the backbone-independent predictions.

Threonine, isoleucine, and valine (Group 4) each have two γ heavy atoms, which greatly restricts the conformation of the sidechain because of steric interactions of the γ atoms with backbone atoms. These three residues have significantly higher backbone-dependent predictions than other sidechains at 80% (Thr), 81% (Val), and 85% (Ile). They also have r2 populations one third to one half those of the other sidechains. Threonine is improved by considering the backbone conformation more than any other amino acid type from a very poor 46% to 80%. Almost twice as many sidechains are correctly predicted by the backbone-dependent data than by the backbone-independent set. Isoleucine and valine by contrast have much less significant gains, although they are the first and second best predicted sidechains by the backbone-dependent library. The

reason for the different behaviors is that threonine has two conformations of equal weight in the backbone independent library of Table 3 (46% each) with a slight preference for χ_1=-60°. When a prediction is made with the backbone-independent library, the χ_1=-60° rotamer is chosen, which is correct only 46% of the time. By contrast, valine and isoleucine are not evenly distributed, but instead have one predominant χ_1 rotamer (70% χ_1=180° for Val and 74% χ_1=-60° for Ile), which produces a much higher backbone-independent library prediction.

Finally, proline is unique in structure amongst the amino acids, having a ring structure attached to the backbone at both ends at backbone N and Cα. Proline has a bimodal distribution of conformations centered around χ_1 = +25° and -25°. In Table 5, we have only considered these rotamers, rather than the three available to the other 17 sidechains. The improvement with the backbone-dependent library is striking, raising the prediction rate from 55% to 74%. The conformation of proline has been shown previously to be strongly dependent on ϕ [58]; this can also be seen from Table 4a. The backbone can be used to select the appropriate conformation from the two approximately equal populations in the database.

In the last line of the table, statistics are averaged across all of the sidechains in the database of Table 1. 69% of sidechain χ_1 values are predicted within 40° of their crystal structure conformation. Eliminating structures with "high energy" conformations (4% of the total), this value rises to 72%. 22% (r2/tot) and 23% (r2/sum) of sidechains are in the next most populated rotamer classes from the backbone-dependent library. Only 5% of the sidechains are in the least populated classes (i.e. r3/tot=r3/sum=5%). These values can be compared with the results from the backbone-independent rotamer library (Table 3). In that case, only 55% of all residues are correctly predicted within 40°, and 58% of residues eliminating the high energy structures. Overall, there is an improvement of 25% in the number of residues correctly predicted by the backbone-dependent library relative to the backbone-independent library (bb-dep/bb-ind), and an improvement of 14% in the

total number of residues ((bb-dep – bb-ind)/tot). When the database is used in conjunction with homology modeling and energy calculations [11], this improvement is very helpful in structure prediction.

B. Conformational analysis and CHARMM calculation of backbone-independent rotamers

In this section, we analyze the backbone-independent rotamer preferences for some model compounds and the 18 amino acids with sidechains common in proteins. First, we use CHARMM to calculate the energies and conformations of local minima of the molecules X-CH_2-CH_2-CH_3 and X-CH_2-CH_2-CH_2-CH_3 where X is a methyl group (X=CH_3), a backbone nitrogen atom (X=N), or a backbone carbonyl carbon atom (X=C). When X=CH_3 these molecules are identical to butane and pentane. These model compounds were chosen to investigate the energy of gauche and syn-pentane interactions of sidechain carbon atoms with backbone N and C atoms as well as with other sidechain methyl and methylene groups. Also we compare the CHARMM results on the X=CH_3 molecules with *ab initio* and experimental results on butane and pentane. Second, we analyze the rotamer preferences of the amino acids common in proteins as well as the sidechains -CH_2-CH_3 and -CH_2-CH_2-CH_3 of 2-amino butanoic acid and 2-amino pentanoic acid (3-letter abbreviations of Abu and Ape respectively). Abu with a single Cγ is compared with results for the polar Ser and Cys residues. Ape with a single Cγ and a single Cδ is used to model sidechains with more than two degrees of freedom (Glu, Gln, Met, Lys, Arg). For each sidechain we count the number of gauche and syn-pentane interactions in its various rotamer conformations and then use CHARMM to calculate the energy of these rotamers and the relative populations of the rotamers. These are compared to the experimental results in the backbone-independent rotamer library in Table 3.

(i) *Gauche interactions in proteins: Calculation of the rotamers of* $X\text{-}CH_2\text{-}CH_2\text{-}CH_3$ (*X =* CH_3*, backbone N, backbone C)*

The largest amount of experimental and theoretical work on heavy atom gauche interactions in organic molecules has been performed on butane. In Table 6, we summarize *ab initio* calculations performed by Wiberg and Murcko at the MP3/6-31g*//MP2/6-31g* level [30], calculations with the CHARMM22 alkane potential [59], and energies derived from infrared and Raman spectra [31, 32]. The CHARMM results fall approximately midway between the experimental values and the *ab initio* results. The gauche minima are 0.8 to 0.9 kcal/mol above the global energy minimum trans configuration. The C-C-C-C dihedral deviates from the canonical 60° (or -60°) values by 5.2° and 6.5°, because of the inequality of the C-C-C-C and C-C-C-H interactions, which favors larger C-C-C-C dihedrals to avoid the larger steric repulsion of the terminal carbon atoms relative to the hydrogen atoms. The anti+, anti-, and syn conformations are saddle point conformations between the gauche+, trans, and gauche- minima. The anti conformations with two C-C-C-H cis (0°) interactions and one H-C-C-H cis interaction have energies of approximately 3.5 kcal/mol in both the *ab initio* and CHARMM calculations. The syn conformation with one C-C-C-C cis interaction and four H-C-C-H cis interactions has an energy of 6.0 in *ab initio* and 5.3 kcal/mol in CHARMM. The experimental barrier is 4.6 kcal/mol, but it is subject to a large uncertainty [31, 32].

The γ heavy atoms of amino acid sidechains can have gauche+, gauche-, and trans minima with respect to both the backbone nitrogen and carbonyl carbon of the same residue. These are indicated in the Newman projection of a single residue peptide fragment in Figure 1. To investigate these interactions separately, we calculated the χ_1 energy minima for the pseudo-molecules $N\text{-}CH_2\text{-}CH_2\text{-}CH_3$ and $C\text{-}CH_2\text{-}CH_2\text{-}CH_3$, where N and C are the CHARMM backbone atom types NH1 and C. In Table 7, we divide the energies of the minima for butane and the two pseudo-molecules just described into angle,

dihedral, electrostatic, and van der Waals energy terms to investigate the origins of the gauche-trans energy differences in the CHARMM potential . The total energies are plotted as function of the dihedral in Figure 2. The gauche conformations for the pseudo-molecules N-CH_2-CH_2-CH_3 and C-CH_2-CH_2-CH_3 have energies of 0.67 kcal/mol and 0.51 kcal/mol, compared to 0.85 for butane. The primary energy differences are in the dihedral and electrostatic energy terms. Since butane has the larger X=CH_3 as its backbone atom, the dihedral X-C-C-C is further from the potential minimum than for the X=N and X=C molecules, which raises the dihedral energy. The torsional dihedrals involved are equivalent in the CHARMM22 potential [51]. When X = N, there is also a large positive electrostatic energy from the repulsion of the negatively charged carbon of the methyl group (-0.27e) and the backbone nitrogen. Similarly, when X = CH_3 there is also a repulsion between the terminal carbons. When X = C, with a zero charge, the electrostatic energy is 0.17 kcal/mol less than the X=CH_3 compound and 0.3 kcal/mol lower than the X=N compound. There are also differences in the angle and the van der Waals terms. When X = C, there is a larger angle term, since the X-CH_2-CH_2 angle equilibrium value is smaller when X = C (108°) than when X = CH_3 (111°), and X = N (113.5°). This means that the terminal heavy atoms of the X=C compound are pushed away from each other by steric repulsion only with the addition of angle strain. The angle strain is lower in the other two compounds with the larger X-CH_2-CH_2 equilibrium angles. The smaller angle also raises the van der Waals energy by bringing the terminal atoms closer together in the X=C molecule than in the X=CH_3 and X=N molecules. This is evident in the higher barriers in the X = C curves in Figure 2.

(ii) *Syn-pentane interactions in proteins: Calculation of the rotamers of X-CH_2-CH_2-CH_2-CH_3 (X = CH_3, backbone N, backbone C)*

In Table 8, the various minima and saddle point structures and energies of pentane, with two dihedral degrees of freedom, are listed. The single gauche ({t,g+} and

{t,g-}), anti ({t,a+} and {t,a-}), and syn ({t,s}) interactions in pentane have similar energies to those in butane. The energies of the {g+,a-}, {g+,a+}, {g+,g+}, {g-,a-}, {g-,a+}, and {g-,g-} are approximately the sums of the individual {t,x} conformer energies. But the {g+,g-}, {g+,s}, {g-,g+}, and {g-,s} conformations are all approximately 1.8 kcal/mol higher than the summed {t,x} components. In the final column of Table 8, the additional energy is designated "p". For example, {g+,g-} has an energy of 2g+p or 2 gauche interactions + the additional "pentane" interaction because of the closeness of the terminal carbons of pentane in this conformation. The {g±,s} conformers have an energy of g+s+p to indicate the gauche and syn components plus the additional energy due to the steric clash of the terminal pentane carbons. The larger dihedral values in the {g+,g-} and {g±,s}, up to 90° instead of 60°, should be noted. Wiberg and Murcko [30] have calculated the {t,t}, {t,g+}, {g+,g+} and {g+,g-} energies, and found somewhat lower energy differences between {g+,g+} and {g+,g-} with {t,t} and {t,g+}. The 2g energy (1.36) is not quite double the single g energy (0.86). The *ab initio* p component can be calculated as either E {g+,g-} - 2E {t,g+} or 1.61 kcal/mol, or as E {g+,g-} - E {g+,g+} or 1.97 kcal/mol. The p component to the {g+,g-} conformer energy is crucial to the relative energies of sidechain rotamers as will be described below.

In Table 9, the nine minimum energy conformations of X-CH_2-CH_2-CH_2-CH_3, where X = CH_3, N, or C are presented with their energies. The total energies are plotted as a function of χ_2 for the χ_1 = +60°, 180° and -60° rotamers in Figures 3a, 3b, and 3c respectively. In this case, the charge distribution of the X = N compound lowers the energy of the {t,g+} and {t,g-} rotamers (Figure 3b), apparently because of the interaction of the terminal methyl group with positively charged C$\alpha$$H_2$. For {g-,t} and {g+,t}, the electrostatic energy is highest in the N compound, since the γ methylene group is now gauche to the nitrogen. The higher angle and van der Waals energy also raise the X = C energy well above the X = N conformer. In each of the four gauche-gauche conformations, the X = C molecule has a higher van der Waals and angle energy, while the

bulky CH_3 group pushes the pentane rotamers to still higher dihedral values and higher dihedral energy. The small size of N and the large N-C-C angle allows the {g+,g-} and {g-,g+} conformations to adopt dihedral values closer to ±60°, and hence smaller dihedral energies. The X = N "p" interaction is only 1.4 kcal/mol above the {g+,g+} energy, while for X = CH_3 and C, the differences are 1.8 and 1.6 kcal/mol.

(iii) *2-Amino butanoic acid, Serine, and Cysteine*

We calculated the backbone-independent minimum energy rotamers for models of 2-amino butanoic acid (sidechain -CH_2-CH_3; also called 2-amino butyric acid, IUPAC abbr. *Abu* [60]), serine, and cysteine. These calculations were performed with a backbone consisting only of the atoms N, Cα, Hα, and C (with charges -0.16e, 0.07e, 0.09e, and 0.00e respectively), without amino hydrogens or the carbonyl oxygen present (Figure 1). In Table 10, we analyze the results for these residues in terms of the χ_1 and $\chi_1{}^C$ dihedrals N-Cα-Cβ-Xγ and C-Cα-Cβ-Xγ, where X = C, O, or S (columns 1 and 2). These dihedrals can be used to locate the gauche interactions present between backbone heavy atoms N and C and the terminal γ atom of each residue (columns 3 and 4). The energy "E" is the sum of these gauche interactions (column 5), and the net energy, "ΔE,", is the excess gauche interaction over the lowest "E" rotamer (column 6). The energies calculated with CHARMM are given next for the three residues, Abu, Ser, and Cys, as well as the calculated distributions, p_i, for each rotamer i. These are compared to the distributions found in proteins from the 184 chain protein database used in this paper. Finally, the χ_1 values for these sidechains in their CHARMM minima are listed in the last three columns of the table.

The Newman projection in Figure 1 and the first four columns in Table 10 show that when χ_1 = -60° or 180° there is one gauche interaction between a γ heavy atom and the backbone – with N and C respectively. When χ_1 = 60°, however, the γ atom has gauche interactions with both N and C. The net g interaction when χ_1 = 60° yields higher

CHARMM energies for Abu, Ser, and Cys. The largest differences in energy between the g+ and t conformations compared to the g- rotamer are in the angle and van der Waals energies of serine of approximately 0.35 and 0.7 kcal/mol respectively (not shown in Table 10). This is due apparently to the C/Oγ interaction, since the C-Cα-Cβ angle and the Cα-Cβ-Oγ angles are close to 109° with larger force constants than N-Cα-Cβ and Cα-Cβ-Cγ in Abu.

The calculated percentages from the CHARMM energies for Abu – 15%, 37%, and 49% agree approximately with χ_1 populations in the PDB for unbranched chains, such as methionine and arginine, which have g+, t, and g- populations of 10%, 30%, and 60% compared to 14%, 37%, and 49% for Abu. Of course, the PDB populations are averaged over backbone conformations available in proteins, and reflect the importance of other interactions in determining sidechain populations detailed in Section C below. But if we look at the backbone-independent populations from the PDB as *approximately* averaging out the effect of backbone conformation, we can get a rough idea whether the backbone has a smaller or larger effect on particular sidechain rotamer populations. For example, the percentages for cysteine, which does not participate in hydrogen bonding with its own backbone to an appreciable extent are in reasonable agreement with the whole PDB percentages. The calculated values for serine, however, are quite different from the PDB populations, because of hydrogen bonding to the backbone oxygen which significantly increases the population of $\chi_1 = 60°$ to a level well above $\chi_1 = -60°$.

Finally, we have calculated the percentages for Abu using a Boltzmann weighted sum over energies calculated every 5° from 0° to 120° for the g+ rotamer, from 120° to 240° for the t rotamer, and from -120° to 0° for the g- rotamer. The results were nearly identical to the calculations from the energy minima alone with a decrease of 0.3% and 0.4% in the g- and t populations and an increase of 0.7% in the g+ rotamer.

(iv) *Valine*

Valine has a methyl group at χ_1 and a methyl group at χ_1+120°. Therefore, for each rotamer position, four interactions need to be accounted for – backbone N and C with each γ carbon. The gauche interactions are listed in Table 11, and the result is a net single gauche interaction when χ_1 = 60° and -60°. The three conformations for valine are depicted as Newman projections in Figure 4. The CHARMM energies from the model compound are 0.6, 0 and 0.6 kcal/mol for g+, t, and g- conformations with calculated percentages of 21%, 58%, and 22%. The PDB percentages show a shift from the calculated values of 13% from g+ to t, because of backbone conformation effects detailed below which interfere with g+ rotamers.

(v) *Threonine*

Threonine (Table 12) has a hydroxyl group at χ_1 and a methyl group at χ_1 - 120° (see Figure 4), and so one has to keep track of interactions between backbone N and C and both Oγ and Cγ. As in valine, one conformation (in this case, g-) has an energy of 2g, while the other two conformations have energies of 3g. The net single gauche interaction gives the g+ and t rotamers CHARMM energies 0.6 and 0.5 kcal/mol above the g- conformation. The PDB results differ significantly from the calculated percentages, as in serine and valine, because of electrostatic and steric interactions with other backbone atoms, which raises the population of g+ significantly from t.

(vi) *2-Amino pentanoic acid, Methionine, Glutamic acid, Glutamine, Arginine, and Lysine*

In Table 13, an analysis of the interactions of single Cγ-single Cδ sidechains is shown with the CHARMM calculated energies and populations of 2-amino pentanoic acid (sidechain = $-CH_2CH_2CH_3$; also called norvaline and 2-amino valeric acid; IUPAC abbr. *Ape* or *Avl* [60]; in this paper, we will use *Ape*). The presence of δ heavy atoms means that in certain minimum energy conformations there will be {g+,g-} interactions of

the δ carbons with backbone N and C. These are shown in the Newman projections in Figure 5 and in the energy curves in Figure 6. In Figure 6, the energies are plotted as a function of χ_2 for the χ_1 = +60°, 180°, and -60° rotamers. The {g+,g-} interactions occur when χ_1 is 60° or -60° and χ_2 is -60° or 60° respectively. They also occur when the dihedral C-Cα-Cβ-Cδ (χ_1^C or χ_1 -120°) is 60° or -60° (when χ_1 = 180° or 60°) and χ_2 is -60° or 60° respectively. The CHARMM calculated energies and distributions for Ape demonstrate that the {60°,60°}, {60°,-60°}, {180°,-60°}, and {-60°,60°} are essentially forbidden conformations for single-γ, single-δ sidechains with energies between 2 and 3 kcal/mol higher than the remaining five conformations. It is worth noting the deviations in χ_1 and χ_2 from +60°, 180°, and -60° for these high energy conformations in their CHARMM calculated minima. The deviations are up to 19° in χ_1 and 24° in χ_2. The experimental averages for Met, Glu, Gln, Arg, and Lys for these conformations in Table 3 also show these large deviations. Schrauber et al. [57] have recently listed these strained conformations as distinct rotamers apart form the usual g+, t, and g- conformations. They are not in fact distinct rotamers, but simply have average χ angles 10-30° from the usual minima because of the strain due to the syn-pentane interactions.

Of the five low energy conformations, the {χ_1, χ_2} pairs {-60°,180°} and {180°,180°} have the lowest CHARMM energies and the highest representation in proteins for the amino acids Met, Glu, Gln, Arg, and Lys. For these five amino acids, the {-60°,180°} conformation is more common than the {180°,180°}, suggesting the influence of other backbone atoms on the choice of rotamer.

(vii) *Isoleucine*

Isoleucine has a CH_2CH_3 group at χ_1 and a methyl group at χ_1-120°. In Table 14, the dihedrals between N and C and both γ-carbons are listed. For different values of χ_2 (Cα-Cβ-Cγ_1-Cδ), the dihedral Cγ_2-Cβ-Cγ_1-Cδ equal to χ_2 +120° is also listed. When χ_2 is 180° or -60°, this dihedral adds another gauche interaction. Because Ile has only one

Cδ (located at {χ_1,χ_2}), it has the same "p" interactions as Ape, Met, etc. So the same four conformations {60°,60°}, {60°,-60°}, {180°,-60°}, and {-60°,60°} are disallowed. All nine rotamers are shown as Newman projections in Figure 7. The energies are plotted in Figure 8.

The {-60°,180°} conformation has the fewest gauche interactions and the lowest calculated energy. The other four conformations each have one excess gauche interaction over the {-60°,180°} rotamer, but energies of only 0.1 to 0.3 kcal/mol higher. The calculated percentages range from 15% to 24% among the five. But the PDB results are quite different with a shift away from both the χ_1 = +60° and 180° conformations to the χ_1 = -60° due to backbone-dependent interactions (see section C).

(viii) *Leucine*

Leucine has a single γ-carbon, a δ-carbon at χ_2 (Cδ1), and a δ-carbon at χ_2+120° (Cδ2) (note: CHARMM reverses $C\delta_1$ and $C\delta_2$ from the PDB definitions; the library and calculations presented here use the PDB definitions). This doubles the number of "p" interactions from 4 to 8: backbone N and C with leucine $C\delta_1$ and $C\delta_2$ times {g+,g-} and {g-,g+} combinations. As shown in Table 15 and the Newman projections for leucine in Figure 9, this occurs when {χ_1,χ_2} is equal to {60°,60°}, {60°,180°}, {60°,-60°} (twice), {180°,180°}, {180°,-60°}, {-60°,60°}, and {-60°,-60°}. These can also be seen in the energy plots in Figure 10. The two higher minima at χ_2 = 180° (χ_1 = +60° and -60°), the two higher minima at χ_2 = -60° (χ_1 = +60°, 180°), the two higher minima at χ_2 = +60° at 2 kcal/mol (1 p interaction each), and the highest local minimum at χ_2 = +60° (χ_1 = +60°) (2 p interactions) for a total of 8. The remaining two conformations {180°,60°} and {-60°,180°} are the only allowed conformations, and have approximately the same calculated energies and calculated percentages. The PDB search shows that these two conformations are much more common than the other seven with "p" interactions,

although the {-60°,180°} conformation is twice as prevalent (55%) as the {180°,60°} conformation.

(ix) *Phenylalanine, tyrosine, histidine, and tryptophan*

Aromatic amino acids (as well as aspartic acid and asparagine covered in the next subsection) each have a single γ carbon and two δ heavy atoms – Cδ1 and Cδ2 in Phe, Tyr, and Trp, and Nδ1 and Cδ2 in His – at χ_2 and χ_2+180° respectively. With an analysis of the g and p interactions, we can predict the likely values for χ_2 before looking at the energy plots and populations. The analysis is presented in Table 16.

As usual, χ_1 = +60° rotamers have one more g interaction than 180° or -60° rotamers, since Cγ has gauche interactions with both backbone N and C. Since it is known experimentally that the χ_2 minima do not occur at the usual +60°, 180°, or -60° values, we have tabulated not only the ±60° and 180° positions for χ_2 but other values of χ_2 as well in order to determine why the potential looks as it does. These include syn and anti conformations for either the δ1 or δ2 atoms. When χ_2 = 0° or 180° there are "p" interactions along with the "s" interactions when χ_1 is -60° or 60°, as was observed for pentane. These are probably of smaller magnitude than the "p" interactions caused by combinations of g+ and g- dihedrals or when the interacting groups are both methyl and/or methylene groups rather than CH groups in the aromatic rings.

For all three χ_1 rotamers, there is a syn interaction (χ=0°) between one of the δ atoms and Cα when χ_2 is 180° or 0°. In butane, a syn interaction was worth 5.3 kcal/mol in CHARMM. With the additional g and p interactions when $\{\chi_1,\chi_2\}$ = {±60°,0°} in pentane, the energy rises to 7.9 kcal/mol ("g+s+p" in Table 7). There is also is a g+a term in the energy when χ_2 is 60°, 120°, -120°, or -60°, which has an energy of 4.3 kcal/mol in pentane. The predicted χ_2 interactions are symmetric about χ_2 = 90° when χ_1 = +60°, but when χ_1 = 180° or -60° there are p interactions when χ_2 = 120° (between δ2 and backbone C) and χ_2 = 60° (between δ_1 and backbone N) respectively.

In Figures 11a and 11b, the potentials as a function of χ_2 are plotted for Phe and Tyr g+, t, and g- χ_1 rotamers. As predicted by the interactions listed in Table 16, the $\chi_1 = +60°$ potential is approximately symmetric about $\chi_2 = \pm 90°$. The 2g+2p+s interaction at $\chi_2 = 0°$ and 180° causes the high barrier between $\chi_2 = 90°$ and -90°. The minima at ±90° are exactly between the lower 3g+p+a energies. The difference in energy between the saddlepoint energy at χ_2=0° and the minimum near χ_2=90° is approximately what would be predicted by the values of g, s, p, and a for pentane in Table 7. That is, if g = 0.9, s = 5.5, p = 1.7, and a = 3.5 kcal/mol then the difference in energy is {2g+2p+s}-{3g+p+a} = p+s-a-g = 1.7+5.5-3.5-0.9 = 2.8 kcal/mol. The CHARMM energy difference for Phe itself in Figure 11a is 2.5 kcal/mol with the minimum at $\chi_2 = 95°/-85°$ for Phe and Tyr. It indicates that the p interaction when $\chi_2 = 0°$ and χ_1 or χ_1^C is ±60° is necessary to explain the energies of Phe and Tyr. The experimental average for χ_2 is listed in Table 3 for the $\chi_1 = +60°$ Phe and Tyr rotamers as 92.1°±12.5° and 88.5°±13.4°, both of which are in close agreement with the backbone-independent CHARMM calculations in Figure 11a and 11b and Table 16.

When χ_1 is 180°, the $\chi_2 = 60°$ and 120° conformations do not have an equal number of g, p, s, and a interactions with the backbone and the rest of the sidechain. The $\chi_2 = 60°$ structure is lower than $\chi_2 = 120°$ by one p interaction, and the CHARMM minimum for Phe occurs at $\chi_2 = 75°/-105°$ instead of at ±90°. Because the 2g+p+a and the g+p+s interactions at 120° and 180° are both of high energy and nearly equal in energy, the maximum energy falls roughly in between them at $\chi_2 = 150°$ and -30°. Similarly, when $\chi_1 = -60°$, the $\chi_2 = 60°$ and 120° conformations do not have an equal number of interactions with the backbone and the rest of the sidechain. In this case, the $\chi_2 = +60°$ has an additional p interaction, and so the CHARMM minimum falls higher than 90°. In this case the deviation from 90° is only 8° in CHARMM, while the $\chi_1 = 180°$ deviation from $\chi_2 = 90°$ was 15°. The reason is that when $\chi_1 = -60°$ the interaction occurs between a δ carbon and backbone N while in the $\chi_1 = 180°$ case, it

occurs between a δ carbon and the larger backbone C. The experimental distributions in Table 3 back up this claim: the χ_2 averages for the χ_1 = 180° and -60° rotamers are 77.6 ° and 98.6° respectively for Phe and 76.1° and 101.1° for Tyr respectively. Both the direction and the magnitudes of the deviations from 90° match the CHARMM calculations and the conformational analysis in Table 16. The CHARMM calculated percentages for Phe and Tyr are also listed in Table 16 and compared to the experimental distributions of Table 3. The CHARMM results are the same for Phe and Tyr, and the PDB results listed are the averages of the Phe and Tyr distributions from Table 3. They differ only by 1-2%. The CHARMM calculation underestimates the percentage of g+ rotamers (calc.=2% and PDB=13%) and overestimates the percentage of g- rotamers (calc.=69% and PDB=53%). This is probably because the CHARMM g+ energy of 2.2 kcal/mol relative to the g- rotamer is too high.

The analysis for histidine is complicated by the three different protonation states and the non-identical chemical identity of Nδ1 and Cδ2. The χ_2 potentials for the χ_1 = +60°,180°, and -60° rotamers are plotted for the three protonation states of histidine in Figures 11c (proton on Nδ1 or "HSD"), Figure 11d (proton on Nε2 or "HSE"), and 11e (protons on both Nδ1 and Nε2 or positively charged "HSP"). The potentials demonstrate that the δ positions are not equivalent. The relative energies of the minima can be explained by the electrostatic interactions between Nδ1 and the backbone N which is negatively charged. When Nδ1 is protonated (HSD, Figure 11c), the interaction between Nδ1/Hδ1 and backbone N is a stabilizing energy which lowers the energy minimum compared to the χ_2+180° conformation. These atoms approach each other when the dihedrals are either {+60°,-60°} or {-60°,+60°}, i.e. when a p interaction is listed in Table 16 between Nδ1 and backbone N. This occurs when χ_1 = +60° and χ_2 = -60° (and 0°) and when χ_1 = -60° and χ_2 is 60° (and 0°). In the χ_1 = +60° case, the χ_2 = -60° minimum is much lower in energy than the χ_2 = ~+80° minimum. When χ_1 = -60°, the χ_2 = +60° minimum is much lower than the χ_2 = -60° structure. The

remaining χ_1 rotamer, $\chi_1 = 180°$, has approximately equal energies for the $\chi_2 = +60°$ and -60° conformations.

When Nδ1 is not protonated (HSE, Figure 11d), it carries a net negative charge that raises the energy when Nδ1 approaches backbone N. In this case, the {+60°,-60°} conformation is higher in energy than the {+60°,+60°}, which is the opposite of the behavior of Hsd as expected. Similarly, the {-60°,+60°} energy is higher than the {-60°,+60°} energy. But for Hse, the $\chi_1 = 180°$ minima are not of equal energy, but the $\chi_2 = 60°$ is 2 kcal/mol lower than the $\chi_2 = -60°$. Finally, the potentials for doubly protonated histidine in Figure 11e (Hsp) show that the two minima follow the same pattern as Hsd, but the differences in energy are less extreme. This occurs probably because the favorable interaction of Nδ1 with backbone N is balanced by competing interactions of Nε2/Hε2 on the opposite side of the ring and backbone N.

The experimental dihedrals for χ_2 for histidine in Table 3 are quite similar to those of Phe and Tyr, showing that the analysis of the steric interactions for aromatic sidechains accurately predicts the histidine rotamers. The potentials indicate that the CHARMM minima are all close to $\chi_2 = \pm 60°$, while the experimental χ_2 dihedrals are 93.6°, 80.3°, and 101.3° for the χ_1 g+, t, and g- rotamers.

The potentials for tryptophan are shown in Figure 11f, and the CHARMM calculated energies, percentages, and χ_2 dihedrals of the minima are listed in Table 16. The potentials differ in form from the other aromatic amino acids because of the large indole ring. In this case, when χ_2 is 180°, the indole ring comes quite near to the Cα and Hα atoms and the energy is therefore much higher than the $\chi_2 = 0°$ conformation. For $\chi_1 = +60°$, the indole ring will also be quite close to the backbone N and C atoms (see Figure 5). Otherwise, the tryptophan potentials show features similar to the Phe and Tyr potentials: the $\chi_1 = +60°$ minima are near ±90° (91° and -80°; experimental averages 85.2° and -87.9°), while the $\chi_1 = 180°$ minima are at 81° and -103° (exp. 68.7° and -97.7°) and the $\chi_1 = -60°$ minima are shifted back to 92° and -88° (exp. 94.8° and -49.7°).

The -49.7° average is caused by a number of structures with χ_2 near 0°. In the calculated potential for χ_1 = -60°, there is a shallow minimum at χ_2 = 0° that is 0.7 kcal/mol above the global minimum.

(x) *Aspartic acid and asparagine*

A similar analysis for aspartic acid and asparagine is shown in Table 17 and Figures 12a and 12b. The oxygen atoms in aspartic acid are negatively charged and experience electrostatic repulsion when they are close in space to the backbone N. In asparagine, the oxygen atom Oδ_1 also has a destabilizing interaction with backbone N, while the amide Nδ_2-H2 has a stabilizing dipole/charge interaction with backbone N. In Table 17, we call these interactions +e and -e respectively, indicating destabilizing interactions with the oxygens (+e) and a stabilizing interaction with the amide (-e) with the value of e>0. The distribution of χ_2 values for each χ_1 is not clear from the averages in Table 3, so in Figure 13a and 13b we have plotted the experimental $\{\chi_1,\chi_2\}$ points from the PDB survey.

The interactions in Table 17 are the same as those listed in Table 16, except for the additional electrostatic interactions when backbone N and the δ_1 and δ_2 atoms exist in {+60°,-60°}, {-60°,+60°}, {±60°,0°} conformations. For aspartic acid, when χ_1 = +60° the only deep minimum occurs when χ_2 = -60° or 120° which because of the symmetry of the δ atoms are the same structure. The experimental distribution, however, shows a heavy distribution between χ_2 = -30° and +30° and near ±180°. When χ_1 = 180°, the close approach of the oxygens and backbone N does not occur, and its CHARMM potential energy minimum and maximum are closer together in energy. As predicted by the steric interactions, the minima occur when the interactions sum to 2g+a (-120° and +60°) between g+p+s (180° and 0°) and 2g+p+a (-60° and 120°) peaks. There are local minima in the CHARMM potential at all of these positions as shown in Table 17 and in Figure 12a, but most of these are quite shallow (except for -60° and 120°). The

experimental distribution in Figure 13a can be compared to the potential in Figure 12a. In the experimental distribution, there are not many structures with χ_2 in the high energy regions of the potential between -90° and -30° and between 90° and 150°.

The χ_1 = -60° potential is similar to the χ_1 = +60° potential with only two deep minima, in this case when χ_2 is -60° and 120° when the electrostatic interactions are weaker than in other conformations. This agrees with the experimental data in Figure 13a, where most of the conformations are between χ_2 = -90° and 0° and between 90° and 180°. But in these regions, there are more structures near the χ_2 = 0° and 180° limits than the χ_2 = ±90° ends, perhaps indicating the influence of other backbone atoms.

The potentials for asparagine in Figure 12b are quite different from aspartic acid. The differences occur because of the attractive nature of the Nδ2/backbone N interaction compared to the repulsive nature of the Oδ1/backbone N interaction. For Asn, the χ_1 = +60° potential has significantly lower energy near χ_2 = 180° than near χ_2 = 0°, which is also true of the experimental distribution. But the deep minimum is near χ_2 = 120°. In the experimental distribution, however, there are few structures near χ_2 = 120°. The χ_1 = 180° potential has a broad minimum from -20° to 120° and another between -140° to -60°. The experimental distribution is evenly distributed from χ_2 = -180° and +90°. There are relatively few structures in the high energy region between χ_2 = 120° and 180°. Finally, when χ_1 = -60°, the low energy minima are predicted to occur along with the negative electrostatic energies when χ_2 = ±180° and -120°. The minimum at χ_2 = -77° occurs between the electrostatic energy minimum at χ_2 = -120° and the steric minimum at χ_2 = -60°. This minimum is much lower in energy than the alternate minimum at χ_2 = 120°. In the experimental distribution, however, most of the structures lie between χ_2 = -90° and 0° and between 90° and 180° as in Asp. While a large portion of crystal structures may have the wrong χ_2 orientation (±180°), the distribution is shifted by 90° and not the 180° rotation expected about the Cγ-Cδ bond.

(xi) *Summary*

The analysis in this section has included enumerating the gauche and syn-pentane interactions between sidechain atoms and backbone atoms whose positions are independent of the backbone conformation dihedrals ϕ and ψ. Knowing the approximate energy values of these interactions from the high level *ab initio* calculations and experimental data on butane and pentane, we can estimate the relative order of the populations of the likely rotamers. Such back-of-the-envelope calculations do not consider electrostatic effects as well as subtle differences in van der Waals radii of the atoms involved in gauche and syn-pentane interactions. Nevertheless, they are useful, since a priori we do not know how well the CHARMM potential will represent all of the interactions involved in these model compounds.

Estimates of the backbone-independent rotamer distributions from both the sum of g/p interactions and the CHARMM calculations may not match the experimental distributions for two reasons. The first is that since the backbone conformation has a large effect on the rotamer preferences, the backbone-independent distributions are determined both by the relative energies of the three rotamers in each backbone conformation and by the populations of sidechains in different backbone conformations. The second is that the conformational analysis does not consider electrostatic effects and the CHARMM potential may not be entirely accurate in representing the relative energies of sidechain rotamers. When the predictions do not match the experimental backbone-independent rotamer library, we can not be sure whether it is the influence of the backbone or whether is just inaccurate modeling by the backbone-independent steric analysis and CHARMM calculations.

In general, the conformational analysis does a good job at picking out which conformations are unlikely. Whenever a p interaction is found for a given rotamer, both the CHARMM calculated probability and the experimental population are quite low. This is true for sidechains with alkane δ heavy atoms – Met, Glu, Gln, Met, Arg, Lys,

Ile, and Leu – whose rotamer populations with p interactions are only a few percent. Amongst the allowed rotamers for these sidechains, the conformational analysis is unable to predict the distributions accurately, since all g and p interactions are considered equivalent. For example, in the single γ/single δ sidechains in Table 13, the conformational analysis predicts a single gauche interaction for the {g+,t}, {t,g+}, and {g-,g-} conformers. However, since not all gauche interactions are equivalent because of the differing van der Waals radii and bond lengths and angles, the CHARMM energies for these three rotamers are 0.6, 0.2, and 0.4 kcal/mol. These small differences in energy result in a calculated distribution of 10%, 18%, and 15% respectively. The CHARMM results do not match the order of the experimental distribution either, where these three conformations account for approximately 7%, 9%, and 18% of Met, Glu, Gln, Arg, and Lys sidechains. There is a general shift toward g- χ_1 rotamers from t rotamers in the experimental data compared to the CHARMM results and the conformational analysis. As we will see in the next section, this has to do with unfavorable backbone-conformation dependent interactions of the t rotamers with backbone atoms not present in the model compounds used in this section.

While the steric analysis and CHARMM calculations do not work perfectly for the alkane sidechains, they do even more poorly with the short polar sidechains Ser, Thr, Asp, and Asn. The fact that simply counting g and p interactions does not succeed is not surprising, since electrostatic interactions were ignored entirely (except for Asp and Asn). CHARMM also does not do very well for these sidechains. In fact for Ser, the most preferred rotamer experimentally (43% g+) has the highest energy (1.8 kcal/mol) and lowest population (4%) according to CHARMM. This discrepancy is unlikely to be due to backbone conformation effects, and is more likely to be due to problems in the potential.

In sum, the analysis of backbone-independent rotamers is useful in determining the allowed and disallowed rotamers of the sidechains. It has also aided in explaining

systematic deviations of the χ_1 and χ_2 dihedrals from +60°, 180°, and -60° because of the influence of syn-pentane interactions on some sidechain conformations. It will be clear from the analysis of the backbone-dependent rotamers that the backbone conformation plays a strong role in determining which of the allowed rotamers is preferentially selected in different areas of the ϕ,ψ map, resulting in altered distributions from what one would expect from a backbone-independent analysis alone.

C. Conformational analysis of backbone-dependent rotamers

Rotamer populations are affected not only by backbone N_i and C_i of residue i whose positions are independent of ϕ and ψ, but also by the positions of other backbone atoms, especially C_{i-1}, O_i, and N_{i+1}, whose positions are dependent on ϕ and ψ. These three atoms are all connected to γ heavy atoms by two sp^3 hybridized atoms (Cβ and Cα) and one sp^2 hybridized atom (backbone N_i and C_i). There are therefore two dihedral degrees of freedom separating the γ heavy atoms with the backbone atoms C_{i-1}, O_i, and N_{i+1} (See Figure 14 for the Newman projection of the sidechain with the dipeptide backbone added. The relevant dihedrals are also listed in Figure 14). Hence, there is the possibility of "p" interactions between these atoms and sidechain γ atoms when the connecting dihedrals are in {g+,g-} or {g-,g+} combinations. The interactions are of smaller magnitude, since bond angles at the sp2 hybridized atoms are 120° instead of 109.5°. They are nevertheless of sufficient magnitude to alter the backbone-independent rotamer distributions calculated in the last section.

The position of C_{i-1} relative to Cβ is determined by the dihedral we call "ϕ^β" (C_{i-1}-N-Cα-Cβ), which is approximately equal to ϕ - 120°. When $\{\phi^\beta,\chi_1\}$ is equal to {-60°,60°} or {60°,-60°} there is a "p" interaction between C_{i-1} and Cγ (or Oγ or Sγ). In addition, in valine there is a p interaction when $\{\phi^\beta,\chi_1+120°\}$ is equal to {-60°,60°} or {60°,-60°}. In threonine and isoleucine, similar constraints occur on $\{\phi^\beta,\chi_1-120°\}$ for Cγ2. These interactions are summarized in Table 18, for the cases when χ_1 is equal to

60°, 180°, and -60°. In the last three columns, the p interactions are summed for the three cases – when there is a single γ heavy atom at χ_1, when there are γ heavy atoms at χ_1 and χ_1-120° (Thr and Ile), and when there are γ heavy atoms at χ_1 and χ_1+120° (Val).

The analysis in Table 18 shows that interactions between γ heavy atoms and C_{i-1} occur in conformations with ϕ near ±180° and +60°. Sidechains with a single γ heavy atom (Abu, Ape, Ser, Cys, Asn, Asp, Leu, Phe, Tyr, His, Trp, Met, Gln, Glu, Lys, and Arg) have a p interaction when ϕ = ±180° and χ_1 = -60° and when ϕ = +60° and χ_1 = +60°. Threonine and isoleucine with γ atoms at χ_1 and χ_1 -120° have p interactions when ϕ = 180° and χ_1 = 60° and -60°, leaving only χ_1 = 180° a likely conformation near ϕ = 180°. When ϕ = 60°, only χ_1 = -60° is allowed. Valine has only χ_1 = +60° allowed when ϕ is near 180°, and χ_1 = 180° allowed near ϕ = 60°.

The backbone dihedral ψ affects χ_1 rotamer preferences through p interactions between N_{i+1} and O_i and the γ heavy atom. The relevant dihedrals in these two cases are $\psi^{N\beta}$, $\psi^{O\beta}$, and χ_1C or N_{i+1}-C-Cα-Cβ, O-C-Cα-Cβ, and C-Cα-Cβ-Xγ (where X = C, O, or S). These dihedrals and their combined effects on rotamer interactions are listed in Table 19

Given the ϕ,ψ distribution of peptide units in proteins, ψ has a much greater role in determining rotamer frequencies than ϕ. Single γ heavy atom sidechains exhibit steric hindrances when {ψ,χ_1} = {±180°,180°}, {-60°,60°}, {0°,180°}, and {120°,60°}. Threonine and isoleucine are more restricted, and have only the following conformations *allowed* when ψ takes on the first value of each pair: {±180°,60°}, {-60°,-60°}, {0°,60°}, and {120°,-60°}. The corresponding allowed conformations for valine are {ψ,χ_1} = {±180°,-60°}, {-60°,180°}, {0°,-60°}, and {120°,180°}.

The steric analysis in Tables 18 and 19 provides a rationale for which χ_1 rotamers are *not* allowed in certain regions of the ϕ,ψ map. These are depicted schematically in Figure 15a for all sidechains with a single γ heavy atom, in Figure 15b for Val, and in Figure 15c for Ile and Thr. In these figures, the ϕ,ψ map is divided into 14 regions where

certain rotamers are "allowed" and others are "disallowed." That is, in certain regions, one or more rotamers may interact with backbone atoms N_{i+1}, O_i, and C_{i-1} in a syn-pentane configuration. The disallowed rotamers are indicated by arrows in the right margin and beneath the chart. The atoms involved in the p interaction are listed, and the arrows indicate the ϕ or ψ values of the interaction. Since these interactions tend to occur over a range of ϕ or ψ values (±20-40°), the regions are drawn as contiguous. The rotamers which are "disallowed" in each region, tend to occur with very low frequency in the backbone-dependent rotamer library of Table 4. The remaining "allowed" rotamers are listed in the boxes in each region of the map. The regions are labeled A1-A7 and B1-B7 (see Figure 15a,b,c for the ϕ,ψ limits of each region), and cover the same portion of the ϕ,ψ map as Table 4a.

The allowed rotamers for sidechains with a single γ heavy atom (except Pro) are shown schematically in Figure 15a. In Regions A1-A7, g- conformations are not allowed because of a p interaction between C_{i-1} and the γ heavy atom when ϕ is near -180°. When ψ is near 180° in Regions A1, B1, A7, and B7, t rotamers are not allowed because of a p interaction between N_{i+1} and the γ heavy atom. The result is that in Regions A1 and A7, only g+ rotamers are allowed. In Regions B1 and B7, both g+ and g- rotamers can occur. At ψ=0°, a p interaction occurs between O_i and the γ heavy atom of t rotamers, so that in Region A4 only g+ rotamers are allowed and in Region B4 both g+ and g- rotamers are allowed. When ψ=120° and -60°, there are p interactions between the γ heavy atom of g+ rotamers and O_i and N_{i+1} respectively. Regions A2 and B2 centered at ψ=120° and A5 and B5 centered at ψ=-60° are therefore expected to have few g+ rotamers. This means that in A2 and A5 only t rotamers are allowed, and in B2 and B5 both t and g- rotamers are allowed. Finally, in A3 and A6 only g- rotamers are disallowed because of the C_{i-1}/γ interaction at ϕ=-180°. In B3 and B6 all three rotamers are allowed.

The preferred and forbidden rotamers for ϕ>-40° have not been plotted in Figure 15a. The only additional interaction for single γ heavy atom sidechains is between the γ

heavy atom and C_{i-1} when ϕ=+60° and χ_1=+60°. This means that in the extensions of the B1, B3, B4, B6, and B7 regions to values of ϕ > -40°, there are unlikely to be g+ rotamers in the region about ϕ=60°.

The distribution of valine rotamers predicted from the conformational analysis of Tables 18 and 19 is shown in Figure 15b. Since valine has a methyl group at χ_1 and at χ_1+120°, its choice of sidechain conformation is more restricted by the values of ϕ and ψ than the single γ sidechains. In the A regions, both g- and t rotamers have p interactions between C_{i-1} and Cγ1 and Cγ2 respectively. Near ψ=180°, 0°, and -180° (Regions A1/B1, A4/B4, and A7/B7), both t and g+ rotamers have p interactions with backbone N_{i+1} or O_i. This leaves g- in regions B1, B4, and B7. But in A1, A4, and A7, all three rotamers have p interactions with the backbone. If valines exist in these regions, then it is likely that the rotmaer with the smallest p interaction with the backbone will be chosen (generally with O_i). In B2 and B5, only t rotamers are likely. In A2 and A5, again all three rotamers are disallowed. When ϕ is near +60°, both g+ and g- rotamers have p interactions with C_{i-1}. This would present a difficulty at values of ψ where t rotamers have p interactions with N_{i+1} and O_i (near ψ=±180° and 0°). Valine is not found in these regions.

The analysis for isoleucine and threonine resembles the analysis for valine just described. Since Ile and Thr have γ heavy atoms at χ_1 and χ_1-120°, their g+, t, and g- rotamers correspond to the g-, g+, and t rotamers of valine. Hence, the map in Figure 15c for Ile and Thr looks like Figure 15b for valine the three rotamers permuted.

The rotamer distributions as a function of ϕ and ψ in the backbone-dependent rotamer library in Table 4a can be compared to the predictions of Figure 15a, b, and c. With the additional interactions at ϕ=+60°, the preferred rotamers predicted from the enumeration of steric interactions can be compared with the limited experimental data for ϕ>-40° in Table 4b.

(i) *Serine*

The Ser distribution in Table 4a shows that in Regions A1 and B1 g+ rotamers predominate. In A1, there are very few g- rotamers as expected. There are t rotamers in both regions, but their numbers are quite small at ψ=180° and increase as ψ decreases from 180° to 150° in the region. In Regions A2 and B2, t rotamers are most common, with very few g+ rotamers. In A2 there are very few g- rotamers and there is a larger number in B2. Further down in Region B4, there are very few t rotamers as expected, and there are more g+ rotamers than g- rotamers. g+ has an extra gauche interaction with backbone N and C than does g-. The preference for g+ in Region B4 is likely to be due to favorable electrostatic interactions of the g+ rotamer with the carbonyl oxygen. In Region B5, g- rotamers predominate. There are also significant t and g+ rotamers in the α-helix region (the right hand half of B5). Again, the shift toward g+ rotamers is likely to be due to favorable electrostatic interactions missing from the simple analysis of Figure 15a.

In Table 4b, there are mostly g+ and t rotamers in the extension of B1 and B7 into this region. In the rest of the map there is a mix, with g- rotamers predominating. The populations are very low in all of these blocks, and so it is not clear that any statistically justified conclusions can be made for rotamer preferences in this region.

(ii) *Cysteine*

In general, Cys adheres more closely to the description in Figure 15a than Ser does, because of the weaker electrostatic interactions of the SH dipole with the backbone than the OH dipole of Ser. While the B1 rotamers of Ser were mostly g+, those in Cys are mostly g-. There are also relatively few t rotamers in Regions A1 and B1 in contrast to Ser. The B4 and B5 rotamers are mostly g-, with few t rotamers in B4 and few g+ rotamers in B5. t rotamers predominate along the bottom and right side of B5.

(iii) *Sidechains with a single γ and single δ heavy atom (Glu, Gln, Met, Lys, and Arg)*

Sidechains with a single γ and single δ heavy atom (Glu, Gln, Met, Lys, and Arg) all behave similarly to one another. For all five sidechains, g+ rotamers populate Region A1 as expected. Region B1 is populated mostly by g- rotamers with very few g+ and t rotamers. The population of t rotamers increases monotonically as ψ decreases from 180° toward 150° and below. The g+ rotamers are fewer in number than g- because of the extra gauche interaction Cγ has with the backbone when χ_1 is +60°. In Regions A2 and B2 t rotamers predominate with relatively few g+ rotamers. The reason for the preference for t rotamers over g- rotamers is not obvious, but is probably due to secondary structure interactions with backbone and sidechain atoms further along the chain. Region B4 for these sidechains is populated mostly with g- rotamers, because t rotamers are not allowed and g+ rotamers have the extra gauche interaction. The rotamers in Region B5 are mostly g- from ϕ=-140° to -90°. g+ rotamers are unlikely in this region because of the steric clash of Cγ and N_{i+1}. Thornton has also noted an unfavorable steric interaction between g+ rotamers and O_{i-3} in the α-helix region [61]. Above ϕ=-80°, there is a significant increase in t rotamers and in many instances the t rotamer populations are higher than the g- populations. The reasons for this are not entirely clear, but are likely to be due to interactions of the sidechain with atoms N-terminal to the residue in question because of the ϕ dependence. One possibility is a weak steric interaction between Cγ and the HN atom of the same residue which would occur when ϕ is between -60° and 0° (HN-N-Cα-Cβ = 0°-60°) and χ_1 is -60°. Also, ψ may have some effect. When ψ is near -60° or +120° and ϕ is between -60° and 0°, N_{i+1} would come very close to the γ atoms of g- and g+ rotamers. This will require further investigation.

The preferences in Table 4b are also worth noting for these sidechains. In the extension of Region B5 from ϕ=-30° to ϕ=0°, there are mostly t rotamers for these sidechains as in the Table 4a maps. Most of the rest of the occupied positions are

populated mostly by g- rotamers, except when ϕ>130°. When $\phi \geq$140°, t rotamers predominate in B2, B3, and B5, and g+ rotamers predominate in B1 and B7. This is to be expected, since the p interaction of the γ heavy atom with C_{i-1} occurs near ϕ=180° when χ_1 is near -60°. The regions where $\phi \geq$140° should be considered as extensions of Regions A1 to A7 with the same constraints on the rotamer preferences.

(iv) *Leucine*

The distribution of most favored rotamers in Table 4a for leucine resembles the single g/single d sidechains discussed in the previous section. The major difference occurs in the distribution of g+ rotamers in Regions B1-B7. Because the leucine Cd's have p interactions with backbone N and C when c1=+60° regardless of the value of c2, Leu has very few g+ rotamers anywhere in the f,y map above f=-150°. This means that in Regions B1 and B4, the rotamers are almost entirely g-. In Regions B2 and B5, there are both t and g- rotamers, with t rotamers predominating in B2 and g- rotamers predominating in B5, except when ϕ>-80° where t rotamers outnumber g- rotamers as in the single γ/single δ sidechains.

(v) *Aromatic sidechains (Phe, Tyr, His, and Trp)*

The distributions for aromatic sidechains differ from the single γ/single δ sidechains in the B2, B4, and B5 regions. In B2, the single γ/single δ sidechains were predominantly t rotamers at all values of ϕ. All four aromatic sidechains have higher g- populations than t populations in the region of B2 from ϕ=-130° to ϕ=-90° with t rotamers predominating on either side (ϕ<-130°; ϕ>-90°). The ϕ dependence suggests interactions of the δ atoms of the aromatic sidechains with atoms N-terminal to the Phe, Tyr, His, or Trp sidechain. In Region B4, there are more g+ rotamers than g- rotamers in some cells of Table 4a when ϕ>-80°. This shift occurs to a lesser extent in the single γ/single δ sidechains. Also, the shift to t rotamers in Region B5 when ϕ>-80° is larger in

the aromatic sidechains than in the single γ/single δ residues. Again, the ϕ dependence is notable.

(vi) *Aspartic acid and asparagine*

The data in Table 4a show that Asp and Asn behave similarly in most regions of the ϕ,ψ map. One exception is in A1 where Asp has more t rotamers than g+ rotamers when ϕ<-160°, while Asn has the usual preference for g+ rotamers. Asp is the only sidechain in which this occurs, violating the analysis in Figure 15a. In B1 for both sidechains, g+ rotamers predominate as in serine. Electrostatic interactions must favor this conformation over the g- rotamer. The distribution in A2 and B2 is mostly t rotamers for both sidechains, except when ϕ>-130° and ψ>130° for both sidechains where there is a strong preference for g- rotamers. The A3 and B3 regions also tend toward t rotamers, much more so than do the single γ/single δ sidechains which are mostly g- in these regions. In A4 and B4, g+ rotamers are more highly populated than g-, which is also in contrast to the results on single γ/single δ residues. Region B5 contains mostly g- rotamers and there are fewer t rotamers in the lower right region of B4 compared to other sidechain types. A7 and B7 are more highly populated than in other sidechains types, and these Asp and Asn sidechains are mostly g+.

The regions in Table 4b for Asp and Asn are more highly populated than they are for other sidechain types, especially in the α_R region near ϕ,ψ=60°,60°. The α_R region contains almost entirely g- rotamers for Asp and Asn. Again, as ϕ approaches +180°, there is a switch from g- rotamers to t and g+ rotamers because of the steric conflict between γ and C_{i-1}.

(vii) *Valine*

The map in Figure 15b predicts the distribution for Val in Table 4a fairly well. Regions B1, B4, and B7 contain almost entirely g- rotamers, while B2 and B5 contain

mostly t rotamers. There is only one valine in Region A3, A4, and A5, none in A6, and nine in A7 (at ϕ=-150° close to the B7 region). The Val sidechains in A1 and A2 occur mostly in the region of the border between A1 and A2 at ψ=150°. This is not surprising, since there p interactions between the Val γ carbons at ψ=180° and 120°. The sidechains in this border area are mostly g+ rotamers, because the g- and t rotamers would have an additional p interaction with C_{i-1} at ϕ=-180°.

(viii) *Isoleucine and Threonine*

The experimental data for Ile in Table 4a follows Figure 15c well. In this case, t rotamers predominate in the border area between A1 and A2 instead of g+ rotamers in valine. B1, B4, and B7 have mostly g+ rotamers and B2 and B5 mostly g- rotamers. Thr is similar to Ile in all regions except B5 where there are more g+ rotamers than g- when ϕ<-80°. Serine shows the same tendency toward g+ rotamers over g- rotamers in this region, because of electrostatic interactions with the backbone.

(ix) *Proline*

The experimental data for proline shows a strong dependence on ϕ. The reason for this has been discussed extensively in Chapter 1 of this thesis and by Cung et al. [58]. When ϕ is below -70°, the Cβ atom is lowered below the plane of the peptide bond (where below and above refers to the opposite and the same side of the peptide bond as the carbonyl carbon respectively). Ring strain forces the γ carbon above the plane of the ring, resulting in positive χ_1 values. The opposite situation occurs when ϕ>-60°. In this case, Cβ is above the plane of the ring, and Cγ is below. In this conformation, χ_1 is less than 0°. The effect is evident in Table 4a (PRO), where the g- rotamers (Cγ-exo) only predominate when ϕ>-60°. In addition, we expect that near ψ=120° and ψ=-60° there will be steric interactions between Cγ and O_i and N_{i+1} respectively when the proline residue is in a Cγ-endo conformations (χ_1>0°). The shift toward Cγ-exo as a function of

ϕ is more dramatic at these values of ψ. The magnitude of this effect is difficult to discern since the population of prolines at other values of ψ is low.

(x) *Summary*

The analysis in Figure 15 is useful for determining which rotamers are *not* likely for sidechains in certain areas of the ϕ,ψ map. For almost all sidechains and areas of the ϕ,ψ map, the analysis is successful in predicting rotamers which are unlikely to occur in the experimental data. Variations occur among the likely two or three rotamers that are not explained by such a simple analysis. In some cases, these are easily rationalized by arguments of electrostatic interactions or steric interactions of sidechain atoms with the rest of the protein. We have not attempted to analyze these variations in detail by locating the particular sidechain/protein interactions responsible. In most cases, the populations are 60/40 or 70/30 between the two allowed rotamers, so the differences in energy are small – a few tenths of a kcal/mol. Nevertheless, a number of factors responsible for the variations can be summarized. First, χ_1=g+ has one extra gauche interaction with backbone N and C (backbone conformation independent interaction), so that between g+ and g-, g- is favored by 0.8 kcal/mol or so. This is why when ψ is near 180°, 0°, and -180°, g- rotamers are preferred for most sidechain types when ϕ>-140°. This holds for single Cγ/Cδ sidechains (Arg, Lys, Met, Glu, Gln). Second, for short polar sidechains, such as Ser, Asp, and Asn, there are electrostatic interactions with the backbone carbonyl and NH dipoles of the same residue and the carbonyl of the preceding residue and the NH of the succeeding residue. For these three residues, g+ is preferred across the map at ψ=180°, 0°, and -180°. His also has some g+ preferences near ψ=0° and ϕ > -80° although the populations are quite close to the g- populations and the region is not well populated. There are also likely to be electrostatic interactions of the δ atoms of His with the backbone, which would depend on the χ_2 orientation.

The situation with the g-/t areas of the map is more complex. There regions are comprised of ϕ,ψ points with ψ values within 20° of ψ=120° (B2) and -60° (B5), where g+ rotamers are forbidden. The g- and t rotamers are approximately of equal energy in a backbone independent context. For single Cγ/single Cδ sidechains, there are tendencies toward t across the map from ϕ=-180° to ϕ=0° at both ψ=120° and -60°. There is a narrowing of the vertical width of the t regions near ϕ=-120°, ψ=120°. In Arg, there is actually a brief break, but the populations of g- and t are almost equal here. There are also breaks at ϕ=-70° or -80° and ψ=-60°.

Aromatics have a much larger break in the t regions near ϕ=-120° and ψ=120°, where g- rotamers are preferred. t rotamers are preferred on either side of the break (ϕ<-130° and ϕ>-90°). The same may also be true at ψ=-60°, but the populations are quite low in this region. This can be explained by the backbone independent analysis of aromatics (see Table 16): t rotamers have a larger energy than g- rotamers because the δ carbons have p interactions with the backbone C rather than backbone N as g- rotamers do. The energy difference in gauche energies with N and C is 0.5 kcal/mol, whereas for Abu it is only 0.2 kcal/mol.

Asp, Asn, Ser, and Leu have t rotamers straight across ψ=120° (B2) and -60° (B5). They also do not show a narrowing of the region at ψ=120°. Again for the polar sidechains (Asp, Asn, Ser) electrostatic interactions must disfavor g- rotamers in the middle (ϕ=-120° to -90°) with polar sidechain δ atoms. It is notable that His shows a less well defined break at ϕ=-120°, ψ=120°. The backbone independent calculations show a much lower energy for Asp t rotamers than g- rotamers. The reverse is true for Asn. Leu has a slightly lower t energy than g- energy in the backbone-independent conformations. These must involve interactions of δ atoms with the backbone N and C (conformation independent atoms). Finally, in some cases there will also be secondary and tertiary structure effects, which we have not attempted to analyze.

D, CHARMM analysis of backbone-dependent rotamers

We calculated the energies of the χ_1 rotamers of several sidechains to investigate how well the CHARMM potential describes the experimental distributions of the backbone-dependent rotamer library. This was done by minimizing the energy of the dipeptide of each residue type (N-acetyl-Xxx-N′-methylamide) with ϕ and ψ constrained to values in 10° increments. The minimizations were done for each χ_1 rotamer separately, and the populations at 300K were calculated according to a Boltzmann weighting. The calculations are described for each sidechain in the following sub-sections.

(i) *2-Amino butanoic acid*

The adiabatic surfaces of the three rotamers of 2-amino butanoic acid as a function of ϕ and ψ are shown in Figure 16a (χ_1 = 60°), 16b (χ_1 = 180°), and 16c (χ_1 = -60°). To produce these figures, we first calculated the lowest energy rotamer (among χ_1 = 60°,180°,-60°) for each value of ϕ and ψ at 10° intervals. We then subtracted the minimum energy from all three rotamer energies. The resulting data sets were then used to make the contour plots shown in Figure 16. The result is that when a particular rotamer is the lowest energy conformation of the three for particular values of ϕ and ψ, the energy of that rotamer is 0.0. When a rotamer is not the minimum energy, the difference in energy from the minimum energy rotamer is plotted. This allows us to remove the influence of the backbone/backbone interactions as well as interactions of the backbone with Cβ.

When χ_1 = 60°, we expect a p interaction when ϕ = 60° (Cγ and C_{i-1}) (Table 18) and when ψ = -60° (Cγ and N_{i+1}) and 120° (Cγ and O). These are shown clearly as the vertical ridge and the two horizontal ridges in Figure 16a. All three are somewhat displaced to ϕ = 90°, ψ = 100°, and ψ = -80°. These displacements are common for {g+,g-} conformations as was demonstrated by the *ab initio* and CHARMM calculations

in Tables 7 and 8 for pentane. The $C\gamma/C_{i-1}$ interaction is the largest, approximately 3.2 kcal/mol, while the $C\gamma/N_{i+1}$ and $C\gamma/O$ interactions are 1.6 and 1.2 kcal/mol respectively.

In Figure 16b, the interactions of $C\gamma$ and N_{i+1} when $\psi = \pm 180°$ and of $C\gamma$ and O when $\psi = 0°$ are demonstrated by the horizontal ridges in the center and the top and bottom of the contour plot. In this case, the ridges have a large peak near $\phi = 0°$, because the other two rotamers have strong interactions near $\phi = 180°$ and -180° ($\chi_1 = -60°$) and $\phi = 90°$ ($\chi_1 = 60°$). These are subtracted from the $\chi_1 = 180°$ rotamer plot near $\phi = \pm 180°$ and 90°, since the $\chi_1 = 180°$ rotamer is the lowest in energy in these regions. This leaves behind the peaks in the horizontal ridges at $\phi = 0°$. When ϕ is near 120°, the $\chi_1 = 180°$ rotamer is actually the lowest energy conformation, despite the p interaction between $C\gamma$ and O. Because oxygen has a significantly smaller van der Waals radius than carbon, the $C\gamma/O$ p energy is smaller than the $C\gamma/C_{i-1}$ energy. Figure 16c confirms the $C\gamma/C_{i-1}$ interaction when $\phi = \pm 180°$ and $\chi_1 = -60°$. Finally, it is worth noting the very weak horizontal ridges in the $\chi_1 = -60°$ map (Figure 16c) at ψ = -180°, -60°, 0°, 120°, and 180° that extent from ϕ = -180° to ϕ = -30°. These are exactly the locations of the ψ ridges in the $\chi_1 = 180°$ (ψ = -180°, 0°, and 180°) and $\chi_1 = +60°$ (ψ = -60° and 120°) maps.

It is clear from the three maps that the $\chi_1 = -60°$ conformation of Abu is the most favored in much of the heavily occupied portions of the Ramachandran map. $\chi_1 = 180°$ has steric interactions with backbone atoms in extended backbone conformations and near $\psi = 0°$, while $\chi_1 = 60°$ has steric conflicts in β-sheets as well as α-helices.

To understand the role of the backbone conformation on the rotamer energies and populations, it is useful to compare the results of CHARMM calculations on the backbone-independent rotamers of Abu (Table 9) and that of the backbone-dependent rotamers. In Table 20 (ABU) and Table 21 (ABU) are the energies and probabilities calculated with CHARMM for Abu as a function of ϕ (from -180° to -40°) and ψ (from -180° to 180°). (Note that in the energy table, the energies are in units of 0.1 kcal/mol. Divide each number by 10 to get the energy in kcal/mol.) The backbone-independent

calculation showed that the g+ rotamer had an energy 0.7 kcal/mol above the g- rotamer, and the t rotamer was 0.2 kcal/mol above the g- rotamer. This resulted in a population distribution of 49% g-, 37% t, and 14% g+. The backbone-dependent calculations, however, show that g- can be 4-5 kcal/mol above g+ when ϕ and ψ are near -180° and 180°. In other positions of the table, χ_1 = 180° conformations can be 1-2 kcal/mol above the other rotamers.

The populations in Table 21 (ABU) vary significantly from cell to cell, and the most prevalent rotamers alternate among the three conformations for each $\{\phi,\psi\}$ pair in a repeating pattern. The interactions responsible for these population shifts have been discussed in terms of the energy contour plots in Figures 16a-c. While there are many regions with one rotamer having a very large population (>80%), the transitions are continuous with intermediate $\{\phi,\psi\}$ values having approximately equal populations between two and sometimes three rotamers. One example of this phenomenon is in the α-helix region, near $\{\phi,\psi\}$ = -60°,-60°. In the range of $-80 \leq \phi \leq -40°$ and $\psi \geq -30°$, g+ is the most preferred rotamer. The differences in energy are a few tenths of a kcal/mol, and the population shift from 50/10/40 to 30/30/40 and to 10/50/40. In the next few subsections, we will discuss a number of amino acid sidechains found in proteins and how well the CHARMM potential reproduces their experimental distributions.

(ii) *Serine and cysteine*

The calculated energies for serine and cysteine are given in Table 20 (SER) and 20 (CYS), and the calculated probabilities are listed in Tables 21 (SER) and 21 (CYS). For these two amino acids (as well as for threonine), the hydroxyl and sulfhydryl hydrogens were placed in each of the three likely positions (dihedral Cα-Cβ-O/Sγ-Hγ = 60°,180°,-60°) and minimized for each χ_1 rotamer and each value of ϕ and ψ in the tables. The lowest energy with respect to the position of Hγ was used in the tables for each $\{\phi,\psi,\chi_1\}$ entry.

The smaller size of the OH and SH groups and the electrostatic interaction with the backbone alters the calculated population distributions compared to the Abu sidechain. The tendency toward g+ rotamers across the top of the table is absent in Ser and only occurs when $\phi \geq -110°$ in Cys. g- rotamers as the most preferable conformation are almost entirely absent from the Cys calculated map, and exist in the Ser calculated map only in two triangular regions {$-180° \leq \phi \leq -160°$, from $20° \leq \psi \leq 120°$ to $70° \leq \psi \leq 100°$} and {$-70° \leq \phi \leq -40°$, from $0° \leq \psi \leq 30°$ to $-50° \leq \psi \leq 140°$}.

The calculated distributions can be compared with the experimental distributions in Table 4a for Ser and Cys. The CHARMM potential does a poor job in representing these sidechains. The experimental preference in regions A1 and B1 is g+, while the calculated preference is t. The correct conformation is predicted in this region by the Abu maps, indicating that either the van der Waals interaction between $O\gamma$ and N_{i+1} is not large enough or the electrostatic interaction is too strong. In regions A2 and B2, the experimental preference is for t, and the calculated preferences are mostly for t, except at less negative ϕ ($\geq -120°$). In region B4, the databank is populated mostly by g+ rotamers, which corresponds to the calculated predictions. The calculated distributions, however, are much more skewed towards the g+ conformation than the experimental results, indicating again that perhaps the electrostatic attraction to the backbone is too strong. Finally, in region B5 there is a shift from g+ to g- rotamers in the database, but no such shift in the calculated conformations. Abu, on the other hand, does show this shift from g+ to g- and t (almost equally) in this region, indicating that the experimental change is caused by steric effects which are insufficient in the serine calculation or overwhelmed by electrostatic factors. In most of the map, the Abu calculations correspond relatively well to the experimental distributions. In region B4 the tendency toward g+ rotamers is much stronger in the serine experimental data than in the Abu calculated probabilities due to electrostatic interactions with the backbone, but the prediction of most favored rotamers is correct. Apparently the hydrogen bonds to other groups in the protein and solvent

mask the electrostatic interactions of the sidechain with its own backbone. The result is that serine's experimental distribution is similar to that of a non-polar sidechain, determined primarily by van der Waals interactions with backbone atoms.

In the CHARMM maps for cysteine, A1 and A2 contain predominantly t rotamers except at values of ϕ>-110° where the g+ populations are higher. The tendency toward g+ is very strong in regions B2-B5. In the experimental data, however, A1 is populated by g+ rotamers, A2 by g- rotamers, B2 by t rotamers, B4 and B5 mostly by g- rotamers and some t rotamers. The Abu map agrees with the experimental results in Regions A1 and B2. In the other regions where the most common rotamer does not correspond to the experimental cysteine results, the experimental preferences are nevertheless well populated in the calculated Abu map.

We repeated the minimizations for serine and cysteine with a dielectric constant of 78 instead of 1 as in the previous calculations (results not shown). The calculations with ε=78 showed that the component of the energy responsible for the unusual behavior was the electrostatic contribution. When ε was 78, the maps showed that g+ rotamers were preferred only in Region A1, the lower half of A3, A4, the lower half of A6, and A7. t rotamers were preferred in A2, the upper half of A3, A5, and the upper half of A6. g- rotamers were preferred throughout the B regions. Given g+ and t rotamers in some B regions of Table 4a, the appropriate value of ε probably somewhere in between 1 and 78. This indicates that interactions with other elements of the protein and the solvent (orientational component of the dielectric effect) have a strong determining effect on the conformation of serine and cysteine. It is difficult to determine whether the parameter set should be adjusted or whether the effects of the rest of the protein and the solvent are sufficient to explain the discrepancy between the ε=1 calculations and the experimental data in Table 4a.

(iii) *Valine*

The calculated map for valine in Table 21 (VAL) is very similar to the experimental distribution in Table 4a. The only difference is that the border between regions B4 and B5 is between ψ = -30° and -20° in the experimental map, but between -50° and -40° in the calculated map.

While the backbone-independent calculations (Table 10) produced energies of the g+ and g- rotamers of 0.6 kcal/mol above the t rotamer, the energies in Table 20 (VAL) for valine show generally much larger energy differences. The interactions with the backbone-conformation dependent atoms (C_{i-1}, N_{i+1}, and O) predominate in determining the conformation of valine rather than the interactions with backbone N and C. There would otherwise be equal populations of g+ and g- as shown in Table 10. However, in the database only 8% of valines are g+ while 22% are g-. The backbone conformation-dependent interactions are the cause of the incorrect prediction.

In the case of valine, the steric analysis of Tables 18 and 19 as well as the CHARMM calculations reproduce the experimental data very well. Without the complications of hydrogen bonds and other electrostatic interactions of serine and cysteine, the description in terms of gauche and syn-pentane interactions is apparently correct.

(iv) *Threonine*

Threonine has an experimental distribution similar to valine's, except that because of the definitions of χ_1, the g+, t, and g- conformations of valine are the t, g-, and g+ rotamers of threonine. The existence of electrostatic interactions with the backbone alters the experimental distribution of Thr only in Region B5, where there are more g+ rotamers than g- rotamers in part of the region. Otherwise, the map looks much like the Val experimental (Table 4a) and calculated (Table 21) maps after permutation of the rotamers.

The CHARMM calculated distributions, however, are quite different from the experimental preferences for Thr. The calculated map has t rotamers in A2 rather than g-

rotamers, t and g+ rotamers in B2, rather than g- rotamers, and g+ rotamers in all of B5, instead of a combination of g+ and g- rotamers as seen in the experimental distribution. The energies of the g- and t rotamers in regions B3, B4, and B5 are all 3-9 kcal/mol above the g+ rotamers. Local electrostatic interactions are apparently overestimated by the CHARMM potential.

To check this, we repeated the calculations with ε=78 as we did with serine and cysteine. The results were much more in accord with the experimental data with g+ rotamers in B1, B3, B4, B6, and B7. t rotamers were found to be preferred in A1, A3, A4, A6 and A7. g- rotamers were predominant in the remaining regions – A2, B2, A5, and B5.

(v) *Phenylalanine and tyrosine*

The calculated Phe and Tyr calculations do not predict g+ and g- rotamers in regions A1 and A2 respectively, instead having entirely t rotamers. In B2, the calculated rotamers are t up to ϕ=-90°, and then switch to g- when ϕ>-90°. The experimental distribution shows a tendency toward t rotamers when ϕ<-120°, g- rotamers when -120°≤ϕ≤-90°, and t rotamers again when ϕ≥-90°. It is clear that the p interaction between N_{i+1} and Cγ when ψ is near 180° is not large enough to prevent t rotamers from being populated all along the top of the calculated ϕ,ψ map. Similarly, the p interaction when ψ is -60° and χ_1 is +60° also does not seem to prevent the prevalence of g+ rotamers in region B4 and B5 in the calculated map. The experimental distribution has very few g+ rotamers in this region.

We repeated the phenylalanine map with ε=78, and found that the results were closer to the experimental distribution. In Regions A1, A4, and A7 the expected g+ rotamers were of lowest energy. The other A regions would be populated with t rotamers while the B regions were entirely g- rotamers. Again, the experimental case is in between the ε=1 and the ε=78 calculations. Whether adjustments in the potential should be made

will be determined by performing minimizations of phenylalanine residues in the full protein context to observe how well the potential performs when other interactions are present. Gelin and Karplus [46] using rigid rotations of sidechain χ angles found that the protein environment had a strong effect on the choice of rotamer for large sidechains such as Phe.

(vi) *Sidechains with single $\gamma$and single δatoms*

As for the backbone-independent calculations in Table 12, 2-Amino pentanoic acid (Ape) can be used as a model for the longer sidechains with single γ and single δ heavy atoms – glutamic acid, glutamine, methionine, lysine, and arginine. We have calculated the rotamer distributions as a function of ϕ and ψ of the dipeptide of 2-amino pentanoic acid by minimizing the energy of all nine possible χ_1,χ_2 rotamers. In Table 20 (APE) and 21 (APE) are the calculated energies and probabilities for Ape, which can be compared with the experimental distributions in Table 4a.

With relatively few exceptions in the experimental data (not shown), χ_2 in these sidechains is usually near 180°. This is also true of the Ape calculated probabilities, with the exception of rotamers near ψ = 120°, where χ_2 is split between 60° and 180°, with a tendency toward 60° conformers. The distribution of χ_1 in these sidechains follows a pattern very similar to Phe and Tyr sidechains. But in this case, the calculated probabilities follow the experimental variations quite well. In particular, the g+ rotamers in region A1 are present in the Ape calculations, and absent in the Phe calculations. Also, the switch from g- to t rotamers between regions B1 and B2 and also B4 and B5 are well reproduced by the calculations. The only incorrect predictions are when ϕ is between -130° and 110°. The calculated rotamers are g+ because of the steric interaction of Cγ and N_{i+1}. The experimental rotamers are primarily g-.

(vii) *Leucine*

Because of p interactions between its Cδ atoms and backbone N and C, the only allowed conformations for leucine are $\{\chi_1,\chi_2\}$ = {180°,60°} and {-60°,180°}. There are, however, some χ_1 = +60° conformations in the upper left corner of the ϕ,ψ map (region A1). This is shown by the experimental data in Table 4a.

As was the case for the other alkyl sidechains Val and Ape, the calculated probabilities for Leu are fairly close to the experimental preferences. All of the trends in the five regions are reproduced by the calculations shown in Table 21 (LEU). The approximate equality of the t and g- rotamers near ϕ,ψ = -70°,-50°, and the shift from a slight preference for g- to a preference for t as ψ decreases is shown in the energy calculations as well as the experimental distribution.

(viii) *Isoleucine*

According to the backbone-independent calculations in Table 13, isoleucine has 5 conformations without p interactions between sidechain atoms and backbone N and C. The experimental distribution in Table 3 is consistent with this prediction. The four remaining conformations exist in relatively few proteins. Isoleucine has a distribution of χ_1 angles fairly similar to valines, with the substitutions g+ $\rightarrow$t, t $\rightarrow$g-, and g- $\rightarrow$ g+. The calculated populations are quite similar to the experimental distributions (Table 21 (ILE)), indicating that the prediction of steric constraints on the positions of Cγ1 and Cγ2 by the backbone conformation as described above for valine and as listed in Table 18 and 19 are reasonable. It is clear from Table 19 that the main reason why there are so few t rotamers for isoleucine is that the χ_1 = 180° conformation results in p interactions of the Cγ atoms with the backbone atoms C_{i-1}, N_{i+1}, and O when ψ = -180°, -60°, 0°, 120°, and 180°. Given that these interactions occur in a certain range around these values for ψ, most of the map is covered.

The calculated map does show a preference for $\chi_2 = -60°$ conformations when ψ is near 120°, while the experimental preference is for $\chi_2 = 180°$. The calculated energies are within a few tenths of a kcal/mol.

(ix) *Proline*

The calculated rotamers for proline are shown in Table 21 (PRO). The effect of ϕ is well reproduced by the CHARMM calculations, showing that at more negative values of ϕ only the endo ($\chi_1>0°$) rotamer is populated. At less negative values, the exo ($\chi_1<0°$) rotamer predominates. The effect of ψ is also evident, although the effect is somewhat weaker than in the experimental data. When $\phi = -60°$, the rotamer population shifts from predominantly exo to nearly equal populations at $\psi = 120°$ instead of 140°. At $\phi = -60°$, the shift back to endo rotamers does not occur at lower values of ψ. It does occur when $\phi = -70°$, such that the shift from endo to exo occurs at $\psi = -20°$.

(x) *Summary*

In general, the CHARMM calculation of the backbone-dependent rotamers works well for alkyl sidechains such as Val, Ile, Leu, Pro, Met, Glu, Gln, Arg, and Lys. It does not work well for the short polar sidechains Ser, Cys, Thr, Asp, and Asn or the aromatic sidechains Phe, Tyr, Trp, and His. We are faced with the dilemma of whether the discrepancies are due to defects in the potential or secondary and tertiary structure effects missing from the simple dipeptide calculations. In the calculation of the backbone-independent rotamers, we were faced with a similar choice between errors in the potential and effects of the local backbone conformation. For some of the sidechains whose ϕ,ψ preferences were calculated with CHARMM, some of the discrepancies were quite large. For example, the fact that the serine and threonine calculated maps are quite similar would seem to indicate that CHARMM overestimates the electrostatic interactions with the backbone. Even the presence of the methyl group making syn-pentane interactions with the backbone was not enough to alter the CHARMM calculated maps to make Ser and

Thr look different. Both serine and threonine have very heavily skewed calculated distributions to certain rotamers due mostly to favorable electrostatic interactions. The test will be whether the CHARMM22 potential can predict the correct sidechain orientation of individual polar or aromatic sidechains when the full protein structure is included in the calculation. If the results are still quite poor, then the parameters will need to be adjusted. If necessary, the experimental data in the backbone-dependent rotamer library will be quite useful in scaling these and other interactions in the CHARMM potential to represent protein intramolecular interactions better.

IV. Discussion

The ability of the backbone-dependent rotamer library to predict 72% of sidechain conformations correctly in high resolution protein structures has a number of implications. The first is the practical use of the library for homology modeling [11], drug and ligand design, *de novo* protein design, and rational choices for site-directed mutagenesis studies. Knowing the probability that a sidechain will adopt a given orientation within a known (if even approximately) protein structure is a boon to experimentalists seeking to design peptides and proteins that will adopt predictable conformations. When substituting one sidechain for another in mutagenesis studies, it has been difficult to know what effect the new sidechain will have on neighboring sidechains if it should adopt a different conformation. The library makes it much easier to predict whether the new sidechain is *likely* to adopt a new conformation, and what that conformation would be. When designing drugs and other ligands for proteins, especially when a structure of a test ligand/protein complex is already available, the effect of altered sidechains on the drug conformation may be readily judged by whether the new sidechain is likely to adopt a new conformation or retain one similar to the original sidechain. In *de novo* protein design, the library affords a simple method for predicting packing

constraints that the backbone exerts on sidechains and may aid in choosing appropriate hydrophobic sidechains for filling a newly designed protein core.

One of the more surprising results of comparing CHARMM calculations and the experimental backbone-dependent distributions of rotamers is that the conformational analysis of the effect of the backbone on a single Cγ atom and CHARMM results for the single Cγ sidechain Abu do a better job than full CHARMM calculations of the dipeptides of such sidechains as Phe, Tyr, His, Trp, Ser, Thr, Cys, Asp, and Asn. The reasons for this are several, and their relative importance will require further analysis. The first and most obvious is that the CHARMM22 parameters have been derived from information on simple organic molecules with functional groups similar to the sidechain portions of each peptide unit. Charges and van der Waals parameters of polar sidechains in particular have been derived from *ab initio* studies on the interactions of these functional groups with water [51]. Whether non-bonded parameters as well as dihedral parameters are capable of accurately representing interactions of the sidechain with the backbone of the same peptide unit has not been studied in detail. The library and the conformational analysis provide information that can be utilized in further parameter development for molecular mechanics force fields.

The second cause of the discrepancy is that without the rest of the protein in the calculations, packing forces of particular secondary structures are absent. These forces may be stronger than the electrostatic interactions of the sidechain with the local backbone atoms. Sidechains such as Ser and Thr were found to behave more like the alkyl sidechains Abu and Val respectively, indicating that their polar nature and hydrogen bonding capabilities do not alter their distribution significantly from their non-polar analogues. This does not imply that individual Ser and Thr sidechain conformations are not determined by hydrogen bonds to other elements of the protein structure. Rather, these sidechains must still pay the price of steric conflicts with the local backbone fold (such as those of Abu and Val) in addition to satisfying hydrogen bonding requirements.

The distribution that results is a consequence of satisfying both the steric effects of the local backbone and the hydrogen bonding.

One of the remaining challenges is to determine what the additional forces on sidechains are that determine the conformations they will adopt. The library is able to predict 72% of χ_1's of sidechains that have structures within 40° of one of the canonical sp3-sp3 rotamers (+60°, 180°, -60°). From the steric analysis and the CHARMM calculations at ε=78, it seems that most of these are sterically the lowest energy minima (i.e., ignoring electrostatic interactions in polar sidechains) for the sidechain given the local backbone dihedral angles ϕ and ψ. Another 23% of sidechains adopt the next most likely local minimum energy structure (Table 5), and the final 5% the highest energy, least likely local minimum. 4% of all sidechains are not within 40° of one of the g+, g-, and t rotamer conformations. It is not known whether these are correctly placed in crystal structures or an error of interpretation of electron density. Given the high energy price paid (3-6 kcal/mol) and that proteins are only marginally stable ($\Delta G_{folding}$ = 5-20 kcal/mol [62]), it would seem a protein would not be able to afford many such sidechain conformations, at least not more than one or two. A challenge is presented to crystallographers to justify such structures with carefully analyzed electron density data and explanations of the interactions that make such unlikely conformations energetically feasible or necessary.

But for the 23% of sidechains in the second most likely χ_1 conformations, it is not known what provides the impetus for the choice of rotamer. The possibilities are limited to sidechains and backbone atoms distant in the sequence and sidechain and backbone atoms of the peptide units on either side of the given sidechain. The former involves packing of secondary structures against each other, and has been well studied [61]. If the latter provides at least part of the answer, then it may be possible to improve the predictions of the backbone-conformation dependent rotamer library for residue i by including ϕ_{i-1}, ψ_{i-1}, χj_{i-1}, ϕ_{i+1}, ψ_{i+1}, χj_{i+1}, and the chemical identities of residues i-1 and i+1. Two approaches can be used to study the effect of the neighboring residues on given

residue. The first is to calculate the steric clashes that occur in the database proteins when a sidechain in a higher energy rotamer is placed in its lowest energy rotamer (i.e., what would be predicted by Table 4). The second is to analyze the likely interactions that occur between atoms with three, four, and five dihedral degrees of freedom separating them. The analysis could be performed by first looking at simple alkyl chains such as hexane and heptane, and then continued by studying steric interactions that occur in tripeptides and tetrapeptides using CHARMM.

While most of the steric effects in backbone-independent and dependent sidechain conformations described in this paper have been studied previously in protein structures [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] and by energy calculations [3, 6, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46], we believe the present analysis is more thorough and consistent in its approach. In particular, the conformational analysis of gauche and syn-pentane interactions has provided a simple organizing principle for explaining and predicting the effects of backbone conformation on sidechain conformation that has not previously been used extensively on peptides. We hope the library will be useful in a number of applications as well as furthering our understanding of protein conformation, its determinants, and the process of protein folding.

References

1. R. Chandrasekaran and G. N. Ramachandran. Studies on the conformation of amino acids. XI. Analysis of the observed side group conformations in proteins. *Int. J. Protein Res.*, **2**, 223-233 (1970).

2. V. Cody, W. L. Duax, and H. Hauptman. Conformational analysis of aromatic amino acids by x-ray crystallography. *Int. J. Pept. Protein Res.*, **5**, 297-308 (1973).

3. J. Janin, S. Wodak, M. Levitt, and B. Maigret. Conformations of amino acid side-chains in proteins. *J. Mol. Biol.*, **125**, 357-386 (1978).

4. T. N. Bhat, V. Sasisekharan, and M. Vijayan. An analysis of side-chain conformation in proteins. *Int. J. Peptide Protein Res.*, **13**, 170-184 (1979).

5. M. N. G. James and A. R. Sielecki. Structure and refinement of penicillopepsin at 1.8 Å resolution. *J. Mol. Biol.*, **163**, 299-361 (1983).

6. E. Benedetti, G. Morelli, G. Nemethy, and H. A. Scheraga. Statistical and energetic analysis of side-chain conformations in oligopeptides. *Int. J. Peptide Protein Res.*, **22**, 1-15 (1983).

7. J. W. Ponder and F. M. Richards. Tertiary templates for proteins: Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.*, **193**, 775-792 (1987).

8. M. J. McGregor, S. A. Islam, and M. J. E. Sternberg. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.*, **198**, 295-310 (1987).

9. M. J. Sutcliffe, F. R. F. Hayes, and T. L. Blundell. Knowledge based modeling of homologous proteins, part II: rules for the conformations of substituted sidechains. *Protein Engineering*, **5**, 385-392 (1987).

10. P. Tuffery, C. Etchebest, S. Hazout, and R. Lavery. A new approach to the rapid determination of protein side chain conformations. *J. Biomol. Str. Dynam.*, **8**, 1267-1289 (1991).

11. R. L. Dunbrack Jr. and M. Karplus. Backbone-dependent rotamer library for proteins: Application to sidechain prediction. *J. Mol. Biol.*, **230**, 543-571 (1993).

12. T. A. Jones and S. Thirup. Using known substructures in protein model building and crystallography. *EMBO J.*, **5**, 819-822 (1986).

13. L. S. Reid and J. M. Thornton. Rebuilding Flavodoxin from Cα coordinates: A test study. *Proteins: Structure, Function, and Genetics*, **5**, 170-182 (1989).

14. J. Desmet, M. DeMaeyer, B. Hazes, and I. Lasters. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*, **356**, 539-542 (1992).

15. L. Holm and C. Sander. Database algorithm for generating protein backbone and side-chain coordinates from a Cα trace: Application to model building and detection of coordinate errors. *J. Mol. Biol.*, **218**, 183-194 (1991).

16. L. Holm and C. Sander. Fast and simple Monte Carlo algorithm for side chain optimization in proteins: Application to model building by homology. *Proteins*, **14**, 213-223 (1992).

17. E. L. Eliel. Conformational analysis – The last 25 years. *J. Chem. Ed.*, **52**, 762-767 (1975).

18. E. L. Eliel, N. A. Allinger, S. J. Angyal, and G. A. Morrison. *Conformational Analysis*. (New York: Interscience, 1965).

19. M. Hanack. *Conformation Theory*. Organic Chemistry. A. T. Blomquist and H. Wasserman ed. (New York: Academic Press, 1965).

20. R. M. Pitzer. The barrier to internal rotation in ethane. *Acc. Chem. Res.*, **16**, 201-210 (1983).

21. D. H. R. Barton. Interactions between non-bonded atoms and the structure of cis-decalin. *J. Chem. Soc.*, 340-342 (1948).

22. D. H. R. Barton. The conformation of the steriod nucleus. *Experientia*, **6**, 316-320 (1950).

23. D. H. R. Barton. The stereochemistry of cyclohexane derivatives. *J. Chem. Soc.*, 1027-1040 (1953).

24. H. Eyring. Steric hindrance and collision diameters. *J. Am. Chem. Soc.*, **54**, 3191-3203 (1932).

25. J. D. Kemp and K. S. Pitzer. The entropy of ethane and the third law of thermodynamics. Hindered rotation of methyl groups. *J. Am. Chem. Soc.*, **59**, 276-279 (1937).

26. K. S. Pitzer. Chemical equilibria, free energies, and heat contents for gaseous hydrocarbons. *Chem. Rev.*, **27**, 39-57 (1940).

27. K. S. Pitzer. The vibration frequencies and thermodynamic functions of long chain hydrocarbons. *J. Chem. Phys.*, **8**, 711-720 (1940).

28. C. A. J. Hoeve. Unperturbed mean-square end-to-end distance in polyethylene. *J. Chem. Phys.*, **35**, 1266-1267 (1961).

29. A. Abe, R. L. Jernigan, and P. J. Flory. Conformational energies of n-alkanes and the random configuration of higher homologs including polymethylene. *J. Am. Chem. Soc.*, **88**, 631-639 (1966).

30. K. B. Wiberg and M. A. Murcko. Rotational barriers. 2. Energies of alkane rotamers. An examination of gauche interactions. *J. Amer. Chem. Soc.*, **110**, 8029-8038 (1988).

31. D. A. C. Compton, S. Montero, and W. F. Murphy. Low-frequency Raman spectrum and assymetric potential function for internal rotation of gaseous *n*-butane. *J. Phys. Chem.*, **84**, 3587-3591 (1980).

32. J. R. Durig and D. A. C. Compton. Analysis of torsional spectra of molecules with two internal C_{3V} rotors. 12. Low frequency vibrational spectra, methyl torsional potential function, and internal rotation of *n*-butane. *J. Phys. Chem.*, **83**, 265-268 (1979).

33. F. Becker. Theoretische Behandlung des Einflusses sterischer Effekte auf die Reaktivitaet aliphatischer Verbindungen I. *Z. Naturforsch.*, **14a**, 547-556 (1959).

34. G. N. Ramachandran and V. Sasisekharan. Conformations of polypeptides and proteins. *Adv. Protein Chem.*, **23**, 283 (1968).

35. G. N. Ramachandran. Molecular forces in protein structure and crystallography. *Int. J. Protein Research*, **1**, 5-18 (1969).

36. P. K. Ponnuswamy and V. Sasisekharan. Studies on the conformation of amino acids. II. Potential energy of conformations of glycine, alanine, and aminobutyric acid. *Int. J. Prot. Res.*, **2**, 37-45 (1970).

37. P. K. Ponnuswamy, A. V. Lakshminarayanan, and V. Sasisekharan. Studies on the conformation of amino acids. XIII. Conformations of the arginine side group. *Biochim. Biophys. Acta*, **229**, 596-602 (1971).

38. P. K. Ponnuswamy and V. Sasisekharan. Studies on the conformation of amino acids. IV. Conformations of serine, threonine, cysteine, and valine. *Int. J. Prot. Res.*, **3**, 1-8 (1971).

39. P. K. Ponnuswamy and V. Sasisekharan. Studies on the conformation of amino acids. V. Conformations of amino acids with δ-atoms. *Int. J. Prot. Res.*, **3**, 9-18 (1971).

40. P. K. Ponnuswamy and V. Sasisekharan. Studies on the conformation of amino acids. IX. Conformations of butyl, seryl, threonyl, cysteinyl, and valyl residues in a dipeptide unit. *Biopolymers*, **10**, 565-582 (1971).

41. V. Sasisekharan and P. K. Ponnuswamy. Studies on the conformation of amino acids. X. Conformations of norvalyl, leucyl, aromatic side groups in a dipeptide unit. *Biopolymers*, **10**, 583-592 (1971).

42. P. N. Lewis, F. A. Momany, and H. A. Scheraga. Energy parameters in polypeptides. VI. Conformational energy analysis of the N-Acetyl N'-methyl amides of the twenty naturally occurring amino acids. *Israel. J. Chem.*, **11**, 121-152 (1973).

43. S. S. Zimmerman, M. S. Pottle, G. Nemethy, and H. A. Scheraga. Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP. *Macromolecules*, **10**, 1-9 (1977).

44. S. S. Zimmerman and H. A. Scheraga. Influence of local interactions on protein structure. II. Conformational energy studies of N-acetyl-N'-methylamides of Ala-X and X-Ala dipeptides. *Biopolymers*, **17**, 1849-1869 (1978).

45. S. S. Zimmerman and H. A. Scheraga. Influence of local interactions on protein structure. IV. Conformational energy studies of N-acetyl-N'-methylamides of Ser-X and X-Ser dipeptides. *Biopolymers*, **17**, 1885-1890 (1978).

46. B. R. Gelin and M. Karplus. Side-chain torsional potentials: Effect of dipeptide, protein, and solvent environment. *Biochemistry*, **18**, 1256-1268 (1979).

47. B. R. Gelin and M. Karplus. Sidechain torsional potentials and motion of amino acids in proteins: Bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA*, **72**, 2002-2006 (1975).

48. H. H.-L. Shih, J. Brady, and M. Karplus. Structures of proteins with single-site mutations: A minimum perturbation approach. *Proc. Natl. Acad. Sci. USA*, **82**, 1697-1700 (1985).

49. B. Tidor and M. Karplus. Simulation analysis of the stability mutant R96H of T4 lysozyme. *Biochemistry*, **30**, 3217-3228 (1991).

50. B. Pullman and A. Pullman. Molecular orbital calculations on the conformation of amino acid residues of proteins. *Adv. Protein Chem.*, **28**, 347-526 (1974).

51. A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, D. T. Nguyen, T. Ngo, B. Prodhom, B. Roux,

M. Schlenkrich, J. Smith, R. Stote, J. Straub, J. Wiorkiewicz-Kuczera, and M. Karplus. Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. *To be submitted*.

52. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, **4**, 187-217 (1983).

53. L. Kuchnir. Ph. D. dissertation, Harvard University (1993).

54. A. V. Finkelstein, A. Y. Badretdinov, and A. M. Gutin. Why do protein architectures have Boltzmann-like statistics? *To be submitted*.

55. N. L. Summers and M. Karplus. Construction of side-chains in homology modelling: Application to the C-terminal lobe of rhizopuspepsin. *J. Mol. Biol.*, **210**, 785-811 (1989).

56. J. J. Wendoloski and F. R. Salemme. PROBIT: A statistical approach to modeling proteins from partial coordinate data using substructure libraries. *J. Mol. Graphics*, **10**, 124-127 (1992).

57. H. Schrauber, F. Eisenhaber, and P. Argos. Rotamers: To be or not to be? An analysis of amino acid side-chain conformations in globular proteins. *J. Mol. Biol.*, **230**, 592-612 (1993).

58. M. T. Cung, B. Vitoux, and M. Marraud. Flexibility of Pro-Pro sequences: IR and NMR experiments. *New Jour. of Chem.*, **11**, 503-510 (1987).

59. A. D. MacKerell, S. Fischer, and M. Karplus. Derivation of empirical energy parameters for alkanes. *To be submitted*.

60. IUPAC-IUB Joint Committee on Biochemical Nomenclature. Nomenclature and symbolism for amino acids and peptides. Recommendations 1983. *Eur. J. Biochem.*, **138**, 9-37 (1984).

61. J. Thornton. Protein structures: The end of the protein folding pathway. In *Protein Folding*. T. E. Creighton ed. (New York: W. H. Freeman, 1993).

62. P. L. Privalov. Physical basis of the stability of the folded conformations of proteins. In *Protein Folding*. T. E. Creighton ed. (New York: W. H. Freeman, 1993).

Table 1

List of Brookhaven Protein Databank files used in backbone dependent rotamer library

Code-chain	Resol (Å)	Name
P1aaj	1.8	APOAMICYANIN
1aap-A	1.5	PROTEASE INHIBITOR DOMAIN OF ALZHEIMER'S AMYLOID BETA-PROTEIN PRECURSOR
P7aat-A	1.9	ASPARTATE AMINOTRANSFERASE: PYRIDOXAL-5'-PHOSPHATE-FORM (PH 7.5)
P1aba	1.45	GLUTAREDOXIN MUTANT (V15G, Y16P)
P1abg	1.7	SULFATE-BINDING PROTEIN WITH SULFATE
P1abh	1.7	PHOSPHATE-BINDING PROTEIN WITH PHOSPHATE
8abp	1.49	L-ARABINOSE-BINDING PROTEIN (M108L) COMPLEX WITH D-GALACTOSE
P7acn	2.0	ACONITASE COMPLEX WITH ISOCITRATE
2act	1.7	ACTINIDIN (SULFHYDRYL PROTEINASE)
1acx	2.0	ACTINOXANTHIN
P1ads	1.6	ALDOSE REDUCTASE WITH BOUND NADPH
1ak3-A	1.9	ADENYLATE KINASE ISOENZYME-3, PHOSPHOTRANSFERASE
1alc	1.7	ALPHA-LACTALBUMIN
1ald	2.0	ALDOLASE A
2alp	1.7	ALPHA-LYTIC PROTEASE
3app	1.8	ACID PROTEINASE (PENICILLOPEPSIN)
2apr	1.8	ACID PROTEINASE (RHIZOPUSPEPSIN)
2aza-A	1.8	AZURIN (OXIDIZED)
3b5c	1.5	CYTOCHROME B5 (OXIDIZED)
1bbp-A	2.0	BILIN BINDING PROTEIN
3bcl	1.9	BACTERIOCHLOROPHYLL-A PROTEIN
3blm	2.0	BETA-LACTAMASE
4bp2	1.6	PROPHOSPHOLIPASE A2 (PHOSPHATIDE-2-ACYL HYDROLASE)
3c2c	1.68	CYTOCHROME C2 (REDUCED)
1ca2	2.0	CARBONIC ANHYDRASE II (CARBONATE DEHYDRATASE)
1ccr	1.5	CYTOCHROME C
2ccy-A	1.67	CYTOCHROME C(PRIME)
2cdv	1.8	CYTOCHROME C3
2cga-A	1.8	CHYMOTRYPSINOGEN A
P3chy	1.7	CHE-Y
2ci2-I	2.0	CHYMOTRYPSIN INHIBITOR 2 (CI-2)
3cla	1.75	TYPE III CHLORAMPHENICOL ACETYLTRANSFERASE WITH CHLORAMPHENICOL
P1cll	1.7	CALMODULIN (VERTEBRATE)
P1clm	1.8	CALMODULIN FROM PARAMECIUM TETRAURELIA (WILD TYPE)
P1cmb-A	1.8	E. COLI MET APOREPRESSOR (METJ)
P2cmd	1.87	MALATE DEHYDROGENASE

3cms	2.0	CHYMOSIN B (FORMERLY KNOWN AS RENNIN) MUTANT (V111F)
2cna	2.0	CONCANAVALIN A
1cob-A	2.0	SUPEROXIDE DISMUTASE (CO SUBSTITUTED)
1cox	1.8	CHOLESTEROL OXIDASE
5cpa	1.54	CARBOXYPEPTIDASE ALPHA (COX)
1cpc-A	1.66	C-PHYCOCYANIN
1cpc-B	1.66	C-PHYCOCYANIN
2cpp	1.63	CYTOCHROME P450CAM (CAMPHOR MONOOXYGENASE) WITH BOUND CAMPHOR
4cpv	1.5	CALCIUM-BINDING PARVALBUMIN (PI=4.25)
1crn	1.5	CRAMBIN
1csc	1.7	CITRATE SYNTHASE - L-MALATE - CARBOXYMETHYL COENZYME A COMPLEX
1cse-E	1.2	SUBTILISIN CARLSBERG COMPLEX WITH EGLIN-C
1cse-I	1.2	SUBTILISIN CARLSBERG COMPLEX WITH EGLIN-C
1ctf	1.7	L7/L12 50 S RIBOSOMAL PROTEIN (C-TERMINAL DOMAIN)
2cyp	1.7	CYTOCHROME C PEROXIDASE (FERROCYTOCHROME C H2O2 REDUCTASE)
P1dfn-A	1.9	DEFENSIN
8dfr	1.7	DIHYDROFOLATE REDUCTASE
P3dni	2.0	DEOXYRIBONUCLEASE I (DNASE I)
P1dri	1.7	D-RIBOSE-BINDING PROTEIN
3ebx	1.4	ERABUTOXIN B
1ecd	1.4	HEMOGLOBIN (ERYTHROCRUORIN, DEOXY)
P1end	1.6	T4 ENDONUCLEASE V
4enl	1.9	ENOLASE (2-PHOSPHO-D-GLYCERATE HYDROLASE) (HOLO)
2er7-E	1.6	ENDOTHIA ASPARTIC PROTEINASE (ENDOTHIAPEPSIN) COMPLEX WITH H-261
3est	1.65	NATIVE ELASTASE
2fb4-H	1.9	IMMUNOGLOBULIN FAB
2fcr	1.8	FLAVODOXIN
4fd1	1.9	FERREDOXIN
3fgf	1.6	BASIC FIBROBLAST GROWTH FACTOR (HBFGF)
P1fia-A	2.0	FACTOR FOR INVERSION STIMULATION (FIS)
1fkf	1.7	FK506 BINDING PROTEIN COMPLEX WITH IMMUNOSUPPRESSANT FK506
P1fus	1.3	RIBONUCLEASE F1
P1fxd	1.7	FERREDOXIN II
2gbp	1.9	D-GALACTOSE/D-GLUCOSE BINDING PROTEIN (GGBP)
1gcr	1.6	GAMMA-II CRYSTALLIN
1gd1-O	1.8	HOLO-D-GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE
P1gky	2.0	GUANYLATE KINASE ATP:GMP-PHOSPHOTRANSFERASE
1gox	2.0	GLYCOLATE OXIDASE
1gp1-A	2.0	GLUTATHIONE PEROXIDASE
1gpb	1.9	GLYCOGEN PHOSPHORYLASE B (T STATE)
P1gpr	1.9	GLUCOSE PERMEASE (DOMAIN IIA)
3grs	1.54	GLUTATHIONE REDUCTASE, OXIDIZED FORM (E)
2had	1.9	HALOALKANE DEHALOGENASE (PH 6.2)

1hip	2.0	OXIDIZED HIGH POTENTIAL IRON PROTEIN (HIPIP)
2hmq-A	1.66	HEMERYTHRIN (MET)
1hne-E	1.84	HUMAN NEUTROPHIL ELASTASE COMPLEX WITH MSACK
1hoe	2.0	ALPHA-AMYLASE INHIBITOR HOE-467A
2hpr	2.0	HISTIDINE-CONTAINING PHOSPHOCARRIER PROTEIN HPR MUTANT (M51V,S83C)
5hvp-A	2.0	HIV-1 PROTEASE COMPLEX WITH ACETYL-PEPSTATIN (NY5 STRAIN)
4i1b	2.0	INTERLEUKIN-1BETA (IL-1BETA)
P4icb	1.6	BOVINE CALBINDIN D9K (MINOR A FORM)
1ifb	1.96	INTESTINAL FATTY ACID BINDING PROTEIN (APO FORM 1)
P1ifc	1.19	INTESTINAL FATTY ACID BINDING PROTEIN (APO FORM 2)
P3il8	2.0	INTERLEUKIN 8
9ins-A	1.7	INSULIN
9ins-B	1.7	INSULIN
1l58	1.65	LYSOZYME (MUTANT WITH PRO 143 REPLACED BY ALA)
6ldh	2.0	M4 APO-LACTATE DEHYDROGENASE
P1lec	2.0	FOURTH LECTIN ISOLATED FROM GRIFFONIA SIMPLICIFOLIA
1lh4	2.0	LEGHEMOGLOBIN (DEOXY)
2lhb	2.0	HEMOGLOBIN V (CYANO,MET)
P1lld-A	2.0	CHOH (D)-NAD (A) L-LACTATE DEHYDROGENASE T-STATE (C199S), WITH NADH
P1lmb-A	1.8	LAMBDA REPRESSOR-OPERATOR COMPLEX
P1lte	2.0	LECTIN COMPLEX WITH LACTOSE
2ltn-A	1.7	PEA LECTIN
2ltn-B	1.7	PEA LECTIN
P1lts-A	1.95	HEAT-LABILE ENTEROTOXIN (LT); CHOLERA-LIKE TOXIN; AB5 TOXIN
P1lts-C	1.95	HEAT-LABILE ENTEROTOXIN (LT); CHOLERA-LIKE TOXIN; AB5 TOXIN
P1lts-D	1.95	HEAT-LABILE ENTEROTOXIN (LT); CHOLERA-LIKE TOXIN; AB5 TOXIN
1mbd	1.4	MYOGLOBIN (DEOXY, PH 8.4)
P2mcm-A	1.5	MACROMOMYCIN
P1mdc	1.8	MANDUCA SEXTA FATTY ACID BINDING PROTEIN (MFB2)
P1mee-A	2.0	MESENTERICOPEPTIDASE WITH EGLIN-C PEPTIDYL PEPTIDE HYDROLASE
2mhr	1.7	MYOHEMERYTHRIN
2mlt-A	2.0	MELITTIN
P2msb-A	1.7	MANNOSE-BINDING PROTEIN A (LECTIN DOMAIN) WITH CA++ AND MAN6GLCNAC2ASN
P3mt2	2.0	METALLOTHIONEIN ISOFORM II
P1noa	1.5	NEOCARZINOSTATIN
P1npc	2.0	NEUTRAL PROTEASE
1omd	1.85	ONCOMODULIN
P1omp	1.8	D-MALTODEXTRIN-BINDING PROTEIN
1ova-A	1.95	OVALBUMIN (EGG ALBUMIN)
5p21	1.35	C-H-RAS P21 PROTEIN (1-166) WITH GPPNP
2pab-A	1.8	PREALBUMIN (HUMAN PLASMA)
P5pal	1.54	PARVALBUMIN (ALPHA LINEAGE)
9pap	1.65	PAPAIN CYS-25 OXIDIZED

1paz	1.55	PSEUDOAZURIN (OXIDIZED CU++ AT PH 6.8)
7pcy	1.8	PLASTOCYANIN
4pep	1.8	PEPSIN
1pgx	1.66	PROTEIN G TYPE 7 (B2 DOMAIN)
P1pii	2.0	N-(5'PHOSPORIBOSYL)ANTHRANILATE ISOMERASE INDOL-3-GLYCEROL-PHOS. SYNTHASE
P2por	1.8	PORIN
P1ppo	1.8	PROTEASE OMEGA (CYS-25 WITH BOUND MERCURY)
1ppt	1.37	AVIAN PANCREATIC POLYPEPTIDE
2prk	1.5	PROTEINASE K
3psg	1.65	PEPSINOGEN
5pti	1.0	TRYPSIN INHIBITOR (CRYSTAL FORM II)
4ptp	1.34	BETA TRYPSIN, DIISOPROPYLPHOSPHORYL INHIBITED
1r69	2.0	434 REPRESSOR (AMINO-TERMINAL DOMAIN) (R1-69)
P1rbc	2.0	RIBONUCLEASE-S MUTANT WITH MET 13 REPLACED BY ALA (M13A)
1rbp	2.0	RETINOL BINDING PROTEIN
1rdg	1.4	RUBREDOXIN
1rei-A	2.0	BENCE-JONES IMMUNOGLOBULIN REI VARIABLE PORTION
P6rlx-A	1.5	RELAXIN
P6rlx-B	1.5	RELAXIN
1rms	1.9	RIBONUCLEASE MS COMPLEXED WITH 3'-GUANYLIC ACID
P1rnb	1.9	BARNASE COMPLEXED WITH DEOXY-DINUCLEOTIDE INHIBITOR (D(GPC))
1rnh	2.0	SELENOMETHIONYL RIBONUCLEASE H
2rnt	1.8	LYS 25-RIBONUCLEASE T1 COMPLEX WITH GUANYLYL-2',5'-GUANOSINE
P1rop-A	1.7	ROP: COLE1 REPRESSOR OF PRIMER
3rp2-A	1.9	RAT MAST CELL PROTEASE II (RMCPII)
7rsa	1.26	RIBONUCLEASE A (PHOSPHATE-FREE)
2rsp-A	2.0	ROUS SARCOMA VIRUS PROTEASE (RSV PR)
5rub-A	1.7	RUBISCO (RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE)
1sar-A	1.8	RIBONUCLEASE SA
P2scp-A	2.0	SARCOPLASMIC CALCIUM BINDING PROTEIN
3sga-E	1.8	PROTEINASE COMPLEXED WITH TETRAPEPTIDE ACE-PRO-ALA-PRO-PHE-ALDEHYDE
3sgb-E	1.8	PROTEINASE B FROM STREPTOMYCES GRISEUS WITH TURKEY OVOMUCOID INHIBITOR
3sgb-I	1.8	PROTEINASE B FROM STREPTOMYCES GRISEUS WITH TURKEY OVOMUCOID INHIBITOR
1sgt	1.7	TRYPSIN (SGT)
P1sha-A	1.5	V-SRC TYROSINE KINASE TRANSFORMING PROTEIN (SH2) WITH PHOSPHOPEPTIDE A
1sn3	1.8	SCORPION NEUROTOXIN (VARIANT 3)
1snc	1.65	STAPHYLOCOCCAL NUCLEASE WITH CA++ AND 3',5'-DEOXYTHYMIDINE BISPHOSPHATE
2sod-B	2.0	CU,ZN SUPEROXIDE DISMUTASE
2tec-E	1.98	THERMITASE COMPLEX WITH EGLIN-C
P1ten	1.8	THE THIRD FIBRONECTIN TYPE III REPEAT OF HUMAN TENASCIN
1tgn	1.65	TRYPSINOGEN
1thb-A	1.5	HEMOGLOBIN (T STATE, PARTIALLY OXYGENATED)
1thb-B	1.5	HEMOGLOBIN (T STATE, PARTIALLY OXYGENATED)

P1thg	1.8	LIPASE TRIACYLGLYCEROL HYDROLASE
2tmn-E	1.6	THERMOLYSIN COMPLEX WITH N-PHOSPHORYL-L-LEUCINAMIDE
5tnc	2.0	TROPONIN-C
1ton	1.8	TONIN
P1trb	2.0	THIOREDOXIN REDUCTASE NADPH: OXIDIZED-THIOREDOXIN OXIDOREDUCTASE
P1tro-A	1.9	TRP REPRESSOR OPERATOR COMPLEX
2trx-A	1.68	THIOREDOXIN
2tsc-A	1.97	THYMIDYLATE SYNTHASE COMPLEX WITH D-UMP AND AN ANTI-FOLATE
1ubq	1.8	UBIQUITIN
1utg	1.34	UTEROGLOBIN (OXIDIZED)
9wga-A	1.8	WHEAT GERM AGGLUTININ (ISOLECTIN 2)
2wrp-R	1.65	TRP REPRESSOR (ORTHORHOMBIC FORM)
6xia	1.65	D-XYLOSE ISOMERASE (GLUCOSE ISOMERASE)
P1yea	1.9	ISO-2-CYTOCHROME C (REDUCED STATE)
P1yeb	1.95	B-2036 COMPOSITE CYTOCHROME C (REDUCED STATE)
1ypi-A	1.9	TRIOSE PHOSPHATE ISOMERASE (TIM)
P2zta-A	1.8	GCN4 LEUCINE ZIPPER
451c	1.6	CYTOCHROME C551 (REDUCED)
256b-A	1.4	CYTROCHROME B562 (OXIDIZED)

The code is the Protein Databank Code prefixed by "P" if the file is a preliminary entry, available by anonymous ftp from the Brookhaven National Labs (pdb.pdb.bnl.gov). The chain used from each file is appended to the code; if there is no chain indicated, then the single chain in the file is used.

Table 2

Limits for rotamer library χ angles

#	χ1 limits	χ2 limits
Ser, Thr, Cys, Val, Phe, His, Tyr		
1	60°±60°	
2	180°±60°	
3	-60°±60°	
Lys, Arg, Met, Gln, Glu, Ile, Leu		
1	60°±60°	60°±60°
2	60°±60°	180°±60°
3	60°±60°	-60°±60°
4	180°±60°	60°±60°
5	180°±60°	180°±60°
6	180°±60°	-60°±60°
7	-60°±60°	60°±60°
8	-60°±60°	180°±60°
9	-60°±60°	-60°±60°
Trp		
1	60°±60°	90°±90°
3	60°±60°	-90°±90°
4	180°±60°	90°±90°
6	180°±60°	-90°±90°
7	-60°±60°	90°±90°
9	-60°±60°	-90°±90°
Asp, Asn		
1	60°±60°	-60°±30°
2	60°±60°	0°±30°
3	60°±60°	30°±30°
4	180°±60°	-60°±30°
5	180°±60°	0°±30°
6	180°±60°	30°±30°
7	-60°±60°	-60°±30°

8	−60°±60°	0°±30°
9	−60°±60°	30°±30°
	Pro	
1	40°±40°	−40°±40°
3	−40°±40°	40°±40°

χ_1 and χ_2 ranges for each defined rotamer for the amino acid sidechains. The numbers in the left-hand column are used in Table 4 to illustrate the preferred rotamers in different positions on the ϕ–ψ map.

Table 3

Backbone-independent rotamer library

Residue	Total #	Rot. #	χ_1 limits	χ_2 limits	#	%	χ_1 Average	χ_1 RMS	χ_2 Average	χ_2 RMS
SER	2396	1	60°±60°		1037	43.3	64.7	15.6		
		2	180°±60°		589	24.6	-178.3	19.9		
		3	-60°±60°		770	32.1	-65.1	17.0		
CYS	594	1	60°±60°		79	13.3	60.8	13.3		
		2	180°±60°		159	26.8	-177.5	11.2		
		3	-60°±60°		356	59.9	-65.3	12.9		
MET	617	1-3	60°±60°		50	8.1	63.5	15.2		
		4-6	180°±60°		195	31.6	-171.0	18.0		
		7-9	-60°±60°		372	60.3	-68.0	13.0		
MET		1	60°±60°	60°±60°	6	1.0	68.5	19.3	85.5	11.9
		2	60°±60°	180°±60°	41	6.6	62.7	14.0	-176.2	16.3
		3	60°±60°	-60°±60°	3	0.5	65.7	18.7	-85.0	16.4
		4	180°±60°	60°±60°	63	10.2	-166.1	16.5	72.9	18.5
		5	180°±60°	180°±60°	117	19.0	-173.4	18.3	176.9	16.5
		6	180°±60°	-60°±60°	15	2.4	-173.0	18.4	-90.6	12.7
		7	-60°±60°	60°±60°	8	1.3	-89.1	13.3	74.0	21.1
		8	-60°±60°	180°±60°	219	35.5	-69.9	13.1	-177.7	15.6
		9	-60°±60°	-60°±60°	145	23.5	-63.9	10.7	-65.3	16.4
GLU	1833	1-3	60°±60°		194	10.6	58.0	22.1		
		4-6	180°±60°		604	33.0	-173.8	19.4		
		7-9	-60°±60°		1035	56.5	-67.6	17.0		
GLU		1	60°±60°	60°±60°	9	0.5	52.1	28.4	86.3	29.5
		2	60°±60°	180°±60°	134	7.3	60.3	21.3	-178.1	22.7
		3	60°±60°	-60°±60°	49	2.7	52.3	22.0	-80.0	16.7
		4	180°±60°	60°±60°	117	6.4	-169.0	20.1	68.1	18.7
		5	180°±60°	180°±60°	454	24.8	-175.6	18.3	179.1	19.1
		6	180°±60°	-60°±60°	33	1.8	-165.7	25.8	-80.3	25.6
		7	-60°±60°	60°±60°	130	7.1	-65.2	23.5	77.9	18.5
		8	-60°±60°	180°±60°	641	35.0	-67.7	14.9	178.9	17.9
		9	-60°±60°	-60°±60°	258	14.1	-68.3	17.9	-66.8	19.3
GLN	1204	1-3	60°±60°		99	8.2	61.5	21.2		
		4-6	180°±60°		408	33.9	-173.1	18.7		

		7-9	-60°±60°		697	57.9	-66.0	15.7		
GLN		1	60°±60°	60°±60°	9	0.7	48.5	26.5	90.2	15.5
		2	60°±60°	180°±60°	76	6.3	65.6	18.1	-179.1	17.4
		3	60°±60°	-60°±60°	13	1.1	48.6	23.8	-82.8	16.6
		4	180°±60°	60°±60°	121	10.0	-174.9	19.3	68.8	15.8
		5	180°±60°	180°±60°	259	21.5	-172.7	18.0	178.3	17.0
		6	180°±60°	-60°±60°	23	1.9	-170.1	22.1	-81.5	27.5
		7	-60°±60°	60°±60°	48	4.0	-74.5	18.4	82.4	18.0
		8	-60°±60°	180°±60°	460	38.2	-66.0	14.4	179.6	16.9
		9	-60°±60°	-60°±60°	189	15.7	-64.0	17.1	-67.4	19.1
ARG	1258	1-3	60°±60°		117	9.3	62.8	17.0		
		4-6	180°±60°		409	32.5	-173.7	18.5		
		7-9	-60°±60°		732	58.2	-67.0	15.7		
ARG		1	60°±60°	60°±60°	9	0.7	67.9	17.9	87.9	13.8
		2	60°±60°	180°±60°	104	8.3	62.5	16.9	177.3	21.4
		3	60°±60°	-60°±60°	3	0.2	53.7	15.4	-98.7	9.0
		4	180°±60°	60°±60°	72	5.7	-174.5	19.0	70.0	17.8
		5	180°±60°	180°±60°	319	25.4	-174.0	17.4	179.0	19.2
		6	180°±60°	-60°±60°	17	1.4	-164.4	30.7	-83.4	26.8
		7	-60°±60°	60°±60°	28	2.2	-81.3	26.3	72.0	33.0
		8	-60°±60°	180°±60°	565	44.9	-67.7	15.0	-177.5	19.3
		9	-60°±60°	-60°±60°	139	11.0	-61.2	13.2	-74.5	18.7
LYS	2045	1-3	60°±60°		167	8.2	61.3	21.5		
		4-6	180°±60°		698	34.1	-173.1	18.5		
		7-9	-60°±60°		1180	57.7	-68.4	17.3		
LYS		1	60°±60°	60°±60°	14	0.7	59.6	29.9	78.0	18.8
		2	60°±60°	180°±60°	142	6.9	62.4	19.4	-177.0	20.0
		3	60°±60°	-60°±60°	11	0.5	49.5	29.5	-80.6	23.6
		4	180°±60°	60°±60°	156	7.6	-173.7	20.0	73.9	18.6
		5	180°±60°	180°±60°	504	24.6	-174.1	17.2	177.6	21.2
		6	180°±60°	-60°±60°	37	1.8	-156.8	21.8	-81.0	25.3
		7	-60°±60°	60°±60°	74	3.6	-86.7	19.6	75.5	29.5
		8	-60°±60°	180°±60°	834	40.8	-68.6	16.3	-178.1	21.8
		9	-60°±60°	-60°±60°	262	12.8	-62.2	16.0	-71.0	21.2
PHE	1312	1	60°±60°		180	13.7	63.3	12.0	92.1	12.5
PHE		2	180°±60°		440	33.5	-177.8	11.1	77.6	17.3
PHE		3	-60°±60°		692	52.7	-66.4	11.7	98.6	29.6
TYR	1274	1	60°±60°		153	12.0	63.7	12.5	88.5	13.4

TYR		2	180°±60°		450	35.3	179.8	11.3	76.1	19.4
TYR		3	-60°±60°		671	52.7	-66.0	11.1	101.1	27.3
HIS	704	1	60°±60°		86	12.2	60.7	14.2	93.6	22.1
HIS		2	180°±60°		237	33.7	-174.4	13.1	80.3	34.3
HIS		3	-60°±60°		381	54.1	-65.4	12.6	101.3	33.2
TRP	440	1,3	60°±60°		69	15.7	61.0	10.5		
		4,6	180°±60°		144	32.7	-179.2	11.8		
		7,9	-60°±60°		227	51.6	-67.1	11.7		
TRP		1	60°±60°	90°±90°	22	5.0	57.1	10.8	85.2	7.7
		3	60°±60°	-90°±90°	47	10.7	62.9	9.8	-87.9	10.6
		4	180°±60°	90°±90°	85	19.3	-179.2	11.2	68.7	25.5
		6	180°±60°	-90°±90°	56	12.7	-179.4	12.3	-97.7	18.0
		7	-60°±60°	90°±90°	169	38.4	-67.1	11.1	94.8	26.2
		9	-60°±60°	-90°±90°	56	12.7	-65.9	12.4	-49.7	37.7
LEU	2613	1-3	60°±60°		46	1.8	62.1	17.8		
		4-6	180°±60°		866	33.1	-173.1	17.0		
		7-9	-60°±60°		1701	65.1	-71.0	15.9		
LEU		1	60°±60°	60°±60°	28	1.1	60.7	12.5	76.6	19.7
		2	60°±60°	180°±60°	16	0.6	63.9	24.8	165.1	25.4
		3	60°±60°	-60°±60°	2	0.1	66.9	11.2	-73.0	35.7
		4	180°±60°	60°±60°	706	27.0	-176.5	15.0	64.6	12.9
		5	180°±60°	180°±60°	139	5.3	-158.0	17.3	-178.1	31.2
		6	180°±60°	-60°±60°	21	0.8	-159.3	16.9	-75.4	28.2
		7	-60°±60°	60°±60°	234	9.0	-94.7	14.4	40.9	26.1
		8	-60°±60°	180°±60°	1404	53.7	-66.5	11.8	175.7	12.0
		9	-60°±60°	-60°±60°	62	2.4	-82.4	16.6	-44.1	26.7
ASP	2050	1-3	60°±60°		379	18.5	63.0	13.3	-3.7	39.2
		4-6	180°±60°		652	31.8	-170.8	15.0	8.3	47.6
		7-9	-60°±60°		1019	49.7	-69.6	13.2	-21.1	39.1
ASP		1	60°±60°	-30°±30°	90	4.4	64.8	16.8	-57.5	16.0
		2	60°±60°	0°±30°	230	11.2	63.6	9.2	1.5	14.2
		3	60°±60°	30°±30°	59	2.9	57.6	18.5	58.2	17.9
		4	180°±60°	-30°±30°	90	4.4	-163.7	19.9	-57.2	18.9
		5	180°±60°	0°±30°	372	18.1	-172.0	12.8	1.4	15.4
		6	180°±60°	30°±30°	188	9.2	-171.4	15.1	58.3	16.1
		7	-60°±60°	-30°±30°	358	17.5	-65.8	14.1	-51.7	15.5
		8	-60°±60°	0°±30°	584	28.5	-71.4	10.3	-12.5	12.1
		9	-60°±60°	30°±30°	75	3.7	-74.3	22.2	68.7	17.1

ASN	1561	1-3	60°±60°		265	17.0	64.5	12.1	3.4	41.1
		4-6	180°±60°		453	29.0	-169.7	15.5	9.3	51.3
		7-9	-60°±60°		843	54.0	-69.7	14.1	-30.3	46.4
ASN		1	60°±60°	-30°±30°	51	3.3	66.0	16.1	-54.9	15.8
		2	60°±60°	0°±30°	145	9.3	65.9	9.7	-1.3	17.1
		3	60°±60°	30°±30°	69	4.4	60.6	12.5	56.4	17.6
		4	180°±60°	-30°±30°	98	6.3	-168.3	17.2	-60.0	18.3
		5	180°±60°	0°±30°	177	11.3	-170.1	14.1	5.0	18.0
		6	180°±60°	30°±30°	177	11.3	-170.2	15.7	54.5	15.9
		7	-60°±60°	-30°±30°	479	30.7	-66.6	12.9	-56.7	15.9
		8	-60°±60°	0°±30°	278	17.8	-73.4	11.3	-11.7	13.9
		9	-60°±60°	30°±30°	84	5.4	-74.7	22.4	68.4	16.2
VAL	2439	1	60°±60°		223	9.1	62.8	26.2		
		2	180°±60°		1699	69.7	174.7	10.7		
		3	-60°±60°		517	21.2	-59.2	16.2		
THR	2125	1	60°±60°		968	45.6	62.4	12.1		
		2	180°±60°		179	8.4	-174.6	20.3		
		3	-60°±60°		978	46.0	-60.1	12.9		
ILE	1792	1-3	60°±60°		266	14.8	63.0	15.1		
		4-6	180°±60°		192	10.7	-171.4	21.1		
		7-9	-60°±60°		1334	74.4	-63.2	10.1		
ILE		1	60°±60°	60°±60°	23	1.3	62.9	34.8	93.4	19.5
		2	60°±60°	180°±60°	236	13.2	62.4	11.1	169.8	14.1
		3	60°±60°	-60°±60°	7	0.4	81.2	12.6	-63.4	26.8
		4	180°±60°	60°±60°	57	3.2	-170.5	26.8	73.0	17.8
		5	180°±60°	180°±60°	128	7.1	-172.4	17.0	168.4	15.0
		6	180°±60°	-60°±60°	6	0.3	-154.3	27.9	-91.6	14.0
		7	-60°±60°	60°±60°	51	2.8	-73.0	18.3	74.8	32.8
		8	-60°±60°	180°±60°	1048	58.5	-64.0	9.1	169.6	14.6
		9	-60°±60°	-60°±60°	234	13.1	-57.7	9.2	-60.9	14.9
PRO	1432	1	40°±40°		790	55.2	21.4	11.3	-25.5	15.3
		3	-40°±40°		642	44.8	-19.7	10.2	29.3	14.0

Table 5
Prediction of PDB sidechain rotamers by backbone-dependent rotamer library

Res.[a]	Tot	Sum	r1/tot	r1/sum	r2/tot	r2/sum	r3/tot	r3/sum	(r1+r2)/ tot	(r1+r2)/ sum	sum/tot	bb-ind/ tot	bb-ind/ sum	bb-dep/ bb-ind	(bbdep-bbind)/ tot
SER	2394	2276	0.58	0.61	0.27	0.28	0.10	0.11	0.85	0.89	0.95	0.42	0.44	1.38	0.16
CYS	593	587	0.67	0.67	0.26	0.26	0.06	0.06	0.93	0.94	0.99	0.59	0.60	1.13	0.08
MET	617	592	0.64	0.67	0.27	0.28	0.05	0.05	0.91	0.95	0.96	0.59	0.61	1.09	0.05
GLU	1833	1699	0.56	0.60	0.31	0.34	0.06	0.06	0.87	0.94	0.93	0.53	0.57	1.05	0.03
GLN	1204	1131	0.61	0.65	0.27	0.29	0.06	0.06	0.88	0.94	0.94	0.55	0.59	1.11	0.06
ARG	1258	1194	0.60	0.64	0.29	0.30	0.06	0.06	0.89	0.94	0.95	0.56	0.59	1.08	0.05
LYS	2045	1905	0.60	0.65	0.27	0.29	0.05	0.06	0.88	0.94	0.93	0.54	0.58	1.12	0.06
PHE	1311	1302	0.73	0.74	0.22	0.22	0.04	0.04	0.96	0.96	0.99	0.52	0.53	1.40	0.21
TYR	1274	1267	0.74	0.75	0.21	0.21	0.04	0.04	0.95	0.96	0.99	0.53	0.53	1.41	0.22
HIS	704	697	0.67	0.68	0.25	0.26	0.06	0.06	0.93	0.94	0.99	0.54	0.54	1.25	0.14
TRP	440	435	0.67	0.68	0.26	0.26	0.06	0.06	0.93	0.94	0.99	0.51	0.52	1.30	0.15
LEU	2612	2428	0.67	0.72	0.25	0.27	0.01	0.01	0.92	0.99	0.93	0.60	0.65	1.10	0.06
ASP	2050	1971	0.72	0.75	0.19	0.19	0.05	0.06	0.91	0.94	0.96	0.48	0.50	1.51	0.24
ASN	1561	1506	0.67	0.69	0.25	0.26	0.05	0.05	0.92	0.95	0.96	0.52	0.54	1.28	0.15
VAL	2439	2341	0.81	0.84	0.10	0.10	0.05	0.06	0.91	0.94	0.96	0.69	0.72	1.18	0.12
THR	2125	2068	0.80	0.82	0.12	0.12	0.05	0.05	0.92	0.95	0.97	0.45	0.46	1.79	0.35
ILE	1791	1751	0.85	0.87	0.09	0.09	0.04	0.04	0.94	0.96	0.98	0.74	0.76	1.15	0.11
PRO	1432	1432	0.74	0.74	0.26	0.26	0.00	0.00	1.00	1.00	1.00	0.55	0.55	1.33	0.18
All	27683	26582	0.69	0.72	0.22	0.23	0.05	0.05	0.91	0.95	0.96	0.55	0.57	1.25	0.14

[a] Definitions used in table headings:

Res.=residue type

Tot.=# of residues of given residue type in proteins database (Table 1)

Sum=# of residues out of tot which have χ_1 within 40° of g+,t, and g- rotamer definitions, i.e. 60°, 180°, -60°. The remaining residues have χ_1 values of 120°±20°, -120°±20°, and 0°±20°. Rotamer definitions for Pro used were χ_1=40°±40° and -40°±40°.

r1=# of residues of given residue type which are within 40° of the most probable χ_1 rotamer in the backbone-dependent rotamer database (Table 4). For each residue, χ_1 is predicted by using Table 4 and the values of ϕ and ψ for that residue. In each case, one sidechain is removed from the bin before the prediction is made to prevent biasing the prediction with the predicted sidechain left in the database.

r2=# of residues of given residue type which are within 40° of the *next* most probable rotamer (after r1) in the backbone-dependent rotamer database.
r3=# of residues of given residue type which are within 40° of the *least* probable rotamer in the backbone-dependent database.
bb-ind=# of residues of given residue type which are within 40° of the backbone-independent library prediction (Table 3).
bb-dep=r1

Table 6

Butane ab initio, CHARMM, and experimental energies

	Ab initio[a]		CHARMM		Experimental[b]
Conformer	χ	ΔE^c	χ	ΔE	ΔE
trans (t)	180.0	0.00	180.0	0.00	0.00
gauche+ (g+)	65.2	0.81	66.5	0.85	0.89±0.03
gauche- (g-)	-65.2	0.81	-66.5	0.85	0.89±0.03
anti+ (a+)	121.6	3.54	120.0	3.48	
anti- (a-)	-121.6	3.54	-120.0	3.48	
syn (s)	0.0	5.97	0.0	5.25	4.6

[a] MP3/6-31G*//MP2/6-31g* (Ref. 30)

[b] From far IR and Raman data (Ref. 31, 32)

[c] Energies in kcal/mol; angles in degrees

Table 7

Energies of X-CH_2-CH_2-CH_3

Conf	χ[a]			Angle energy			Dihedral energy			Electrostatic energy			Van der Waals energy			Total energy		
X:[b]	CH_3	N	C	CH_3	N	C	CH_3	N	C	CH_3	N	C	CH_3	N	C	CH_3	N	C
t	180	180	180	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
g-	-67	-63	-65	-0.02	-0.01	0.12	0.29	0.03	0.05	0.46	0.59	0.29	0.09	0.05	0.13	0.85	0.67	0.51
g+	67	63	65	-0.02	-0.01	0.12	0.29	0.03	0.05	0.46	0.59	0.29	0.09	0.05	0.13	0.85	0.67	0.51

[a] Dihedral X-CH_2-CH_2-CH_3

[b] X=CH_3 (butane); X=N (backbone nitrogen); X=C (backbone carbonyl carbon)

Table 8

Pentane ab initio, CHARMM, and experimental energies

	Ab initio[a]			CHARMM			
Conf.	χ1	χ2	ΔE	χ1	χ2	ΔE	Energy term
t,t	180.0	-180.0	0.00	180.0	180.0	0.00	
t,a-				179.8	-120.0	3.47	a
t,g-				-177.1	-67.6	0.87	g
t,s				180.0	0.0	5.45	s
t,g+	177.2	68.8	0.86	177.1	67.6	0.87	g
t,a+				-179.8	120.0	3.47	a
g+,a-				68.7	-120.0	4.30	g+a
g+,g-				89.8	-71.3	3.51	2g+p
g+,s				82.6	0.0	7.86	g+s+p
g+,g+	63.8	63.4	1.36	65.3	65.3	1.79	2g
g+,a+				69.3	120.0	4.30	g+a
g-,a-				-69.3	-120.0	4.30	g+a
g-,g-				-65.3	-65.3	1.79	2g
g-,s				-82.6	0.0	7.86	g+s+p
g-,g+	-94.6	63.2	3.33	-89.8	71.3	3.51	2g+p
g-,a+				-68.7	120.0	4.30	g+a

[a] MP3/6-31G*//MP2/6-31g* (Wiberg and Murcko, JACS, 110, 8029-8038 (1988))

Table 9

Energies of X-CH_2-CH_2-CH_2-CH_3

Conf	χ_1[a]			χ_2[b]			Angle energy			Dihedral energy			Electrostatic energy			Van der Waals energy			Total energy		
X:[c]	CH_3	N	C	CH_3	N	C	CH_3	N	C	CH_3	N	C	CH_3	N	C	CH_3	N	C	CH_3	N	C
t,t	180	180	180	180	180	-180	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
t,g-	-177	-177	-178	-67	-65	-67	0.01	0.02	-0.02	0.32	0.29	0.31	0.37	0.06	0.33	0.15	0.13	0.16	0.87	0.52	0.83
t,g+	177	177	178	67	66	67	0.01	0.02	-0.02	0.32	0.29	0.31	0.37	0.06	0.33	0.15	0.13	0.16	0.87	0.52	0.82
g-,t	-67	-63	-66	-177	-178	-177	0.01	0.00	0.07	0.32	0.03	0.06	0.37	0.43	0.21	0.15	0.07	0.17	0.87	0.53	0.54
g+,t	67	63	66	177	178	177	0.01	0.00	0.07	0.32	0.03	0.06	0.37	0.43	0.21	0.15	0.07	0.17	0.87	0.53	0.54
g-,g-	-65	-62	-64	-65	-66	-65	0.07	0.05	0.12	0.55	0.31	0.31	0.83	0.63	0.61	0.28	0.21	0.32	1.79	1.24	1.43
g+,g+	65	62	64	65	66	65	0.07	0.05	0.12	0.55	0.31	0.31	0.83	0.63	0.61	0.28	0.21	0.32	1.79	1.24	1.43
g-,g+	-71	-75	-76	89	74	79	0.27	0.25	0.47	1.78	0.89	1.05	0.97	0.92	0.76	0.34	0.45	0.51	3.52	2.63	2.98
g+,g-	89	75	76	-71	-74	-79	0.27	0.24	0.47	1.78	0.89	1.05	0.97	0.92	0.76	0.34	0.45	0.51	3.52	2.63	2.98

[a] Dihedral X-CH_2-CH_2-CH_2

[b] Dihedral CH_2-CH_2-CH_2-CH_3

[c] X=CH_3 (pentane); X=N (backbone nitrogen); X=C (backbone carbonyl carbon)

Table 10

Backbone N,C interactions with χ_1: Single Xγ (Abu, Ser, Cys)

Conformation		Interactions		Energies (kcal/mol)					Probabilities (%)					χ_1 (calculated)		
χ_1 [a]	χ_1^C [b]	N,γ	C,γ	E	ΔE	Calc. Abu	Calc. Ser	Calc. Cys	Calc. Abu	Calc. Ser	Calc. Cys	PDB Ser	PDB Cys	Abu	Ser	Cys
60	-60	g+	g-	2g	g	0.7	1.8	1.6	14	4	5	43	12	58	54	52
180	60		g+	1g		0.2	1.1	0.9	37	12	17	24	25	-176	-173	-169
-60	180	g-		1g		0.0	0.0	0.0	49	84	78	32	62	-64	-63	-63

[a] N-Cα-Cβ-Xγ

[b] C-Cα-Cβ-Xγ

[c] Backbone independent rotamer library (Table 3)

Table 11

Backbone N,C interactions with χ_1: Val (CH_3 at χ_1 and χ_1+120°)

Conformation				Interactions				Energies			Prob. (%)		
χ_1 [a]	χ_1^{C} [b]	$\chi_1^{\gamma 2}$ [c]	$\chi_1^{C,\gamma 2}$ [d]	N,Cγ1	C,Cγ1	N,Cγ2	C,Cγ2	E	ΔE	Calc. Val	Calc. Val	PDB Val	χ_1 (calc.)
60	-60	180	60	g+	g-		g+	3g	g	0.6	21	8	62
180	60	-60	180		g+	g-		2g		0.0	58	70	-179
-60	180	60	-60	g-		g+	g-	3g	g	0.6	22	22	-67

[a] N-Cα-Cβ-Cγ1 = χ_1

[b] C-Cα-Cβ-Cγ1 = χ_1 - 120°

[c] N-Cα-Cβ-Cγ2 = χ_1 + 120°

[d] C-Cα-Cβ-Cγ2 = χ_1

Table 12

Backbone N,C interactions with χ_1: Thr (OH at χ_1 and CH_3 at χ_1-120°)

Conformation				Interactions				Energies			Prob. (%)		
χ_1 [a]	χ_1^C [b]	$\chi_1^{\gamma 2}$ [c]	$\chi_1^{C,\gamma 2}$ [d]	N,Oγ1	C,Oγ1	N,Cγ2	C,Cγ2	E	ΔE	Calc. Thr	Calc. Thr	PDB Thr	χ_1 (calc).
60	-60	-60	180	g+	g-	g-		3g	g	0.6	21	46	52
180	60	60	-60		g+	g+	g-	3g	g	0.5	24	9	-174
-60	180	180	60	g-			g+	2g		0.0	56	45	-58

[a] N-Cα-Cβ-Oγ1 = χ_1

[b] C-Cα-Cβ-Oγ1 = χ_1 - 120°

[c] N-Cα-Cβ-Cγ2 = χ_1 - 120°

[d] C-Cα-Cβ-Cγ2 = χ_1 + 120°

Table 13

Backbone N,C interactions with χ_1, χ_2: Single Cγ, Single Cδ(Ape, Met, Glu, Gln, Arg, Lys)

Conformation			Interactions					Energies			Probabilities (%)							
χ_1	χ_1^C	χ_2	N,Cγ	N,Cδ	C,Cγ	C,Cδ	Cα,Cδ	E	ΔE	Calc. Ape	Calc. Ape	PDB Met	PDB Glu	PDB Gln	PDB Arg	PDB Lys	Ape χ_1 Calc.	Ape χ_2 Calc.
60	-60	60	g+		g-	p	g+	p+3g	p+2g	2.9	0	1	1	2	1	1	47	73
		180	g+		g-			2g	g	0.6	10	7	7	6	8	7	55	173
		-60	g+	p	g-		g-	p+3g	p+2g	2.7	0	1	3	1	0	1	69	-67
180	60	60			g+		g+	2g	g	0.2	18	10	7	10	5	6	-176	63
		180			g+			1g		0.0	28	20	24	21	24	25	-173	169
		-60			g+	p	g-	p+2g	p+g	1.9	1	2	1	2	1	2	-161	-70
-60	180	60	g-	p			g+	p+2g	p+g	1.9	1	2	8	4	3	3	-62	84
		180	g-					1g		0.0	26	34	33	39	45	40	-57	175
		-60	g-				g-	2g	g	0.4	15	22	14	15	11	12	-56	-61

Table 14

Backbone N,C interactions with χ_1, χ_2: Ile (CH_2CH_3 at χ_1 and CH_3 at χ_1-120°)

Conformation						Interactions								Energies			Prob. (%)			
χ_1 N,γ1 [a]	C,γ1 [b]	N,γ2 [c]	C,γ2 [d]	χ_2 α,δ [e]	δ,γ2 [f]	N, Cγ1	N, Cδ	C, Cγ1	C, Cδ	N, Cγ2	C, Cγ2	Cα, Cδ	Cγ2 Cδ	E	ΔE	Calc. Ile	Calc. Ile	PDB Ile	χ_1 (calc.)	χ_2 (calc.)
60	-60	-60	180	60	180	g+		g-	p	g-		g+		p+4g	p+g	2.0	1	2	47	73
				180	-60	g+		g-		g-			g-	4g	g	0.1	20	13	55	173
				-60	60	g+	p	g-		g-		g-	g+	p+5g	p+2g	2.6	0	0	69	-67
180	60	60	-60	60	180			g+		g+	g-	g+		4g	g	0.0	22	3	-176	63
				180	-60			g+		g+	g-		g-	4g	g	0.3	15	8	-173	169
				-60	60			g+	p	g+	g-	g-	g+	p+5g	p+2g	2.5	0	0	-161	-70
-60	180	180	60	60	180	g-	p				g+	g+		p+3g	p	2.0	1	3	-62	84
				180	-60	g-					g+		g-	3g		0.0	24	57	-55	175
				-60	60	g-					g+	g-	g+	4g	g	0.2	18	14	-56	-61

[a] N-Cα-Cβ-Cγ1 = χ_1

[b] C-Cα-Cβ-Cγ1 = χ_1 - 120°

[c] N-Cα-Cβ-Cγ2 = χ_1 - 120°

[d] C-Cα-Cβ-Cγ2 = χ_1 + 120°

[e] Cα-Cβ-Cγ1-Cδ = χ_2

[f] $C\gamma_2$-Cβ-Cγ1-Cδ = χ_2 + 120°

Table 15

Backbone N,C interactions with χ_1, χ_2: Leucine (CH3 at χ_2, χ_2+120°)

Conformation				Interactions								Energies			Prob. (%)			
χ_1 [a]	χ_1 [b]	χ_2 [c]	χ_2 [d]	N, Cγ	N, Cδ1	N, Cδ2	C, Cγ	C, Cδ1	C, Cδ2	Cα, Cδ1	Cα, Cδ2	E	ΔE	Calc. Leu	Calc. Leu	PDB Leu	χ_1 (calc)	χ_2 (calc)
60	-60	60	180	g+			g-	p		g+		p+3g	p+g	2.8	0	1	47	76
		180	-60	g+		p	g-				g-	p+3g	p+g	2.6	1	1	70	171
		-60	60	g+	p		g-		p	g-	g+	2p+4g	2p+2g	5.3	0	0	61	-52
180	60	60	180				g+			g+		2g		0.0	47	27	-178	65
		180	-60				g+		p		g-	p+2g	p	1.9	2	5	-162	169
		-60	60				g+	p		g-	g+	p+3g	p+g	2.1	1	1	-166	-63
-60	180	60	180	g-	p					g+		p+2g	p	1.8	2	8	-76	76
		180	-60	g-							g-	2g		0.0	45	55	-62	176
		-60	60	g-		p				g-	g+	p+3g	p+g	2.0	2	3	-74	-46

[a] N-Cα-Cβ-Cγ = χ_1

[b] C-Cα-Cβ-Cγ = χ_1 - 120°

[c] Cα-Cβ-Cγ-Cδ1 = χ_2

[d] Cα-Cβ-Cγ-Cδ2 = χ_2 + 120°

Table 16

Aromatics (CH at χ_2,χ_2+180°)

Conformation				Interactions								Energies	Calc. ΔE	Calc. %	PDB %	Calc. χ_2	Calc. ΔE	Calc. %	PDB %	Calc. χ_2	Calc. ΔE	Calc. %	PDB %	Calc. χ_2
χ_1	χ_1 C	χ_2 α,δ_1	χ_2 α,δ_2	N,γ	C,γ	γ,δ_1	γ,δ_2	N,δ_1	N,δ_2	C,δ_1	C,δ_2		Phe/ Tyr	Phe/ Tyr	Phe/ Tyr	Phe/ Tyr	Trp +90°	Trp +90°	Trp +90°	Trp +90°	Trp -90°	Trp -90°	Trp -90°	Trp -90°
60	-60	0	-180	g+	g-	s		p		p		2g+2p+s												
		60	-120	g+	g-	g+	a-			p		3g+p+a												
		90	-90	g+	g-							2g	2.2	2	13	95	1.7	2	5	91	1.3	3	11	-80
		120	-60	g+	g-	a+	g-		p			3g+p+a												
		180	0	g+	g-		s		p		p	2g+2p+s												
180	60	0	-180		g+	s				p		g+p+s												
		60	-120		g+	g+	a-					2g+a												
		90	-90		g+							g	0.5	30	34	75	0.4	16	19	81	0.2	22	13	-103
		120	-60		g+	a+	g-				p	2g+p+a												
		180	0		g+		s				p	g+p+s												
-60	180	0	-180	g-		s		p				g+p+s									0.7	10		-1
		60	-120	g-		g+	a-	p				2g+p+a												
		90	-90	g-								g	0.0	69	53	98	0.0	31	38	92	0.4	16	13	-88
		120	-60	g-		a+	g-					2g+a												
		180	0	g-			s		p			g+p+s												

Table 17

Aspartic acid and Asparagine (O at χ_2, O or NH_2 at χ_2+180°)

Conformation				Interactions										Energies					
χ_1	χ_1^C	χ_2 α,δ_1	χ_2 α,δ_2	N,γ	C,γ	γ,δ_2	γ,δ_2	N,δ_1 Asp	N,δ_2 Asp	N,δ_1 Asn	N,δ_2 Asn	C,δ_1	C,δ_2	Asp	Asn	Calc. ΔE Asp	Calc. χ_2 Asp	Calc. ΔE Asn	Calc. χ_2 Asn
60	-60	-180	0	g+	g-		s		p+e		p-e		p	2g+2p+s+e	2g+2p+s-e	7.8	-176		
		-120	60	g+	g-	a-	g+						p	3g+p+a	3g+p+a	6.0	-112	3.5	-113
		-60	120	g+	g-	g-	a+	p+e		p+e				3g+p+a+e	3g+p+a+e				
60	-60	0	-180	g+	g-	s		p+e		p+e		p		2g+2p+s+e	2g+2p+s+e	7.8	5		
		60	-120	g+	g-	g+	a-					p		3g+p+a	3g+p+a	6.0	71		
		120	-60	g+	g-	a+	g-		p+e		p-e			3g+p+a+e	3g+p+a-e			2.1	111
		180	0	g+	g-		s		p+e		p-e		p	2g+2p+s+e	2g+2p+s-e	7.8	-176		
180	60	-180	0		g+		s						p	g+p+s	g+p+s	0.8	-173		
		-120	60		g+	a-	g+							2g+a	2g+a	0.0	-123	2.2	-109
		-60	120		g+	g-	a+					p		2g+p+a	2g+p+a	1.2	-73		
180	60	0	-180		g+	s						p		g+p+s	g+p+s	0.8	8	1.5	16
		60	-120		g+	g+	a-							2g+a	2g+a	0.0	57	1.3	57
		120	-60		g+	a+	g-						p	2g+p+a	2g+p+a	1.2	107		
		180	0		g+		s						p	g+p+s	g+p+s	0.8	-173		
-60	180	-180	0	g-			s		p+e		p-e			g+p+s+e	g+p+s-e				
		-120	60	g-		a-	g+		p+e		p-e			2g+p+a+e	2g+p+a-e				
		-60	120	g-		g-	a+							2g+a	2g+a	1.2	-59	0.0	-77
		0	-180	g-		s		p+e		p+e				g+p+s+e	g+p+s+e				
		60	-120	g-		g+	a-	p+e		p+e				2g+p+a+e	2g+p+a+e				
		120	-60	g-		a+	g-							2g+a	2g+a	1.2	123	1.8	118
		180	0	g-			s		p+e		p-e			g+p+s+e	g+p+s-e				

Table 18

Effect of backbone dihedral ϕ on χ1 rotamers

Conformation					Interactions				Energies		
ϕ [a]	ϕβ [b]	χ1 [c]	χ1- 120°	χ1+ 120°	C^{-1}, β	C^{-1}, γ	C^{-1}, γ-120°	C^{-1}, γ+120°	γ	γ & γ-120°	γ & γ+120°
-180	60	60	-60	180	g+		p			p	
		180	60	-60	g+			p			p
		-60	180	60	g+	p			p	p	p
-60	180	60	-60	180							
		180	60	-60							
		-60	180	60							
60	-60	60	-60	180	g-	p			p	p	p
		180	60	-60	g-		p			p	
		-60	180	60	g-			p			p

[a] C_{i-1}-N-Cα-C = ϕ

[b] C_{i-1}-N-Cα-Cβ = ϕ -120°

[c] N-Cα-Cβ-Xγ

Table 19

Effect of backbone dihedral ψ on χ_1 rotamers

Conformation							Interactions								Energies		
ψ [a]	$\psi^{N\beta}$ [b]	$\psi^{O\beta}$ [c]	χ_1 [d]	χ_1^{C} [e]	χ_1^{C}– 120°	χ_1^{C}+ 120°	N_{i+1}, β	N_{i+1}, γ	N_{i+1}, γ-120	N_{i+1}, γ+120	O, β	O, γ	O, γ–120	O, γ+120	γ	γ, γ-120 Thr, Ile	γ, γ+120 Val
-180	-60	120	60	-60	180	60	g-			p	a+						p
			180	60	-60	180	g-	p			a+				p	p	p
			-60	180	60	-60	g-		p		a+					p	
-120	0	180	60	-60	180	60	s										
			180	60	-60	180	s										
			-60	180	60	-60	s										
-60	60	-120	60	-60	180	60	g+	p			a-				p	p	p
			180	60	-60	180	g+		p		a-					p	
			-60	180	60	-60	g+			p	a-						p
0	120	-60	60	-60	180	60	a+				g-			p			p
			180	60	-60	180	a+				g-	p			p	p	p
			-60	180	60	-60	a+				g-		p			p	
60	180	0	60	-60	180	60					s						
			180	60	-60	180					s						
			-60	180	60	-60					s						
120	-120	60	60	-60	180	60	a-				g+	p			p	p	p
			180	60	-60	180	a-				g+		p			p	
			-60	180	60	-60	a-				g+			p			p
180	-60	120	60	-60	180	60	g-			p	a+						p
			180	60	-60	180	g-	p			a+				p	p	p
			-60	180	60	-60	g-		p		a+					p	

[a] N-Cα-C-N_{i+1} = ψ

[b] N_{i+1}-C-Cα-Cβ = ψ+120°

[c] O-C-Cα-Cβ = ψ-60°

[d] N-Cα-Cβ-Xγ = χ_1

[e] C-Cα-Cβ-Xγ = χ_1-120°

Chapter 5

Homology Modeling of HLA-B35, HLA-B53, HLA-Cw4, HLA-Cw6 and HLA-Cw7 from HLA-B27

Abstract

The structures of six Class I human leukocyte antigens (HLA) are predicted from the x-ray structure of HLA-B27 using a homology modeling scheme that uses information from the template protein (HLA-B27) in combination with a backbone-dependent rotamer library to predict sidechain conformation. Simple rules are developed and presented to determine whether the conformation of each sidechain in the initial model comes from the template protein or from the rotamer library. As a test of the method, we have used it predict the placement of sidechains in HLA-A68 whose structure has been determined previously. The method works quite well, especially for sidechains in the antigen binding site. Of 30 sidechains which extend into the antigen binding site and determine which peptides will bind to HLA-A68, 27 are correctly predicted within 40° by the homology modeling scheme used here.

The method is used to predict the structures of HLA-B35, HLA-B53, HLA-Cw4, HLA-Cw6, and HLA-Cw7. HLA-B53 provides a defense against malaria, while the closely related allele HLA-B35 does not. These two proteins differ by only 5 amino acids, all in the end of the peptide binding groove which binds the C-terminus of foreign peptides. Sequences of peptides eluted from B35 and B53 have been determined by Hill et al. [1]. We have used the predicted structures to rationalize the binding specificities of these two HLA alleles.

Recently, peptides eluted from three HLA-C alleles have been sequenced by Rötzschke et al. [2]. These three alleles share a common identity for the anchor residue at the second position of the peptide, but different residues elsewhere. We have predicted the structures of these three HLA-C alleles, and used them to analyze the relationship between the structure and character of the antigen binding site and the sequences of bound peptides.

I. Introduction

The first major task of immune surveillance is the identification of non-self from self, after which the elimination of non-self can occur. This task is performed in part by two groups of proteins: immunoglobulins produced by B-cells and T-cell receptors produced by CD4+ and CD8+ T-cells. In the first case, both membrane bound and soluble immunoglobulins (or "antibodies") bind to intact protein antigens from foreign organisms and viruses. Once an antibody is bound to its foreign antigen, other functions of the immune system, such as macrophages, natural killer cells, neutrophils, and the proteins of complement can home in on and disable or kill the invading organism. Immunoglobulins are made by approximately 10^7 clonally distinct sets of B-cells, each capable of making an antibody of different sequence and structure. B-cell clones which make antibodies to self-antigens are either eliminated in B-cell development or made anergic by other immune system cells.

Distinct from the "humoral" immune system of antibodies, cellular recognition of infected or cancerous self cells proceeds via recognition of complexes of human leukocyte antigen proteins (HLA) and foreign peptides on affected cells. This recognition is made by helper and killer T-cells via their membrane-bound T-cell receptors which bind to the HLA/peptide complex on infected or cancerous cells. In a fashion similar to B-cells, each T-cell line produces a different T-cell receptor protein on its surface composed of $\alpha\beta$ or $\gamma\delta$ TCR protein heterodimers. Gene rearrangement of the α and β chain genes in $\alpha\beta$ T-cells and the $\gamma$and δ chain genes on $\gamma\delta$ T-cells can produce approximately 10^8 different TCR proteins respectively.

Class I HLA proteins on almost all cell types and Class II HLA proteins on B-cells, macrophages, and other antigen presenting cells routinely bind self peptides (derived from proteolyzed intracellular proteins transported into the endoplasmic reticulum [3, 4]) of 8 to 10 amino acids in length (Class I) [1, 5, 6] or 13-25 amino acids in length (Class II) [7]. These HLA/peptide complexes are placed on the surface of the cell

for presentation to cytotoxic T-cells (reviewed by [8]). T-cells bearing T-cell receptors (TCR's) capable of binding to such complexes with self-peptides are eliminated during T-cell maturation in the thymus during fetal and early post-natal development or are otherwise made anergic [9]. If the bound peptide, however, is from a foreign source such as a bacterial or viral infection, a TCR from a particular T-cell may bind to the HLA protein/foreign peptide complex. A killer T-cell will respond by secreting destructive proteins (e.g. perforins); a helper T-cells will secrete cytokines such as interleukin 2 which will stimulate other immune system cells to respond [10].

Class I and Class II genes are located in a large region of chromosome 6 in humans called the major histocompatibility complex (MHC). For both Class I and Class II, there are three loci – HLA-A, HLA-B, and HLA-C for Class I and HLA-DP, HLA-DQ, and HLA-DR for Class II [11]. In addition, there are a number of other HLA-like genes, such as CD1 whose function is unknown. Each individual therefore produces six Class I proteins and six Class II proteins. Compared to most human genes, the HLA loci are highly polymorphic [12]. To date, 41 HLA-A sequences, 61 HLA-B sequences, 18 HLA-C sequences, 8 HLA-DPA sequences, 31 HLA-DPB, 14 HLA-DQA sequences, 19 HLA-DQB sequences, 2 HLA-DRA sequences, and 72 HLA-DRB have been identified [11].

The polymorphism at particular sites in HLA protein sequences was understood only when the first Class I structure (HLA-A2) was solved by x-ray crystallography [13, 14, 15]. The protein was found to have a domain consisting of two long α-helices on top of a β-sheet on top of two immunoglobulin domains. The helices and sheet form a long groove with dimensions sufficient for accommodating an 8 to 10 amino acid peptide. Most of the polymorphic sites in HLA protein sequences were found to line the floor and sides of the groove as well as the tops of the α-helices. Saper et al. (1991) analyzed the site of HLA-A2 in terms of six "pockets" named A through F. These pockets were of varying depth from 7 to 12 Å and were either hydrophobic (D pocket), neutral and polar

(A,B,F pockets) or positively charged (C and E pockets). Electron density not accounted for by the HLA protein sequence was found in the groove, and was attributed to peptides with heterogeneous sequences. Pockets A and F bind the N and C termini respectively of heterogeneous peptides. Peptides eluted from HLA-A2.1 have been sequenced [16, 17], and tend to have Leu at position 2 and valine at position 9. The remaining positions have a variety of amino acid types present in pooled eluted peptides. Viral peptides which are known to bind to HLA-A2.1 fit the peptide binding motif derived from the pooled peptide data [16]. The structure at 2.6 Å resolution showed electron density extending deep into pocket B and less pronounced density at the right end of the site near pocket F.

Subsequently, the structures of two other HLA alleles, HLA-A68 (formerly "HLA-Aw68") [6, 18, 19] and HLA-B27 [6, 18, 19, 20, 21] have been solved by Wiley and colleagues as well as the mouse Class I protein H-2K^b by two other groups [22, 23]. As discussed in the preceding paragraph, the structure of HLA-A2 has been resolved to 2.6 Å with heterogeneous peptide present in the site [15]. HLA-A68 has been resolved to 1.9 Å with heterogeneous peptide [6], and to 2.8 Å with the influenza nucleoprotein peptide Np 91-99 bound in the site [19]. While HLA-B27 has not been studied with a single peptide, the 2.1 Å structure shows well-resolved electron density for an extended chain peptide of 9 amino acids in length in the peptide-binding groove [21]. The mouse Class I protein H-2K^b has been solved with peptides from vesicular stomatitis virus nucleoprotein (52-59) [22, 23] and Sendai virus nucleoprotein (324-332) [22] to 2.3 and 2.8 Å resolution respectively.

Analysis of a high resolution structure of HLA-A68 with heterogeneous bound peptides [6] and with bound influenza Np 91-99 peptide [19] showed a binding pattern slightly different from the HLA-A2 example. As with the HLA-A2 structure, the peptide was bound by its N-terminus and C-terminus in the deep A and F pockets at each end of the groove formed by the α1 and α2 domain α-helices. Consistent with the sequences of eluted peptides from HLA-A68 [6], there was sufficient space in the

peptide binding site for a small amino acid (Val or Thr) at position 2 (P2) in the shallow B or "45" pocket of HLA-A68 and an arginine or lysine at the last peptide position (P9 or PC) in a deep negatively charged pocket (F pocket). In fact, Np 91-99 has Thr and Arg sidechains at P2 and P9 that bind in the two pockets.

The 2.1 Å crystal structure of HLA-B27 showed well-defined density for a peptide backbone of 9 amino acids [21]. Electron density for an arginine sidechain at the second position of the peptide (P2) could be seen extending deeply into a pocket formed by the amino acids His-9, Thr-24, and Glu-45 (B or 45 pocket). Peptides eluted from B27 on LG-2 lymphoblastoid cells are found to have arginine uniformly at the P2 position [5]. This pocket in HLA-A2 [15] and A68 is much smaller and neutral and the P2 position is usually occupied by a Leu residue in A2 [17] or a Val or Thr residue in A68 [6]. Also in HLA-B27 at P9 there is usually a basic sidechain (Arg or Lys) that binds in a deep pocket formed by residues Leu-81, Tyr-123, Thr-143, Asp-74, Asp-77, and Asp-116. In some cases, the peptide sidechain is hydrophobic, which model building suggests can be accommodated in the pocket if the Asp sidechains form salt-bridges with positive sidechains of the protein nearby [21].

The principles developed from the human Class I proteins were confirmed by studies on the mouse Class I protein H-2K^b by Wilson and colleagues [22, 24]. As with the other Class I structures, there were deep pockets at each end of the antigen binding site that accommodated the N and C termini of the bound octamer peptide. There was also a deep pocket in the center of the peptide binding groove, which was found to contain the tyrosine P5 sidechain of the VSV-8 peptide or the SEV-9 peptide in each of the two crystal structures studied. This deep pocket in the center of the groove is much shallower or absent in the human Class I proteins so far studied [24]. The change in shape is caused by the replacement of large residues in HLA proteins by smaller residues (e.g. F9V, R97V, Y99S) as well as a rotation of Tyr116 away from the center of the groove. Sidechains of peptide SEV-9 residues P2, P3, P6, P7, and P9 were found to

contact the floor and sides of the groove in the B pocket (peptide residue P2; near H-2K^b residues 24, 45), the D pocket (P3 and 99, 156), the C pocket (P6 and 9, 97), the E pocket (P7 and 114, 152), and the F pocket (P9 and 77, 143). Residues P6 and P9 were found to extend vertically down into the site, whereas P2, P3, and P7 were found to be horizontal within the site.

In sum, the analysis of Matsumura et al.(1992) has shown that the known structures of Class I proteins (HLA-A2, HLA-A68, HLA-B27, H-2K^b) all have one or two pockets deep enough to bind only specific sidechains, which they refer to as "anchor" residues. There are also one or two shallower pockets that can accommodate a variety of peptide sidechains. The shape and nature of these pockets varies from Class I protein to Class I protein, altering the specific motifs which bind to each allele. Several sidechains also extend away from the site. These positions are usually found to be variable in sequences of eluted peptides, and presumably interact with the T-cell receptor during cell-mediated immune recognition.

The specificity of Class I proteins for certain peptide sequences can lead to a tendency toward susceptibility or immunity to certain diseases. Some autoimmune diseases such as psoriasis [25] and ankylosing spondylitis [26] are associated with genes for particular HLA alleles such as HLA-B27. Immunity to disease has also been associated with HLA type. One example includes the HLA allele B53 and resistance to malaria, which has been attributed to successful presentation of the liver-stage specific antigen (LSA) peptide of *Plasmodium falciparum* by HLA-B53, but not by the related HLA-B35 allele which differs from HLA-B53 by five amino acids in the peptide-binding site [1]. The structure of these two proteins, however, is not known. In this paper we present models of these two proteins and discuss the interactions with the *P. falciparum* peptides that determine their specificity.

The purpose of HLA-C proteins has been the subject of some controversy [27, 28]. HLA-C proteins are expressed at levels as low as 10% of the HLA-A and B

expression levels on cell surfaces [29, 30, 31]. Their role in antigen recognition has therefore been questioned. However, they have been associated with certain diseases such as acute leukemia [32, 33], psoriasis (HLA-Cw6) [25], and type 2 diabetes [34] and they have been found to be responsible for allorecognition responses [35, 36, 37]. HLA-C recognizes a peptide from HIV gag proteins [38] and peptides from influenza virus and Sendai virus in HLA-Cw3 transgenic mice [39], and has been found to confer recognition of Epstein-Barr virus infected cells to cytotoxic T-lymphocytes [37, 40]. In addition they have been implicated in the response of natural killer cells to infected cells [28, 41]. Their sequence variability has not been studied extensively, but is believed to be lower than HLA-A and HLA-B [27]. For example, HLA-Cw7 minor variants account for ~40% of Caucasian populations [33]. HLA-C alleles have less variation in the antigen binding site than HLA-A and HLA-B alleles, but more variation outside the site [27].

Recently Rötzschke et al. [2] have sequenced pooled peptides eluted from HLA-C proteins HLA-Cw4, HLA-Cw6, and HLA-Cw7 and found sequence motifs which differ from the A and B motifs. At position 2, Tyr and Phe were found to be anchor residues for HLA-Cw4 and Cw7, but not for Cw6. For Cw6, no single residue was dominant at position 2, and Pro and Arg were reproducibly detected at low levels. In all three alleles, hydrophobic sidechains were commonly found at positions 5 and 6. These were referred to as "auxiliary anchors," meaning that several sidechains of a similar character could bind in these positions rather than just one or two. At position 9 in all three alleles, hydrophobic residues were also identified as "anchors."

Since the structures of HLA-B35, HLA-B53, HLA-Cw4, HLA-Cw6, and HLA-Cw7 are not known and because eluted peptide sequence data is available for these alleles, we undertook to model the structures of these five proteins. The purpose was to build the structures and attempt to rationalize the eluted peptide sequence data. We have used the predicted structures to predict the size and chemical nature of the pockets in the peptide binding site in analogy to the studies on other HLA and H-2 proteins [6, 15, 19,

21, 24, 42]. Since C alleles are closer to B alleles than A alleles [27], we used the 2.1 Å structure of HLA-B27 as a starting structure [21]. We believe that one can learn more from studying structures of proteins, even if inaccurate, than from protein sequence comparisons of the various pockets, since it is difficult to visualize the effect of mutations from one allele to another.

The method used is a variation on the methods presented by Summers and Karplus [43] and Dunbrack and Karplus [44]. It uses sidechain coordinate information (Cartesian coordinates and dihedral angles) from the known structure in combination with a backbone-dependent rotamer library . The Cartesian coordinates of the backbone of the structure to be modeled (the "target") are obtained directly from the template protein – in this case, HLA-B27. In cases in which residue types are identical in the template and target sequences, the Cartesian coordinates of the sidechain are transferred from the template. When the residues are not the same, one of two things can happen. If the residues are of similar structure (e.g. Val, Ile, Thr), then dihedral angles (χ_1 and χ_2) can be obtained from the template structure. In cases in which the residues are of very different structure or character (e.g. Trp versus Val), dihedral information for the target sidechain is obtained from a backbone-dependent rotamer library [44]. There are also intermediate cases, when χ_1 is obtained from the template but χ_2 is obtained from the library because of the different nature of the atoms further along in the chain (e.g. Phe versus Arg). The change from the previous methods [43, 44] consists of a simple set of rules for determining which sidechain dihedral angles should be kept from the template structure (HLA-B27) and which should be obtained from the backbone-dependent rotamer library. The rules are described below. In the following section, it is tested on the HLA-B27-> HLA-A68 problem. It is then used to predict the three HLA-C alleles whose peptides were sequenced by Rötzschke et al.(1993) as well as HLA-B35 and HLA-B53 whose peptides were sequenced by Hill et al.(1992).

It will be shown that the peptide binding data is easily rationalized by the predicted structures. In principle, the structures of any HLA allele can be predicted and peptide binding motifs predicted. In the near future, we plan to present the predicted structures of a number of other common HLA alleles, so that they can be analyzed in terms of likely peptide sequence motifs.

II. Methods

The method used for predicting sidechain conformations a homology modeling scheme that combines rules proposed by Summers and Karplus [43] with information derived from a backbone-dependent rotamer library derived by Dunbrack and Karplus [44]. The principle of the combined method is to obtain as much information as possible from the homologous template protein for structurally related sidechains and to use the backbone-dependent rotamer library to provide information about structurally dissimilar sidechains.

A. Construction of Initial Model

(i) *Backbone coordinates.*

As described by Dunbrack and Karplus [Dunbrack, 1993 #55; Chapter 3 of this thesis], homology modeling of an unknown protein structure proceeds from the known Cartesian coordinates of a homologous protein by first using the template to obtain x,y,z coordinates of the backbone atoms of the target protein. In the present case, there are no insertions and only a single amino acid deletion between HLA-B27 and HLA-Cw7 (HLA-Cw*0702). Rather than model the deletion, a glycine residue was placed at position 135 of HLA-Cw7. This residue is located in a strand of the $\alpha 2$ β-sheet outside the antigen binding site and is unlikely to perturb it.

(ii) *Sidechain placement*

Once the backbone coordinates of the target model are obtained, sidechain coordinates for an initial model structure must be determined. For sidechains which are identical in the template and target sequences, the Cartesian coordinates of the template protein sidechain are copied directly to the new structure (method temp/temp, Ref. [44]). When sidechains are not identical in the template and target sequences, information is obtained from the template sidechain dihedral angles and the backbone-dependent rotamer library of Dunbrack and Karplus [44]. Based on a comparison of homologous protein sidechain conformations [43] and an extensive analysis of protein sidechain conformations (Chapter 4 of this thesis), rules have been derived for determining initial sidechain conformations from template sidechain dihedral angles and the backbone-dependent rotamer library. Sidechains have been grouped together based on the patterns of their backbone-dependent rotamer preferences (see Tables 4 and 5 in Chapter 4 of this thesis).

First, the amino acids types have been divided into eight groups based on their structural similarities:

1) Ser, Cys

2) Glu, Gln, Met, Arg, and Lys

3) Phe, Tyr, His, Trp

4) Asp, Asn

5) Leu

6) Thr, Val, Ile

7) Pro

8) Gly, Ala

The first five groups have a single γ heavy atom, but differ in the number and character of their δ and ε atoms. The group 1 sidechains have no δ heavy atoms; group 2 sidechains have a single δ heavy atom; group 3 consists of the aromatic sidechains, where χ_2 is generally +90° or -90°; group 4 contains the short polar sidechains Asp and Asn

with two δ heavy atoms at χ_2 and χ_2+180°; group 5 contains only leucine which has two δ methyl groups at χ_2 and χ_2+120°. Sidechains with two γ heavy atoms make up group 6; proline is by itself in group 7 because of its unique ring structure involving backbone atoms N and Cα; and group 8 consists of the residues without sidechain conformational flexibility – Gly and Ala.

Given these groupings and the case where the template and target sidechain amino acid types are not identical, the following rules are used to obtain information from the template protein where possible and the backbone-dependent library where necessary:

1) When both the template and target sidechains have a single γ heavy atom (both from groups 1-5), χ_1 from the template protein is used for the target sidechain.

2) When both the template and target sidechains have two γ heavy atoms (both from group 6), χ_1 from the template protein is used to determine the heavy atom positions of the target sidechain. When the sidechains involved are Ile and Thr (i.e. Ile → Thr or Thr → Ile), then χ_1 from the template can be used directly. When one of the template sidechain is Val and the target sidechain is Ile or Thr, then χ_1+120° is used. When the opposite occurs (Ile or Thr → Val), χ_1-120° is used. In all cases, the heavy atoms are placed in the same position relative to the backbone, but because of the definitions of χ_1 of Val relative to Ile and Thr, χ_1 can not be used directly.

3) When the template sidechain is from groups 6, 7, or 8 and the target sidechain is from another group (1-5, 6, 7, 8 but not equal to the template group), the backbone-dependent rotamer library is used to place χ_1 for the target sidechain. Similarly, if the template sidechain is from groups 1-5 and the target is from groups 6, 7, or 8, the library is also used to obtain χ_1. When the library is used, bond lengths and angles are obtained from the CHARMM residue topology file, which have been calculated from minimized tetrapeptides of the form Ac-Ala-Xxx-Ala-$NHCH_3$ for each sidechain [44]. With the bond lengths, the bond angles, and the χ angles from the backbone-dependent rotamer library, the sidechain Cartesian coordinates can be determined. Since

we are using the all-hydrogen atom parameter set (MacKerell *et al*., to be published), both heavy atom and hydrogen atom bond lengths and angles are obtained by the tetrapeptide minimizations.

4) When both sidechains are from group 2, χ_2 is obtained from the template sidechain. If the substitution is Glu for Gln or vice versa, then χ_3 is also obtained from the template sidechain. Similarly, if the two sidechains are Lys and Arg (in either order), then χ_3 is also obtained from the template sidechain. In all other cases, χ_3 and χ_4 (in the case of Arg and Lys) are set to 180°, which is the most likely value in a protein databank survey [44].

5) When both sidechains are from group 3, then χ_2 is borrowed from the template. No effort has been made to orient histidine sidechains or to determine their protonation state. In all cases, histidine is arbitrarily protonated on the δ nitrogen (ND1), and a χ_2 value of +90° is used rather than -90°. When a substitution is made from Phe, Tyr, or His to Trp, the library is used to predict χ_2 of Trp from its χ_1 value (previously determined from rules 1-3 above). That is, given χ_1, the most common χ_2 rotamer for that ϕ,ψ,χ_1 combination is used as the value of χ_2 (either +90° or -90°).

6) When a mutation from Asp to Asn (group 4) or vice versa is made, χ_2 from the template sidechain is used for the target sidechain. No effort is made to determine the orientation of the sidechain (by ±180°) in the target protein when the sidechain is an Asn residue.

7) When the template and target sidechains are from different groups and the target sidechain has a χ_2 degree of freedom, the library is used to predict χ_2 from the value of χ_1 determined from rules 1-3. Thus, in all cases where the target sidechain is Trp, Leu, or Ile (and the template sidechain is not identical to the target sidechain), the library is used to predict χ_2. For all sidechains in this category, whether χ_1 is determined from the homologous template sidechain or from the library, χ_2 is chosen from the highest χ_2 rotamer population given ϕ, ψ, and χ_1.

8) Finally, the CHARMM residue topology file is used to set up the remaining coordinates which remain undefined. This involves the Ala sidechains, the backbone hydrogens, and Gly Hα.

(iii) *Hydrogen atom minimization*

From this point on, the method proceeds along the lines proposed by Dunbrack and Karplus [44]. The first structure (structure 0) is build from the Cartesian and internal coordinates just described. Hydrogen atoms are built using the CHARMM residue topology file. In the cases studied here, all of the six cysteines and the three disulfide bonds between them are conserved and are therefore not minimized. The hydrogen atom positions are minimized keeping the heavy atoms fixed using 50 steps of steepest descent and 100 steps of Powell, yielding Structure 0.

B. Refinement of Model

(i) *Sidechain minimizations (sidechain/backbone clashes)*

Steric clashes are determined and all sidechains clashing with the backbone above the same threshold energies as in Dunbrack and Karplus [44] are subjected to the same minimization scheme as previously. Once the minimizations are completed, all the sidechains are moved to their new coordinates. Hydrogen atom minimizations are repeated to produce Structure 1.

(ii) *Sidechain minimizations (sidechain-sidechain clashes except Ile, Thr, Val)*

In the second round, only mutated sidechains with steric clashes with other atoms are subjected to the minimization scheme. This follows the method of Summers and Karplus [43], where a preference for moving sidechains not identical in the template and target structures was proposed. Exceptions are also made for sidechains which are Val,

Ile, or Thr in the target structure, since these are well determined by the library and by homology modeling [Dunbrack, 1993 #55; Summers, 1989 #54; Chapter 4 of this thesis].

(iii) *Repeated sidechain minimizations (all clashes)*

In the third and subsequent rounds, any sidechain with steric clashes with other atoms is subjected to the minimization scheme and moved between rounds until all clashes are removed.

C. Visualization

When application of the model optimization was completed for the HLA proteins studied here, the final structures were used for analyzing the peptide binding site and likely peptide sequence motifs of each allele. The molecular graphics program Quanta was used to visualize the predicted structures. They were examined for their hydrophobic or hydrophilic character. Also, we used the program CHARMM to determine which atoms in the antigen binding site were accessible to 1.4 and 4.0 Å probes. Subsites which were accessible to the 1.4 Å probe but not to the 4.0 Å probe were considered to be potential pockets in the site, which may bind sidechains. More quantitative studies will be made with multiple-copy simulated search methods and ligand docking programs (A. Caflisch, E.-R. Evensen, R. L. Dunbrack, and M. Karplus, in progress).

D. Modeling sidechains at the P2 and P9 positions of bound peptide

The coordinates for the backbone of the peptide bound to HLA-B27 [21] were used to model peptides into the sites of the seven HLA proteins studied in this paper. We substituted 10 different sidechains into the P2 and P9 positions of the peptide using the same sidechain conformation prediction method as was used to build the protein models. Madden et al. have modeled a peptide with Arg at P2 and Lys at P9, and the

sidechain coordinates were used to model other sidechains according to the rules of Section II B above. At each position, we built in the following sidechains in turn: Ala, Glu, Phe, Ile, Leu, Asn, Pro, Arg, Ser, and Trp. These residues were chosen to be a representative set in terms of both size and charge (e.g. Lys and Arg are fairly similar as are Glu and Gln, etc. so Lys and Gln were not modeled).

Having built the sidechains into the peptide and positioned the peptide into the site according to the coordinates for the peptide backbone given by Madden et al., we used the program CHARMM to calculate the interaction energy of the sidechain with the protein as well as with the peptide itself. The peptide sidechain was then briefly minimized for 20 conjugate gradient minimizer steps to relieve van der Waals contacts with the protein or the rest of the peptide. During these minimizations, both the protein coordinates and the peptide backbone coordinates were held fixed. This was done to allow the sidechain to find space within the antigen binding site without disturbing conserved protein/peptide backbone interactions. The interaction energies between the sidechain and the protein and between the sidechain and the peptide were calculated. The results were compared with the unminimized energies, and correlated with the size of the subsite which accommodates the sidechain and the sequence information from eluted peptides for the P2 and P9 positions on the peptide.

III. Results

We first present information about the relatedness of the HLA proteins studied in this paper. In Figure 1, the sequences of HLA-A2, A68, B27, B35, B53, Cw4, Cw6, Cw7 are aligned. It can be readily seen that the HLA-B and HLA-C alleles differ from HLA-A alleles at a large number of sites. HLA-B35 and B53 differ from each other by five amino acids all in the F-pocket end of the antigen binding site (positions 77, 80, 81, 82, 83). In addition, HLA-Cw4 and Cw6 are somewhat closer to each other in sequence than either is to Cw7.

A. The prediction of HLA-A68 from HLA-B27 and comparison with the known structure of HLA-A68

In Table 1, we list the dihedrals of the predicted models of HLA-A68 and compare them with the x-ray structure at 1.9 Å resolution [6]. Column 1 lists the residue numbers of B27 and A68, and columns 2 and 3 list the sequences. Column 4 lists i for each χ_i and column 5 lists its value in the x-ray structure of A68. The next three columns describe Structure 0 of the modeling process, that is after identical sidechain coordinates have been borrowed from the template, dihedrals of similar sidechain types have been transferred from the template, and the backbone-dependent rotamer library has been used to place the remaining dihedrals. In the first column under Structure 0, the predicted value of each χ is listed. The next column lists the difference in χ angles from the x-ray structure of HLA-A68, and the third column lists a "y" if the predicted dihedral is within 40° of the experimental dihedral and an "n" otherwise. The next three columns provide similar information about Structure 1 (after residues conflicting with the backbone have been moved), and the final three columns cover Structure 3 – after all sidechain/sidechain clashes have been removed by the minimization procedure described in the Methods section.

Of the 32 sequence differences between B27 and A68 that result in sidechains to be predicted for A68, 21 are correctly predicted within 40° of the x-ray structure. The rate of 21/32 or 66% correct of non-identical sidechains compares with 164/195 or 84% of identical sidechains. The total prediction rate is 81% of 227 A68 sidechains (see the bottom of Table 1). This compares with an average prediction of 78% of χ_1's of all sidechains from the backbones of six proteins (i.e. no homologous protein information) in our previous work [44].

Of the 42 residues whose χ_1 values differ by more than 40° between the x-ray structure and the HLA-B27 derived model, 31 are identical in the A68 and B27 sequences.

Most of these involve charged sidechains – 5 Asp's, 4 Arg's, 6 Glu's, and 1 Lys for a total of 16 out of 31. A number of these are in unusual conformations either in A68 or in B27. These include "forbidden" rotamers (χ_1,χ_2 = {g+,g+}, {g+,g-}, {t, g-}, {g-, g+}; see Chapter 4 of this thesis; viz. Glu58, Arg111, Glu128 (in both structures), Glu177, Arg181, and Glu198) or non-rotamers (more than 40° from +60°, 180°, or -60°) (viz. Ser2, Glu128, Asp61, Tyr113, Ser132, and Glu198). The remainder consists of changes from one common rotameric conformation to another. Since the sidechain optimization procedure is biased toward moving mutated sidechains, most of these sidechains were not adjusted during the optimization (although several sidechains were moved from incorrect conformations to correct ones). Charged sidechains may be particularly prone to alternate conformations in homologous proteins, often to form particular interactions such as salt bridges or to gain access to solvent. Differences in conserved sidechains may result from interactions with mutated sidechains and subtle changes in backbone conformation that tend to favor one rotamer over another.

Bjorkman et al [14] list 32 amino acids which would have sidechains extending into the peptide binding site (in some alleles, there are glycines or alanines at these positions): 5, 7, 9, 22, 24, 26, 66, 67, 70, 73, 74, 77, 80, 81, 84, 95, 97, 99, 114, 116, 143, 146, 147, 152, 155, 156, 159, 163, 167, and 171. These sidechains have been printed in bold type in Table 1. In A68, there are sidechains (non-Ala, non-Gly) at 30 of these sites (i.e. excluding Ala24 and Gly26). In the model built from HLA-B27 (Table 1), 27 of these sidechains are correctly predicted within 40° (90%). The only exceptions are C67V (i.e. Cys in B27 and Val in A68), D116D, and L156W. In the case of Asp116, the B27 conformation is $\{\chi_1,\chi_2\}$ = {-71°,86°} while the A68 conformation is {65°,-150°}. The prediction from B27, {-71°,85°}, is sterically allowed in the A68 model, and so is not removed by the optimization procedure. In the case of L156W, χ_1 has been minimized to -132° and χ_2 to 57°, whereas the correct conformation is -87°, -69°. Apparently because of the incorrect orientation for χ_2, the value of χ_1 is off by 45°. It is encouraging that the

residues in the antigen binding site are well predicted by the procedure used here, indicating some confidence in this region for the other alleles to be studied.

B. Models of HLA-B35, B53 , Cw4, Cw6, and Cw7 from HLA-B27

(i) *Sidechain dihedral angles*

We followed the same procedure for modeling A68 from B27 to build structures of HLA-B35, B53, Cw4, Cw6, and Cw7. The dihedrals of sidechains which differ from the B27 template in sequence or in conformation as well as all the sidechains which line the antigen binding groove (in bold type) [14] are listed in Table 2 for the model structures (including A68 previously discussed). It is useful to follow the differences in dihedral values to determine how the method proceeded to place the sidechains, and how much confidence we can have in their placement. If the conformations of identical sidechains were different in the 7 alleles in many cases, then we should be suspicious of the predicted structures. Also, if there are substantial numbers of inherently unlikely rotamers (i.e. containing syn-pentane interactions; see Chapter 4 of this thesis; or having non-rotamer χ values), then again the results would be suspicious.

We have examined the sidechains which point into the antigen binding groove, highlighted in bold type in Table 2. There are 31 non-Gly, non-Ala sidechains which point into the site in these 7 alleles. Of these 31, 12 are identical across the 7 alleles (or mutated to Gly or Ala without a sidechain conformation to be modeled). In all 12 cases, the predicted χ angles are either identical to the original B27 structure (i.e. they were not adjusted by the minimization procedure) or are within 40° of the original dihedral value (i.e. after minimization to remove steric clashes with the backbone or other sidechains). Examples of the latter include Tyr7 in B35 and B53, Phe22 in A68, B35, B53, and Cw4, and Tyr84 in Cw4 and Cw6. In cases where some alleles had mutated sidechains but some did not, the identical sidechains either retained exactly the B27 dihedrals or minimized to similar rotamers. Examples of identical sidechain types which minimized to

similar conformations as in the B27 x-ray structure include Tyr99 in Cw6, Trp147 in Cw6 and Cw7, and Leu156 in B35, B53, and Cw7. Even sidechains which were mutated from the B27 amino acid types were found in similar conformations in the six predicted structure, whether minimized or not. In the six alleles, there are a total of 71 mutations in residues in the antigen binding site. In only 9 of these does the prediction differ in χ_1 by more than 40° from the B27 sidechain, viz. Lys66 in Cw4, Cw6, Cw7 (104° in Cw4, Cw6, Cw7 vs. -71° for Ile66 in B27), Asn70 in B35 and B53 (-93° vs. 174° for Lys70), Asp114 in Cw4 and Cw6 (-108°/110° vs. -174° for His114), Ser116 in Cw6 (-160° vs. -71° for Asp116), and Trp156 in A68 (-132° vs. -57° for Leu156). In five of these nine, the predicted conformation is more than 40° away from a standard rotamer and therefore should be considered as likely to be incorrectly placed. In sum, of the 181 sidechains placed in the antigen binding groove (31 positions x 6 models built - 5 mutations to alanine), only 9 or 5% differ by more than 40° in their χ_1 values from the B27 template.

(ii) *Descriptions of the antigen binding sites*

Sequences of eluted peptides, pooled eluted peptides, and viral and parasitic antigens known to bind to HLA-A2, A68, B27, B35, B53, Cw4, Cw6, Cw7 are summarized in Table 3. As discussed in the introduction, anchor residues are those which are common to most if not all peptides found to bind to particular alleles and also tend to bind in deep pockets in the antigen binding site.

The amino acid identities of residues found in the B, C, and F pockets for HLA-A2, A68, B27, B35, B53, Cw4, Cw6, Cw7 are listed in Table 4. This Table is similar to comparisons made by Matsumura et al (1992) for H-2K and HLA proteins. The information in Table 4 can be used to get a first glance at what the character of each pocket in the antigen binding site is likely to be.

The x-ray structure of HLA-B27 is shown from a number of perspectives in Figures 2a, 2b, and 3a-3f. In Figures 2a and 2b the Cα coordinates are shown, and those

residues with sidechains pointing into the site are labeled and numbered. The predicted structures of HLA-A68, B35, B53, Cw4, Cw6, and Cw7 are shown in Figures 4a-4f, 5a-5f, 6a-6f, 7a-7f, 8a-8f, and 9a-9f respectively. The first figure of each series is a stick figure of the Cα coordinates of the antigen binding site as seen from above with only those sidechains which point into the site as defined by Bjorkman et al. [14] (in bold type in Table 2). Positively charged sidechains are drawn in solid heavy lines, negatively charged sidechains in broken heavy lines, polar sidechains in thin solid lines, and hydrophobic sidechains in thin broken lines. The second figure of each series is drawn in the same fashion as the first, but the proteins have been rotated 90° away from the viewer so that the antigen binding site points toward the top of the page.

In Figure c of each series, van der Waals spheres for atoms in the site which are accessible to a 1.4 Å spherical probe are shown. Again, only atoms from the sidechains listed in bold type in Table 2 are shown. The same atoms are shown from the side in Figure d of each series (i.e. after the same 90° rotation as in Figures a and b). Figures e and f show those atoms in the site which are accessible to a 4.0 Å probe, again from the top and side views respectively. In each case, the top and side views allow us to observe the placement of charged and uncharged sidechains in three dimensions. The same code for drawing atoms from charged and uncharged sidechains used in Figures a and b of each series is also used in Figures c through f. Also, all of the figures are drawn with the same scale of 3.24 mm/Å. The seven proteins will be discussed individually, emphasizing the regions of the antigen sites that bind the common anchor residues at positions 2 and 9 in the bound peptides.

HLA-B27. From the x-ray data of Madden et al. [21], the arginine sidechain in position 2 of the peptide can be seen extending down into a pocket formed by the sidechains of His9, Thr24, Glu45, and Cys67. The Glu45 sidechain is responsible for the tendency to bind positively charged arginine at P2 by forming a salt-bridge with the

guanidinium group of Arg. This sidechain can be seen in Figure 3a at the top left of the antigen binding site (dashed heavy line), extending from a β-sheet strand below the α1 helix. In the figure it is just above Glu63 which is on the α1 helix extending further into the site. Near the bottom left of the site is the Glu163 sidechain. Figure 3b shows that the Glu45 sidechain is located towards the bottom of the site with the Glu63 and Glu163 sidechains located higher up in the site. Glu45 is accessible to a 1.4 Å probe, as shown in Figures 3c and 3d. In Figure 3d, one atom of Glu45 is partially visible at the bottom of the site, occluded by a tyrosine sidechain. Figures 3e and f show that the Glu sidechain is not accessible to a 4.0 Å sphere, indicating that the B '45' pocket narrows toward its terminus near residue 45.

At the other end of the site, there is a cluster of three Asp residues, Asp74, Asp77, and Asp116. Asp 74 and Asp77 are on the α1 helix (top of Figure 3a) and Asp 116 is on the lower α2 helix. The carboxylate of Asp116 is 2.8 Å from Lys Nζ at position 9 of the peptide [21]. Asp 116 is also the most deeply buried of the three as can be seen from Figure 3b, and is accessible to a 1.4 Å probe (Figures 3c and 3d). As noted by Madden et al., [21] these sidechains can hydrogen bond to other HLA-B27 sidechains, such that a positively charged P9 residue is not required in the peptide. This is in contrast to Glu45, which is deeply buried in the site and without the flexibility to reach another sidechain to form a salt-bridge.

The electron density of residue P3 extends into a non-polar site near Leu156 [20], which is located at the center of the α2 helix, extending into the site near His114 (i.e. Leu156 is represented by the thin dashed lines in the center of the lower helix just below a histidine residue). This site is lined by the sidechains of Tyr99 and Tyr159. Tyr99 lies across the bottom of the antigen binding site just to the left of His114. It is also visible in the center of Figure 3c. Tyr159 is on the α2 helix and is adjacent and perpendicular to Tyr99.

Madden et al. [20] found that density for the P7 sidechain was bifurcated, extending towards His114 near the α2 helix (bottom of Figure 3a) and Asn97 in the middle of the β-sheet strand just above His114. The polar but uncharged site is flexible enough to accommodate a number of different sidechain types [5] apparently interacting with different subsites within the pocket.

It is also worth noting sidechains which make contact with the bound peptide backbone [21]. These include the highly conserved Tyr7 and Tyr171 which hydrogen bond to the peptide N terminus at the left end of the pocket (Figure 3a). Tyr159 which is also highly conserved across HLA alleles forms a hydrogen bond with P1 backbone oxygen. Glu63 (below Glu45) and Arg62 (not shown in the site) make hydrogen bonds with P2 backbone N and C respectively. The Tyr99 OH contacts P3 backbone N, while the highly conserved Trp147 indole Nε1-H contacts the backbone O of P8. Trp147 is visible in the lower right end of the site in Figure 3a (dashed thin lines). The C-terminus of the peptide is bound by several water molecules as well as the OH of Tyr84 (at the very right end of the site in Figure 3a just above Lys146 in heavy lines) and the OH of Thr143 (just to the left of K146). Lys146 makes hydrogen bonds to Tyr84 and a water molecule that in turn makes a hydrogen bond to the C-terminus carboxylate. Residues Lys146, Thr143, and Tyr84 are all highly conserved.

HLA-A68. The x-ray structure of HLA-A68 has been discussed in detail by Guo et al. [6]. As noted in Section A above, the only residues in the site that are incorrectly predicted from B27 are Val67, Asp116, and Trp156. Val67 replaces a cysteine in B27. The rotation from a χ_1 value of -63° in the x-ray to χ_1 of 180° in the predicted structure has little effect on the shape of the B pocket, since there is still a methyl group in either case pointing into the pocket (Cγ1 in the x-ray and Cγ2 in the predicted structure; Figure 4a). Since His9, Thr24, and Glu45 in B27 have been replaced with Tyr, Ala, and Met in

A68 the B pocket is non-polar. It is also smaller and can only accommodate a Val or Thr P2 sidechain [19].

Asp116 hydrogen bonds to the Arg P9 sidechain of the Np91-99 peptide in the x-ray structure of Silver et al. [19]. The rotation of χ_1 by 120° from the A68 x-ray structure to the predicted structure (which has the same position for Asp116 as in B27) moves the carboxylate ion toward the F pocket. The consequences of this difference in Asp116 position between B27 and A68 are probably minor, since both proteins tend to bind positive charged sidechains such as Arg and Lys at the P9 position.

The incorrectly predicted Trp156 is in a sterically hindered conformation with χ_1 =-132°. The sidechain in this position is vertical in the site, stretching between the helices as can be seen in Figures 4a and 4b. This would block a peptide from binding and is clearly not consistent with peptide binding.

HLA-B35 and HLA-B53. The predicted structures of HLA-B35 and HLA-B53 are shown in Figures 5a through 6f. By comparing these figures with HLA-B27 (Figures 3a-3f), certain features of the B35 and B53 sites become evident. First, B35 and B53 have a Thr residue at position 45 so the pocket is no longer charged. B27 Glu63 which overhangs the site has been replaced with an Asn residue in B35 and B53. In Figures 5c and 6c, the atoms accessible to a 1.4 Å probe are shown from a view looking down into the site. By overlapping these figures with Figure 5a, atoms accessible to the 1.4 Å probe can be identified. Only a single atom of Asn63 and only the OH of Thr45 is accessible to the 1.4 Å probe. Tyr7 at the bottom of the site is almost entirely accessible to a 1.4 Å probe in B27, but is only partly accessible in B35 and B53. Cys67 in B27 has been replaced by a Phe residue in B35 and B53. The plane of the Phe67 ring is vertical and fairly low down in the site, filling the lower portion of the B pocket. Phe67 is not accessible to a 1.4 Å probe (Figure 5c). These changes are consistent with the fact that B35 and B53 tend to bind Pro residues at P2 while B27 binds an arginine.

At the PC or P9 end of the pocket, it can be seen in Figures 5a and 6a that B35 and B53 do not have the Asp74, Asp77, and Asp116 sidechains which characterize the charged P9 pocket of HLA-B27. Instead there is a serine at position 116 and a tyrosine at position 74 in both B35 and 53. B35 and B53 differ in the P9 end of the pocket at residues 77, 80, 81, 82, and 83. Only residues Ser77 and Asn80 in B35 and Asn77 and Ile80 in B53 extend into the site. The pockets are of different shapes, as can be seen by comparing Figures 5c and 6c. The Leu81 sidechain at the bottom of the PC pocket in B35 is accessible to a 1.4 Å probe, while the Ala81 sidechain is not. The Ile95 sidechain, which extends up from the β-sheet floor is also more accessible in B35 than it is in B53. The B35 pocket appears to be deeper or more accessible than the B53 pocket, perhaps explaining why B35 can accommodate a tyrosine sidechain at PC (either P8 or P9 position) while B53 prefers valine or threonine sidechains.

From the data of Hill et al. [1], B35 seems to *require* a tyrosine sidechain at the PC position (P8 or P9 depending on the length of the peptide). The reasons for this are not clear from the structures, but the PC subsite in the protein is large enough to accommodate the tyrosine sidechain. It is possible that smaller hydrophobic sidechains do not adequately fill the site, exposing a hydrophobic sidechain to solvent. The *P. falciparum* peptides which bind to HLA-B53 conferring resistance to liver-stage specific antigens have a valine at P9. Since B35 requires tyrosine, these peptides do not bind to B35. The smaller site in B53 can accommodate valine and bind these peptides.

HLA-Cw4. Instead of having an aromatic residue at position 9 in the B pocket, HLA-Cw4 has a serine. Glu45 in B27 has been replaced by a glycine residue, while Cys67 has been replaced by a tyrosine residue. As can be seen in Figure 7a and 7c, the sidechain of Glu63 forms a salt-bridge with the sidechain of Arg68, which is an alanine in B27. The result is a fairly deep but uncharged pocket that can accommodate a Phe or Tyr residue.

At the other end of the site, Leu81, Leu95, and Phe116 line the F pocket of Cw4. A nearby arginine residue, Arg79, is high up and extends away from the site (Figure 7b). The replacement of Asp116 with Phe results in a hydrophobic pocket large enough to accommodate Leu, Phe, Met or Ile residues which are commonly found in Cw4 eluted peptides [2]. This pocket can be observed in Figure 7c. At the entrance to the pocket are the positively charged residues Lys146 and Arg79, extending up from the site. But the atoms lining the bottom and sides of the site are entirely from hydrophobic residues, which are represented by thin broken lines. These atoms are not accessible to the 4.0 Å probe (Figure 7e), indicating that the pocket is fairly narrow towards its bottom. The site is similar in B27 except that Asp116 is located at the bottom and towards the left side of the pocket (Figure 3c, heavy broken lines), so the pocket is negatively charged, binding arginine and lysine sidechains.

Another significant change is the Arg97 sidechain in Cw4, which is an Asn residue in B27. This sidechain extends up from the floor in the center of the antigen binding site, and is accessible to a 1.4 Å probe. Since Cw4 tends to bind peptides with hydrophobic sidechains at P5 and P6, these sidechains may be placed away from Arg97 toward Asn114 to allow for solvent access to the Arg97 sidechain. P7 is sometimes a Glu residue in peptides bound to Cw4, and it is possible that Arg97 forms a salt-bridge with Glu P7.

HLA-Cw6. Cw6 differs from Cw4 at several key sites. In the B pocket, Cw6 has Asp9 rather than Ser9. Asp9 is deep within the pocket (Figure 8b) and is accessible to a 1.4 Å probe (Figure 8c). Arg is reproducibly detected at the P2 position in Cw6 but not in Cw4. A positively charged P2 sidechain is likely to form a salt-bridge with the accessible Asp9 residue. The pocket narrows towards its bottom, since many of the atoms at the bottom present in Figure 8c are absent in Figure 8e, which shows only atoms accessible to a 4.0 Å probe.

In the center of the site, Arg97 of Cw4 is a tryptophan in Cw6. Trp97 extends up from the β-sheet into the site (Figure 8a). It is accessible to a 1.4 Å probe on one edge of the ring (Figure 8c). In Cw6, there is an aspartic acid residue at position 114 rather than Asn in Cw4. Cw6 binds peptides with arginine and lysine at P7 (as well as Asn, Gln, and Tyr), and these sidechains may form a salt-bridge with Asp114. In Cw4, Glu was sometimes found at P7 possibly forming a salt-bridge with Arg97. These differing specificities are analogous to the bifurcated electron density seen in HLA-B27 for the P7 sidechain toward both positions 97 and 114 in B27. As noted by Matsumura et al. [24], these Class I residues may be in contact with either P6 or P7 sidechains, depending on the structure of the site.

At the PC end of the pocket, Cw4 and Cw6 are essentially identical, and bind similar sets of sidechains at PC.

HLA-Cw7. Cw7 has the same residues in the B pocket as Cw6, except for a substitution of serine at position 99 for tyrosine in Cw6 and B27. This is likely to make the pocket larger and deeper. Ser99 is accessible to the 1.4 Å probe (Figure 9c) as is Asp9. Tyrosine is a P2 anchor residue in Cw7 peptides but not in Cw6 peptides. Arginine is reproducibly found in peptides eluted from both proteins, apparently because of hydrogen bonding to the Asp9 sidechain deep inside the pocket.

Cw7 shares with Cw4 the Arg97 sidechain in the center of the site, again extending upwards underneath a bound peptide. Cw7 like Cw4 can bind peptides with a negatively charged P7 sidechain (Asp), which may form the salt-bridge with Arg97. Cw7 also binds peptides with hydrophobic P7 sidechains. Position 114 in Cw7 is an Asp residue. There is the possibility of a salt-bridge between Arg97 and Asp114, which may allow a more hydrophobic P7 sidechain to be placed in the C pocket.

At the PC end of the site, Cw7 has an Asn at position 80 rather than a lysine residue as in Cw4 and Cw6. This means that the opening of the F pocket is wider in

Cw7. On the other side near the α2 helix, Trp147 which forms a hydrogen bond with the P8 backbone O has been replaced with a leucine residue in Cw7. This is the only known HLA allele without Trp at position 147. The consequences of this are not clear, but it is likely that the PC end of the pocket is more open than in other alleles. Cw7 binds peptides with Tyr and Phe at P9, which could be easily accommodated in the wider pocket.

(iii) *Minimized bound peptides*

As described in the Methods, we modeled bound peptides into the site from the B27 peptide model of Madden et al. [21]. The energies of interaction of the minimized sidechain with the protein and the peptide itself are listed in Table 5 for all 10 sidechains modeled into each of the seven proteins.

For HLA-B27, the total minimized energies (column 6) show that an Arg residue at position 2 results in the lowest energy of the 10 sidechains tested. A very favorable electrostatic interaction between Arg and Glu45 makes the difference between the Arg interaction energy of -132 kcal/mol compared to Glu -17.8 kcal/mol. The large hydrophobic residues tested all caused large steric clashes that were relieved upon minimization but resulted in large bonded interactions in the peptide (bond stretch, bond angle deformation, dihedral angle deformation) (last column). Trp, for example, has a favorable van der Waals interaction with the protein (-13 kcal/mol) but a bonded energy of 172 kcal/mol.

At position 9 in HLA-B27, Arg is again the most favored residue. Arg and Lys are commonly found at P9 in HLA-B27 peptides. But other residues are also acceptable, including Asn, Leu, and Tyr. Trp has a large van der Waals and bonded interaction with the peptide and is the least favored residue of the 10 sidechains tested.

In A68, at P2 Ala, Arg, Asn, Leu, and Glu have favorable total interaction energies with the protein and peptide summed together. Arg and Asn however have very large

bonded energies of 22 and 13 kcal/mol respectively. Ile, Pro, and Trp have large unfavorable interactions within the complex. Table 3 indicates that A68 prefers Val and Thr at P2, which is contrary to the minimization results. It is possible that Val or Thr would fit into the site, while Ile is too large. It is still surprising that Glu fits. At P9, only Trp has a strongly unfavorable interaction energy, although Arg, Pro, Trp, and Tyr have large non-bonded energies. A68 can bind Arg, Lys, Ile, and His at P9.

B35 and B53 tend to bind Pro at P2, as measured experimentally by Hill et al. Pro has weakly unfavorable interaction energies, mostly due to bonded interactions in both B35 and B53. Both P2 sites seem to be fairly small, with Arg, Glu, Ile, Leu, Trp, and Tyr all having large unfavorable interaction energies with the peptide and protein. This would seem to indicate that only small sidechains, such as Pro and Ser may be able to fit into the site and still maintain interactions between the protein and the peptide backbone.

At the P9 site, B35 binds Tyr better than B53 does by 7 kcal/mol. In eluted peptides, B35 has Tyr as a P9 anchor while B53 does not have a strong anchor residue at P9. Most other sidechains appear to bind well in the site, except Trp. Arg has a large bonded energy in both proteins.

In the three HLA-C proteins, the P2 site has a shape that interferes with the backbone conformation derived from the B27 structure. In all three proteins, even Ala at P2 produces very large unfavorable interaction energies. In each case it is Lys66 of the protein that interferes with the sidechain conformation. Before minimization of the P2 sidechain position, Lys66 makes large steric contacts with Cβ of each P2 residue type. From Table 2, Lys66 is shown to be in an unlikely conformation with χ_1 equal to 104°. Without further minimization of the protein or peptide conformation, it is difficult to know whether this can be overcome with only minor adjustments in the protein model, such as the position of the Lys66 sidechain, or whether the peptide must also adjust significantly or both.

At P9, the model Cw4 can not tolerate Arg, Trp, or Tyr. Only Trp has a large energy in Cw6 and Cw7, although the bonded Arg energies are quite high. Apparently, in Cw6 and Cw7 large favorable electrostatic energies overcome the large unfavorable bonded/steric clashes of Arg with the protein. The Trp energy is much larger in Cw6 than in Cw7, which agrees with the structural analysis above that indicated that the Cw7 PC pocket was more open than that in Cw6 and Cw4. Cw7 has Tyr and Leu at P9 as anchor residues, while Cw4 has Leu, Met, and Phe and Cw6 has Leu, Ile, and Val.

IV. Discussion

In this paper, we have applied a more detailed method for homology modeling with the aid of the backbone-dependent rotamer library than described previously [44]. The method is completely automated, and takes only 2-4 hours on an SGI 340 computer. As such, it is highly suited to the modeling problem presented by the extensive polymorphism of HLA Class I and Class II proteins. With only three HLA protein structures determined experimentally, there are more than 100 other HLA proteins whose structures are not known. Many of these have associations with susceptibility to autoimmune diseases as well as resistance to certain infections agents. Modeling the structures of these HLA alleles is therefore of great interest

In this paper, the structures of five HLA proteins have been predicted, and compared with experimental sequence information for the peptides eluted from their antigen binding sites [1, 2]. Examination of the structures of HLA-B35, B53, Cw4, Cw6, and Cw7 and comparison with the known HLA-B27 structure has revealed subsites within the antigen binding groove which are likely to bind sidechains from bound peptides. At least in some cases, the character and size of the pockets within the antigen binding site has been correlated with the identity of sidechains from the eluted peptides. Modeling of the peptide within the site of these proteins has also enabled us to get a measure of the size of these pockets. In some cases, residues exhibiting reasonable

interaction energies with the protein and the rest of the peptide after a brief energy minimization were also found to be anchor residues in eluted peptides. Some residues with large steric clashes were not found at certain positions in eluted peptides. But there were enough conflicting examples to indicate a need for more careful modeling of the peptide within the site.

In the future, we plan to examine the antigen binding sites in further detail with the application of multiple copy simulated search methods as well as closer examination on molecular graphics systems. Model building of the peptide in the site will be particularly useful in attempting to correlate sequence data with the structure of the antigen binding site. In all of the examples studied here, the sequences of peptides bound to the site are known. The more difficult problem would be to examine the site of a modeled HLA structure, and attempt to predict the nature of sidechains of bound peptides. Such predictions could be tested by determining the sequences of eluted peptides. This is a long term goal, and is dependent on a more thorough analysis of the proteins studied in this paper. Nevertheless, the rewards in terms of understanding and possibly preventing certain illnesses are great and form the primary motivation for this study.

Acknowledgments

I would like to thank Dean Madden for providing the x-ray coordinates of HLA-B27 and HLA-A68 and Olaf Rötzschke for providing the peptide sequence data for the HLA-C proteins prior to publication.

References

1. A. V. S. Hill, J. Elvin, A. C. Willis, M. Aidoo, C. E. M. Allsopp, F. M. Gotch, X. M. Gao, M. Takaguchi, B. M. Greenwood, A. R. M. Townsend, A. J. McMichael, and H. C. Whittle. Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature*, **360**, 434-439 (1992).

2. O. Rötzschke, K. Falk, B. Grahovac, D. Schendel, S. Stevanovic, V. Gnau, G. Jung, J. L. Strominger, and H.-G. Rammensee. HLA-C molecules are peptide receptors with allele-specific peptide motifs. *In preparation*, (1993).

3. A. R. M. Townsend, J. Rothbard, F. M. Gotch, G. Bahadur, D. Wraith, and A. J. McMichael. The epitopes of influenza nucleoprotein recognized by cytotoxic T lymphocytes can be defined with short synthetic peptides. *Cell*, **44**, 959-968 (1986).

4. R. DeMars and T. Spies. New genes in the MHC that encode proteins for antigen processing. *Trends Cell Biol.*, **2**, 81-86 (1992).

5. T. S. Jardetzky, W. S. Lane, R. A. Robinson, D. R. Madden, and D. C. Wiley. Identification of self peptides bound to purified HLA-B27. *Nature*, **353**, 326-329 (1991).

6. H.-C. Guo, T. S. Jardetzky, T. P. J. Garrett, W. S. Lane, J. L. Strominger, and D. C. Wiley. Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle. *Nature*, **360**, 364-366 (1992).

7. R. M. Chicz, R. G. Urban, W. S. Lane, J. C. Gorga, L. J. Stern, D. A. A. Vignali, and J. L. Strominger. Predominant naturally processed peptides bound to HLA-DR1 are derived form MHC-related molecules and are heterogeneous in size. *Nature*, **358**, 764-768 (1992).

8. F. M. Brodsky and L. E. Guagliardi. The cell biology of antigen processing and presentation. *Annu. Rev. Immunol.*, **9**, 707-744 (1991).

9. M. M. Davis. T-cell receptor gene diversity and selection. *Annu. Rev. Biochem.*, **59**, 475-496 (1990).

10. A. Townsend and H. Bodmer. Antigen recognition by class I-restricted T lymphocytes. *Annu. Rev. Immunol.*, **7**, 601-624 (1989).

11. J. G. Bodmer, S. G. E. Marsh, E. D. Albert, W. F. Bodmer, B. Dupont, H. A. Erlich, B. Mach, W. R. Mayr, P. Parham, T. Sasazuki, G. M. T. Schreuder, J. L. Strominger, A. Svejgaard, and P. I. Terasaki. Nomenclature for factors of the HLA system 1991. *Tissue Antigens*, **39**, 161-173 (1992).

12. P. Parham, C. E. Lomen, D. A. Lawlor, J. P. Ways, N. Holmes, H. L. Coppin, R. D. Salter, A. M. Wan, and P. D. Ennis. Nature of polymorphism in HLA-A, -B, and -C molecules. *Proc. Natl. Acad. Sci. USA*, **85**, 4005-4009 (1988).

13. P. J. Bjorkman, M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature*, **329**, 506-512 (1987).

14. P. J. Bjorkman, M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*, **329**, 512-518 (1987).

15. M. A. Saper, P. J. Bjorkman, and D. C. Wiley. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol.*, **219**, 277-319 (1991).

16. K. Falk, O. Rötzschke, S. Stevanovic, G. Jung, and H.-G. Rammensee. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature*, **315**, 290-296 (1991).

17. D. F. Hunt, R. A. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A. L. Cox, E. Appella, and V. H. Engelhard. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science*, **255**, 1261-1263 (1992).

18. T. P. J. Garrett, M. A. Saper, P. J. Bjorkman, J. L. Strominger, and D. C. Wiley. Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature*, **342**, 692-696 (1989).

19. M. L. Silver, H.-C. Guo, J. L. Strominger, and D. C. Wiley. Atomic structure of a human MHC molecule presenting an influenza virus peptide. *Nature*, **360**, 367-369 (1992).

20. D. R. Madden, J. C. Gorga, J. L. Strominger, and D. C. Wiley. The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature*, **353**, 321-325 (1991).

21. D. R. Madden, J. C. Gorga, J. L. Strominger, and D. C. Wiley. The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC. *Cell*, **70**, 1035-1048 (1992).

22. D. H. Fremont, M. Matsumura, E. A. Stura, P. A. Peterson, and I. A. Wilson. Crystal structures of two viral peptides in complex with murine MHC Class I H-2K^{b}. *Science*, **257**, 919-926 (1992).

23. W. Zhang, A. C. M. Young, M. Imarai, S. G. Nathenson, and J. C. Sachettini. Crystal structure of the MHC class I H-2K^b molecule containing a single viral peptide: implications for peptide binding and T cell receptor recognition. *Proc. Natl. Acad. Sci USA*, **89**, 8403-8407 (1992).

24. M. Matsumura, D. H. Fremont, P. A. Peterson, and I. A. Wilson. Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science*, **257**, 927-934 (1992).

25. J. Green, M. Montasser, H. C. Low, and J. C. Woodrow. Investigation of the associations of a number of HLA antigens with psoriasis and psoriatic arthritis. *Stat. Med.*, **7**, 443 (1988).

26. M. A. Khan. HLA and ankylosing spondylitis. In *Ankylosing spondylitis: New clinical applications in rheumatology*. J. J. Calabro and C. Dick ed. (Lancaster, England: MTP Press, 1987).

27. J. Zemmour and P. Parham. Distinctive polymorphism at the HLA-C locus: Implications for the expression of HLA-C. *J. Exp. Med.*, **176**, 937-950 (1992).

28. H.-G. Ljunggren and K. Kaerre. In search of the 'missing self': MHC molecules and NK cell recognition. *Immunol. Today*, **11**, 237-244 (1990).

29. D. Snary, C. J. Barnstable, W. F. Bodmer, and M. J. Crumpton. Molecular structure of human histocompatibility antigens: the HLA-C series. *Eur. J. Immunol.*, **8**, 580 (1977).

30. R. Sodoyer, M. Damotte, T. L. Delovitch, J. Trucy, B. R. Jordan, and T. Strachan. Complete nucleotide sequence of a gene encoding a functional human class I histocompatibility antigen (HLA-Cw3). *EMBO J.*, **3**, 879 (1984).

31. D. Guessow, R. S. Rein, I. Meijer, W. Hoog, G. H. A. Seemann, F. M. Hochstenbach, and H. L. Ploegh. Isolation, expression, and the primary structure of HLA-Cw1 and HLA-Cw2 genes: evolutionary aspects. *Immunogenetics*, **25**, 313 (1987).

32. J. D'Amaro, J. J. van Rood, F. H. Bach, A. A. Rimm, and M. M. Bortin. HLA-C associations with acute leukaemia. *Lancet*, **2**, 1176-1178 (1984).

33. C. A. Mueller, R. Hasmann, H. Grosse-Wilde, U. Voegeler, C. Bei-Jun, R. Dopfer, and H. D. Waller. Significant association of acute lymphoblastic leukemia with HLA-Cw7. *Genet. Epidemiol.*, **5**, 453 (1988).

34. L. Groop, S. Koskimies, R. Pelkonen, and E.-M. Tolppanen. Increased frequency of HLA-Cw4 in type 2 diabetes. *Acta. Endocrinol.*, **104**, 475 (1983).

35. B. Malissen, T. Kristensen, C. Goridis, M. Madsen, and C. Mawas. Clones of human cytotoxic T lymphocytes derived from an allosensitized individual: HLA specificity and cell surface markers. *Scand. J. Immunol.*, **14**, 213 (1981).

36. M. Bonneville, J. F. Moreau, E. Blokland, J. Pool, J. P. Moisan, E. Goulmy, and J. P. Soulillou. T lymphocyte cloning from rejected human kidney allograft: Recognition repertoire of alloreactive T cell clones. *J. Immunol.*, **141**, 4187-4195 (1988).

37. D. J. Schendel, C. Reinhardt, P. J. Nelson, B. Maget, L. Pullen, G. W. Bornkamm, and A. Steinle. Cytotoxic T lymphocytes show HLA-C-restricted recognition of EBV-bearing cells and allorecognition of HLA Class I molecules presenting self-peptides. *J. Immunol.*, **149**, 2406-2414 (1992).

38. R. A. Littaua, M. B. A. Oldstone, A. Takeda, C. Debouck, J. T. Wong, C. U. Tuazon, B. Moss, F. Kievits, and F. A. Ennis. An HLA-C-restricted CD8+ cytotoxic T-lymphocyte clone recognizes a highly conserved epitope on human immunodeficienty virus type 1 gag. *J. Virol.*, **65**, 4051-4056 (1991).

39. O. Dill, F. Kievits, S. Koch, P. Ivanyi, and G. J. Haemmerling. Immunological function of HLA-C antigens in HLA-Cw3 transgenic mice. *Proc. Natl. Acad. Sci. USA*, **85**, 5664-5668 (1988).

40. B. P. Chen, V. Lam, E. E. Kraus, . DeMars, and P. M. Sondel. Restriction of Epstein-Barr virus-specific cytotoxic T cells by HLA-A, -B, and -C molecules. *Hum. Immunol.*, **26**, 137-147 (1989).

41. M. Colonna, T. Spies, J. L. Strominger, E. Ciccone, A. Moretta, L. Moretta, D. Pende, and O. Viale. Alloantigen recognition by two human natural killer cell clones is associated with HLA-C of a closely linked gene. *Proc. Natl. Acad. Sci. USA*, **89**, 7983-7985 (1992).

42. D. R. Madden and D. C. Wiley. Peptide binding to the major histocompatibility complex molecules. *Current Opinion in Structural Biology*, **2**, 300-304 (1992).

43. N. L. Summers and M. Karplus. Construction of sidechains in homology modeling. Application to the C-terminal lobe of rhizopuspepsin. *J. Mol. Biol.*, **210**, 785-812 (1989).

44. R. L. Dunbrack Jr. and M. Karplus. Backbone-dependent rotamer library for proteins: Application to sidechain prediction. *J. Mol. Biol.*, **320**, 543-574 (1993).

Table 1. HLA-B27 → HLA-A68 Prediction

					Structure 0			Structure 1			Structure 3		
Res #	B27	A68	χ	Xray A68	χ	Δχ	Cor?	χ	Δχ	Cor?	χ	Δχ	Cor?
2	SER	-	1	-54	-101	-46	n	-101	-46	n	-101	-46	n
3	HSD	-	1	-69	-81	-12	y	-81	-12	y	-81	-12	y
3	HSD	-	2	-81	-80	2	y	-80	2	y	-80	2	y
4	SER	-	1	76	86	10	y	86	10	y	86	10	y
5	**MET**	**-**	**1**	**-172**	**-171**	**1**	**y**	**-171**	**1**	**y**	**-171**	**1**	**y**
5	**MET**	**-**	**2**	**170**	**172**	**2**	**y**	**172**	**2**	**y**	**172**	**2**	**y**
5	**MET**	**-**	**3**	**51**	**58**	**8**	**y**	**58**	**8**	**y**	**58**	**8**	**y**
6	ARG	-	1	-71	-66	5	y	-66	5	y	-66	5	y
6	ARG	-	2	-169	-67	103	n	-67	103	n	-67	103	n
6	ARG	-	3	-138	-174	-36	y	-174	-36	y	-174	-36	y
6	ARG	-	4	144	-96	120	n	-96	120	n	-96	120	n
7	**TYR**	**-**	**1**	**-74**	**-77**	**-3**	**y**	**-77**	**-3**	**y**	**-62**	**12**	**y**
7	**TYR**	**-**	**2**	**94**	**103**	**9**	**y**	**103**	**9**	**y**	**94**	**0**	**y**
8	PHE	-	1	-68	-63	5	y	-63	5	y	-63	5	y
8	PHE	-	2	76	91	15	y	91	15	y	91	15	y
9	**HSD**	**TYR**	**1**	**-64**	**-42**	**22**	**y**	**-42**	**22**	**y**	**-59**	**5**	**y**
9	**HSD**	**TYR**	**2**	**171**	**81**	**90**	**n**	**81**	**90**	**n**	**-81**	**-72**	**n**
10	THR	-	1	-52	-62	-10	y	-62	-10	y	-62	-10	y
11	SER	-	1	-73	-72	2	y	-72	2	y	-73	0	y
12	VAL	-	1	-175	-175	0	y	-175	0	y	-175	0	y
13	SER	-	1	58	46	-12	y	46	-12	y	46	-12	y
14	ARG	-	1	-67	-58	9	y	-58	9	y	-58	9	y
14	ARG	-	2	-167	-171	-4	y	-171	-4	y	-171	-4	y
14	ARG	-	3	-60	-68	-8	y	-68	-8	y	-68	-8	y
14	ARG	-	4	-175	-174	2	y	-174	2	y	-174	2	y
15	PRO	-	1	-22	-17	5	y	-17	5	y	-17	5	y
15	PRO	-	2	41	36	-5	y	36	-5	y	36	-5	y
17	ARG	-	1	179	-70	111	n	-70	111	n	-70	111	n
17	ARG	-	2	171	-171	18	y	-171	18	y	-171	18	y
17	ARG	-	3	73	-59	-131	n	-59	-131	n	-59	-131	n
17	ARG	-	4	-128	117	-115	n	117	-115	n	117	-115	n
19	GLU	-	1	-45	63	109	n	63	109	n	63	109	n
19	GLU	-	2	-161	171	-28	y	171	-28	y	171	-28	y
19	GLU	-	3	164	95	111	n	95	111	n	95	111	n
20	PRO	-	1	0	-27	-27	y	-27	-27	y	-27	-27	y
20	PRO	-	2	3	36	33	y	36	33	y	36	33	y
21	ARG	-	1	175	171	-4	y	171	-4	y	171	-4	y
21	ARG	-	2	72	111	39	y	111	39	y	111	39	y
21	ARG	-	3	-176	176	-8	y	176	-8	y	176	-8	y
21	ARG	-	4	-70	-122	-52	n	-122	-52	n	-122	-52	n
22	**PHE**	**-**	**1**	**174**	**176**	**3**	**y**	**176**	**3**	**y**	**175**	**2**	**y**
22	**PHE**	**-**	**2**	**77**	**75**	**-2**	**y**	**75**	**-2**	**y**	**97**	**20**	**y**
23	ILE	-	1	-62	-64	-2	y	-64	-2	y	-64	-2	y
23	ILE	-	2	177	170	-6	y	170	-6	y	170	-6	y
25	VAL	-	1	-64	-63	0	y	-63	0	y	-63	0	y
27	TYR	-	1	-69	-76	-7	y	-76	-7	y	-76	-7	y
27	TYR	-	2	-88	-100	-12	y	-100	-12	y	-100	-12	y
28	VAL	-	1	168	169	1	y	169	1	y	169	1	y
29	ASP	-	1	-55	-60	-5	y	-60	-5	y	-60	-5	y
29	ASP	-	2	-43	-41	2	y	-41	2	y	-41	2	y
30	ASP	-	1	-64	-56	9	y	-56	9	y	-56	9	y
30	ASP	-	2	-38	-50	-12	y	-50	-12	y	-50	-12	y
31	THR	-	1	-56	-51	6	y	-51	6	y	-51	6	y
32	LEU	GLN	1	177	176	-1	y	176	-1	y	176	-1	y
32	LEU	GLN	2	177	-180	3	y	180	3	y	180	3	y
32	LEU	GLN	3	4	-180	-4	y	180	-4	y	180	-4	y
33	PHE	-	1	67	67	0	y	67	0	y	67	0	y
33	PHE	-	2	99	105	5	y	105	5	y	105	5	y

34	VAL	-	1	-76	-75	1	y	-75	1	y	-75	1	y
35	ARG	-	1	54	54	0	y	54	0	y	54	0	y
35	ARG	-	2	100	97	-3	y	97	-3	y	97	-3	y
35	ARG	-	3	75	71	-4	y	71	-4	y	71	-4	y
35	ARG	-	4	166	174	8	y	174	8	y	174	8	y
36	PHE	-	1	-180	175	-5	y	175	-5	y	-170	10	y
36	PHE	-	2	93	96	3	y	96	3	y	150	-123	n
37	ASP	-	1	-175	-173	1	y	-173	1	y	-173	1	y
37	ASP	-	2	164	-12	5	y	-12	5	y	-12	5	y
38	SER	-	1	64	66	2	y	66	2	y	66	2	y
39	ASP	-	1	-82	-174	-92	n	-174	-92	n	-174	-92	n
39	ASP	-	2	-167	74	61	n	74	61	n	74	61	n
42	SER	-	1	80	-40	-120	n	-40	-120	n	-40	-120	n
43	PRO	GLN	1	-73	-60	13	y	69	142	n	69	142	n
43	PRO	GLN	2	-56	-180	-124	n	-58	-2	y	-58	-2	y
43	PRO	GLN	3	-60	-180	60	n	116	-5	y	116	-5	y
44	ARG	-	1	-55	-77	-22	y	-77	-22	y	-77	-22	y
44	ARG	-	2	-79	-164	-85	n	-164	-85	n	-164	-85	n
44	ARG	-	3	-71	61	132	n	61	132	n	61	132	n
44	ARG	-	4	169	82	-87	n	82	-87	n	82	-87	n
45	GLU	MET	1	-175	180	-5	y	180	-5	y	-168	7	y
45	GLU	MET	2	71	160	89	n	160	89	n	178	108	n
45	GLU	MET	3	110	-180	70	n	-180	70	n	75	-35	y
46	GLU	-	1	-73	-68	5	y	-68	5	y	-68	5	y
46	GLU	-	2	173	165	-8	y	165	-8	y	165	-8	y
46	GLU	-	3	-127	32	-20	y	32	-20	y	32	-20	y
47	PRO	-	1	32	29	-3	y	29	-3	y	29	-3	y
47	PRO	-	2	-40	-39	1	y	-39	1	y	-39	1	y
48	ARG	-	1	-77	-73	5	y	-73	5	y	-73	5	y
48	ARG	-	2	-62	-62	0	y	-62	0	y	-62	0	y
48	ARG	-	3	177	-173	9	y	-173	9	y	-173	9	y
48	ARG	-	4	-93	-95	-2	y	-95	-2	y	-95	-2	y
50	PRO	-	1	-32	-21	12	y	-21	12	y	-21	12	y
50	PRO	-	2	50	37	-13	y	37	-13	y	37	-13	y
51	TRP	-	1	47	52	4	y	52	4	y	52	4	y
51	TRP	-	2	83	83	1	y	83	1	y	83	1	y
52	ILE	-	1	-158	-168	-9	y	-168	-9	y	-168	-9	y
52	ILE	-	2	68	69	1	y	69	1	y	69	1	y
53	GLU	-	1	-81	-57	24	y	-57	24	y	-57	24	y
53	GLU	-	2	171	-179	10	y	-179	10	y	-179	10	y
53	GLU	-	3	21	4	-17	y	4	-17	y	4	-17	y
54	GLN	-	1	58	60	2	y	60	2	y	60	2	y
54	GLN	-	2	-176	-171	5	y	-171	5	y	-171	5	y
54	GLN	-	3	103	-75	2	y	-75	2	y	-75	2	y
55	GLU	-	1	-66	-63	3	y	-63	3	y	-63	3	y
55	GLU	-	2	-63	-63	0	y	-63	0	y	-63	0	y
55	GLU	-	3	161	-18	1	y	-18	1	y	-18	1	y
57	PRO	-	1	-34	-23	11	y	-23	11	y	-23	11	y
57	PRO	-	2	50	37	-13	y	37	-13	y	37	-13	y
58	GLU	-	1	-158	65	-136	n	65	-136	n	65	-136	n
58	GLU	-	2	-115	-73	41	n	-73	41	n	-73	41	n
58	GLU	-	3	115	-16	50	n	-16	50	n	-16	50	n
59	**TYR**	**-**	**1**	**177**	**176**	**-1**	**y**	**176**	**-1**	**y**	**176**	**-1**	**y**
59	**TYR**	**-**	**2**	**93**	**90**	**-2**	**y**	**90**	**-2**	**y**	**81**	**-12**	**y**
60	TRP	-	1	-67	-64	3	y	-64	3	y	-64	3	y
60	TRP	-	2	-69	-67	2	y	-67	2	y	-67	2	y
61	ASP	-	1	-114	-71	43	n	-71	43	n	-71	43	n
61	ASP	-	2	-85	-21	-116	n	-21	-116	n	-21	-116	n
62	ARG	-	1	-176	179	-5	y	179	-5	y	179	-5	y
62	ARG	-	2	174	-163	23	y	-163	23	y	-163	23	y
62	ARG	-	3	73	-177	110	n	-177	110	n	-177	110	n
62	ARG	-	4	-119	-97	22	y	-97	22	y	-97	22	y
63	**GLU**	**ASN**	**1**	**-79**	**-59**	**20**	**y**	**-59**	**20**	**y**	**-55**	**24**	**y**
63	**GLU**	**ASN**	**2**	**-92**	**0**	**-88**	**n**	**0**	**-88**	**n**	**-42**	**-130**	**n**

64	THR	-	1	-60	-59	1	y	-59	1	y	-59	1	y
65	GLN	ARG	1	-76	-48	27	y	-48	27	y	-167	-91	n
65	GLN	ARG	2	-162	173	-24	y	173	-24	y	-62	100	n
65	GLN	ARG	3	-46	-180	-134	n	180	-134	n	-50	-3	y
65	GLN	ARG	4	-85	-180	-95	n	180	-95	n	-168	-83	n
66	**ILE**	**ASN**	**1**	**-73**	**-60**	**13**	**y**	**-60**	**13**	**y**	**-60**	**13**	**y**
66	**ILE**	**ASN**	**2**	**-21**	**0**	**21**	**y**	**0**	**21**	**y**	**0**	**21**	**y**
67	**CYS**	**VAL**	**1**	**-63**	**-180**	**-117**	**n**	**-180**	**-117**	**n**	**-180**	**-117**	**n**
68	LYS	-	1	-75	-57	18	y	-57	18	y	-57	18	y
68	LYS	-	2	-179	-168	11	y	-168	11	y	-168	11	y
68	LYS	-	3	138	-159	64	n	-159	64	n	-159	64	n
68	LYS	-	4	-170	-60	110	n	-60	110	n	-60	110	n
70	**LYS**	**GLN**	**1**	**-176**	**174**	**-10**	**y**	**174**	**-10**	**y**	**174**	**-10**	**y**
70	**LYS**	**GLN**	**2**	**66**	**96**	**30**	**y**	**96**	**30**	**y**	**96**	**30**	**y**
70	**LYS**	**GLN**	**3**	**54**	**-180**	**-54**	**n**	**180**	**-54**	**n**	**180**	**-54**	**n**
71	ALA	SER	1	64	-60	-124	n	-60	-124	n	74	10	y
72	GLN	-	1	-62	180	-118	n	180	-118	n	180	-118	n
72	GLN	-	2	-54	53	107	n	53	107	n	53	107	n
72	GLN	-	3	-54	28	-98	n	28	-98	n	28	-98	n
73	**THR**	**-**	**1**	**-65**	**-51**	**14**	**y**	**-51**	**14**	**y**	**-51**	**14**	**y**
74	**ASP**	**-**	**1**	**-76**	**-80**	**-5**	**y**	**-80**	**-5**	**y**	**-80**	**-5**	**y**
74	**ASP**	**-**	**2**	**-172**	**-1**	**-10**	**y**	**-1**	**-10**	**y**	**-1**	**-10**	**y**
75	ARG	-	1	178	-66	116	n	-66	116	n	-66	116	n
75	ARG	-	2	-166	-173	-8	y	-173	-8	y	-173	-8	y
75	ARG	-	3	51	-166	143	n	-166	143	n	-166	143	n
75	ARG	-	4	77	106	29	y	106	29	y	106	29	y
76	GLU	VAL	1	169	-180	11	y	180	11	y	180	11	y
77	**ASP**	**-**	**1**	**-71**	**-74**	**-3**	**y**	**-74**	**-3**	**y**	**-74**	**-3**	**y**
77	**ASP**	**-**	**2**	**168**	**-16**	**-4**	**y**	**-16**	**-4**	**y**	**-16**	**-4**	**y**
78	LEU	-	1	-80	-78	2	y	-78	2	y	-78	2	y
78	LEU	-	2	-76	-69	8	y	-69	8	y	-69	8	y
80	**THR**	**-**	**1**	**-61**	**-58**	**2**	**y**	**-58**	**2**	**y**	**-58**	**2**	**y**
81	**LEU**	**-**	**1**	**-86**	**-98**	**-12**	**y**	**-98**	**-12**	**y**	**-98**	**-12**	**y**
81	**LEU**	**-**	**2**	**176**	**164**	**-12**	**y**	**164**	**-12**	**y**	**164**	**-12**	**y**
82	LEU	ARG	1	-173	-67	106	n	-67	106	n	-67	106	n
82	LEU	ARG	2	-151	-180	-29	y	-180	-29	y	-180	-29	y
82	LEU	ARG	3	58	-180	122	n	-180	122	n	-180	122	n
82	LEU	ARG	4	60	-180	120	n	-180	120	n	-180	120	n
84	**TYR**	**-**	**1**	**-66**	**-68**	**-2**	**y**	**-68**	**-2**	**y**	**-68**	**-2**	**y**
84	**TYR**	**-**	**2**	**-17**	**179**	**16**	**y**	**179**	**16**	**y**	**179**	**16**	**y**
85	TYR	-	1	-71	-68	3	y	-68	3	y	-68	3	y
85	TYR	-	2	103	102	-1	y	102	-1	y	102	-1	y
86	ASN	-	1	-175	-156	19	y	-156	19	y	-156	19	y
86	ASN	-	2	65	-1	114	n	-1	114	n	-1	114	n
87	GLN	-	1	-69	-64	6	y	-64	6	y	-64	6	y
87	GLN	-	2	163	167	4	y	167	4	y	167	4	y
87	GLN	-	3	-36	-27	9	y	-27	9	y	-27	9	y
88	SER	-	1	161	164	3	y	164	3	y	164	3	y
89	GLU	-	1	-62	-73	-10	y	-73	-10	y	-73	-10	y
89	GLU	-	2	66	-137	157	n	-137	157	n	-137	157	n
89	GLU	-	3	-109	-51	-123	n	-51	-123	n	-51	-123	n
92	SER	-	1	175	168	-7	y	168	-7	y	168	-7	y
93	HSD	-	1	-63	-72	-9	y	-72	-9	y	-72	-9	y
93	HSD	-	2	93	-82	6	y	-82	6	y	-82	6	y
94	THR	-	1	-51	-56	-5	y	-56	-5	y	-56	-5	y
95	**LEU**	**ILE**	**1**	**-74**	**-60**	**14**	**y**	**-60**	**14**	**y**	**-60**	**14**	**y**
95	**LEU**	**ILE**	**2**	**178**	**-180**	**2**	**y**	**-180**	**2**	**y**	**-180**	**2**	**y**
96	GLN	-	1	-69	-75	-6	y	-75	-6	y	-75	-6	y
96	GLN	-	2	162	160	-2	y	160	-2	y	160	-2	y
96	GLN	-	3	-95	-87	8	y	-87	8	y	-87	8	y
97	**ASN**	**MET**	**1**	**170**	**168**	**-2**	**y**	**168**	**-2**	**y**	**168**	**-2**	**y**
97	**ASN**	**MET**	**2**	**158**	**-180**	**22**	**y**	**180**	**22**	**y**	**180**	**22**	**y**
97	**ASN**	**MET**	**3**	**70**	**-180**	**110**	**n**	**-180**	**110**	**n**	**-180**	**110**	**n**
98	MET	-	1	-172	164	-24	y	164	-24	y	164	-24	y

98	MET	-	2	-85	-72	13	y	-72	13	y	-72	13	y
98	MET	-	3	-77	-107	-31	y	-107	-31	y	-107	-31	y
99	**TYR**	**-**	**1**	**67**	**66**	**0**	**y**	**66**	**0**	**y**	**66**	**0**	**y**
99	**TYR**	**-**	**2**	**-98**	**-99**	**-1**	**y**	**-99**	**-1**	**y**	**-99**	**-1**	**y**
101	CYS	-	1	57	51	-6	y	51	-6	y	51	-6	y
102	ASP	-	1	-55	-51	4	y	-51	4	y	-51	4	y
102	ASP	-	2	156	149	-7	y	149	-7	y	149	-7	y
103	VAL	-	1	-68	-53	15	y	-53	15	y	-53	15	y
105	PRO	SER	1	-73	60	133	n	60	133	n	60	133	n
106	ASP	-	1	55	50	-5	y	50	-5	y	50	-5	y
106	ASP	-	2	1	8	7	y	8	7	y	8	7	y
108	ARG	-	1	-81	-68	14	y	-68	14	y	-68	14	y
108	ARG	-	2	168	-164	28	y	-164	28	y	-164	28	y
108	ARG	-	3	48	162	114	n	162	114	n	162	114	n
108	ARG	-	4	-107	-176	-69	n	-176	-69	n	-176	-69	n
109	LEU	PHE	1	174	-174	12	y	44	-130	n	44	-130	n
109	LEU	PHE	2	-98	90	8	y	-90	7	y	-90	7	y
110	LEU	-	1	-119	-116	3	y	-116	3	y	-116	3	y
110	LEU	-	2	153	-62	144	n	-62	144	n	-62	144	n
111	ARG	-	1	68	177	110	n	177	110	n	177	110	n
111	ARG	-	2	-176	77	-107	n	77	-107	n	77	-107	n
111	ARG	-	3	73	-172	115	n	-172	115	n	-172	115	n
111	ARG	-	4	70	-94	-164	n	-94	-164	n	-94	-164	n
113	TYR	-	1	20	-62	-83	n	-62	-83	n	-62	-83	n
113	TYR	-	2	92	95	3	y	95	3	y	95	3	y
114	**HSD**	**ARG**	**1**	**169**	**175**	**6**	**y**	**175**	**6**	**y**	**175**	**6**	**y**
114	**HSD**	**ARG**	**2**	**-112**	**-180**	**-68**	**n**	**-180**	**-68**	**n**	**-180**	**-68**	**n**
114	**HSD**	**ARG**	**3**	**167**	**-180**	**13**	**y**	**-180**	**13**	**y**	**-180**	**13**	**y**
114	**HSD**	**ARG**	**4**	**-93**	**-180**	**-87**	**n**	**-180**	**-87**	**n**	**-180**	**-87**	**n**
115	GLN	-	1	66	47	-19	y	47	-19	y	47	-19	y
115	GLN	-	2	-109	-155	-46	n	-155	-46	n	-155	-46	n
115	GLN	-	3	-86	50	-44	n	50	-44	n	50	-44	n
116	**ASP**	**-**	**1**	**65**	**-71**	**-136**	**n**	**-71**	**-136**	**n**	**-71**	**-136**	**n**
116	**ASP**	**-**	**2**	**-150**	**86**	**56**	**n**	**86**	**56**	**n**	**86**	**56**	**n**
118	TYR	-	1	-180	-175	5	y	-175	5	y	-175	5	y
118	TYR	-	2	63	50	-13	y	50	-13	y	50	-13	y
119	ASP	-	1	-74	-69	4	y	-69	4	y	-69	4	y
119	ASP	-	2	154	-28	-2	y	-28	-2	y	-28	-2	y
121	LYS	-	1	-62	-53	9	y	-53	9	y	-53	9	y
121	LYS	-	2	-146	151	-63	n	151	-63	n	151	-63	n
121	LYS	-	3	-63	-164	-101	n	-164	-101	n	-164	-101	n
121	LYS	-	4	-63	-84	-21	y	-84	-21	y	-84	-21	y
122	ASP	-	1	-62	-76	-14	y	-76	-14	y	-76	-14	y
122	ASP	-	2	-22	179	21	y	179	21	y	179	21	y
123	TYR	-	1	-179	-174	4	y	-174	4	y	-174	4	y
123	TYR	-	2	-99	-103	-4	y	-103	-4	y	-103	-4	y
124	ILE	-	1	-166	-169	-3	y	-169	-3	y	-169	-3	y
124	ILE	-	2	169	-169	22	y	-169	22	y	-169	22	y
126	LEU	-	1	-177	-170	7	y	-170	7	y	-170	7	y
126	LEU	-	2	-173	-175	-3	y	-175	-3	y	-175	-3	y
127	ASN	LYS	1	-73	-65	8	y	-65	8	y	-155	-83	n
127	ASN	LYS	2	168	-180	12	y	180	12	y	-173	19	y
127	ASN	LYS	3	-174	-180	-6	y	180	-6	y	82	-104	n
127	ASN	LYS	4	-99	-180	-81	n	-180	-81	n	-132	-33	y
128	GLU	-	1	-80	177	-103	n	177	-103	n	177	-103	n
128	GLU	-	2	51	-73	-124	n	-73	-124	n	-73	-124	n
128	GLU	-	3	49	-27	103	n	-27	103	n	-27	103	n
129	ASP	-	1	60	57	-4	y	57	-4	y	57	-4	y
129	ASP	-	2	-1	7	8	y	7	8	y	7	8	y
130	LEU	-	1	-54	-49	4	y	-49	4	y	-49	4	y
130	LEU	-	2	-56	-51	6	y	-51	6	y	-51	6	y
131	SER	ARG	1	-72	-59	14	y	-59	14	y	172	-115	n
131	SER	ARG	2	174	-180	6	y	-180	6	y	-79	108	n
131	SER	ARG	3	-174	-180	-6	y	-180	-6	y	-48	126	n

131	SER	ARG	4	-91	-180	-89	n	-180	-89	n	-120	-30	y
132	SER	-	1	-9	55	65	n	55	65	n	52	61	n
133	TRP	-	1	-61	-63	-2	y	-63	-2	y	-63	-2	y
133	TRP	-	2	-78	-84	-6	y	-84	-6	y	-84	-6	y
134	THR	-	1	-57	-66	-9	y	-66	-9	y	-66	-9	y
137	ASP	-	1	58	54	-4	y	54	-4	y	54	-4	y
137	ASP	-	2	26	25	-1	y	25	-1	y	25	-1	y
138	THR	MET	1	-62	-60	2	y	-60	2	y	-60	2	y
138	THR	MET	2	-50	-180	-130	n	180	-130	n	180	-130	n
138	THR	MET	3	-71	-180	-109	n	180	-109	n	180	-109	n
141	GLN	-	1	-170	-80	91	n	-80	91	n	-80	91	n
141	GLN	-	2	43	171	128	n	171	128	n	171	128	n
141	GLN	-	3	-119	-38	-99	n	-38	-99	n	-38	-99	n
142	ILE	THR	1	-55	-64	-9	y	-64	-9	y	-64	-9	y
143	**THR**	**-**	**1**	**-57**	**-61**	**-4**	**y**	**-61**	**-4**	**y**	**-61**	**-4**	**y**
144	GLN	LYS	1	175	-173	13	y	-173	13	y	-173	13	y
144	GLN	LYS	2	172	177	6	y	177	6	y	177	6	y
144	GLN	LYS	3	170	-180	10	y	-180	10	y	-180	10	y
144	GLN	LYS	4	174	-180	6	y	-180	6	y	-180	6	y
145	ARG	HSD	1	-80	-76	4	y	-76	4	y	-76	4	y
145	ARG	HSD	2	-35	90	-55	n	90	-55	n	90	-55	n
146	**LYS**	**-**	**1**	**-50**	**-67**	**-17**	**y**	**-67**	**-17**	**y**	**-67**	**-17**	**y**
146	**LYS**	**-**	**2**	**-176**	**-176**	**0**	**y**	**-176**	**0**	**y**	**-176**	**0**	**y**
146	**LYS**	**-**	**3**	**168**	**177**	**10**	**y**	**177**	**10**	**y**	**177**	**10**	**y**
146	**LYS**	**-**	**4**	**-178**	**-179**	**-1**	**y**	**-179**	**-1**	**y**	**-179**	**-1**	**y**
147	**TRP**	**-**	**1**	**-73**	**-77**	**-4**	**y**	**-77**	**-4**	**y**	**-77**	**-4**	**y**
147	**TRP**	**-**	**2**	**-173**	**-171**	**2**	**y**	**-171**	**2**	**y**	**-171**	**2**	**y**
148	GLU	-	1	-74	-61	13	y	-61	13	y	-73	1	y
148	GLU	-	2	171	177	6	y	177	6	y	171	0	y
148	GLU	-	3	-40	-35	4	y	-35	4	y	-5	34	y
151	ARG	HSD	1	-44	-47	-3	y	-47	-3	y	-47	-3	y
151	ARG	HSD	2	-46	90	-44	n	90	-44	n	90	-44	n
152	**VAL**	**-**	**1**	**165**	**172**	**7**	**y**	**172**	**7**	**y**	**172**	**7**	**y**
154	GLU	-	1	-81	-79	2	y	-79	2	y	-79	2	y
154	GLU	-	2	155	174	19	y	174	19	y	174	19	y
154	GLU	-	3	44	-107	29	y	-107	29	y	-107	29	y
155	**GLN**	**-**	**1**	**-75**	**-70**	**5**	**y**	**-70**	**5**	**y**	**-70**	**5**	**y**
155	**GLN**	**-**	**2**	**167**	**174**	**6**	**y**	**174**	**6**	**y**	**174**	**6**	**y**
155	**GLN**	**-**	**3**	**87**	**-83**	**10**	**y**	**-83**	**10**	**y**	**-83**	**10**	**y**
156	**LEU**	**TRP**	**1**	**-87**	**-57**	**30**	**y**	**-132**	**-45**	**n**	**-132**	**-45**	**n**
156	**LEU**	**TRP**	**2**	**-69**	**90**	**159**	**n**	**57**	**125**	**n**	**57**	**125**	**n**
157	ARG	-	1	-176	-179	-3	y	-179	-3	y	-179	-3	y
157	ARG	-	2	174	161	-12	y	161	-12	y	161	-12	y
157	ARG	-	3	166	76	-90	n	76	-90	n	76	-90	n
157	ARG	-	4	-138	-105	32	y	-105	32	y	-105	32	y
159	**TYR**	**-**	**1**	**169**	**167**	**-2**	**y**	**167**	**-2**	**y**	**167**	**-2**	**y**
159	**TYR**	**-**	**2**	**79**	**74**	**-5**	**y**	**74**	**-5**	**y**	**74**	**-5**	**y**
160	LEU	-	1	-67	-65	2	y	-65	2	y	-65	2	y
160	LEU	-	2	-76	-73	3	y	-73	3	y	-73	3	y
161	GLU	-	1	-69	-68	1	y	-68	1	y	-68	1	y
161	GLU	-	2	-171	174	-15	y	174	-15	y	174	-15	y
161	GLU	-	3	43	-16	121	n	-16	121	n	-16	121	n
163	**GLU**	**THR**	**1**	**-38**	**-60**	**-22**	**y**	**-60**	**-22**	**y**	**-60**	**-22**	**y**
164	CYS	-	1	-175	-163	12	y	-169	6	y	-168	6	y
165	VAL	-	1	168	-176	16	y	-176	16	y	-176	16	y
166	GLU	-	1	-72	-70	2	y	-70	2	y	-70	2	y
166	GLU	-	2	165	153	-12	y	153	-12	y	153	-12	y
166	GLU	-	3	3	28	26	y	28	26	y	28	26	y
167	**TRP**	**-**	**1**	**-79**	**-80**	**-1**	**y**	**-80**	**-1**	**y**	**-80**	**-1**	**y**
167	**TRP**	**-**	**2**	**-92**	**-98**	**-5**	**y**	**-98**	**-5**	**y**	**-98**	**-5**	**y**
168	LEU	-	1	167	173	6	y	173	6	y	173	6	y
168	LEU	-	2	176	179	3	y	179	3	y	179	3	y
169	ARG	-	1	-68	-62	7	y	-62	7	y	-62	7	y
169	ARG	-	2	177	-167	15	y	-167	15	y	-167	15	y

169	ARG	-	3	-159	-48	111	n	-48	111	n	-48	111	n
169	ARG	-	4	130	-122	108	n	-122	108	n	-122	108	n
170	ARG	-	1	165	175	10	y	175	10	y	175	10	y
170	ARG	-	2	-162	-170	-7	y	-170	-7	y	-170	-7	y
170	ARG	-	3	-177	177	-6	y	177	-6	y	177	-6	y
170	ARG	-	4	81	89	8	y	89	8	y	89	8	y
171	**TYR**	**-**	**1**	**-60**	**-64**	**-3**	**y**	**-64**	**-3**	**y**	**-64**	**-3**	**y**
171	**TYR**	**-**	**2**	**148**	**-27**	**5**	**y**	**-27**	**5**	**y**	**-27**	**5**	**y**
172	LEU	-	1	-71	-69	2	y	-69	2	y	-69	2	y
172	LEU	-	2	-65	-63	2	y	-63	2	y	-63	2	y
173	GLU	-	1	-169	-175	-6	y	-175	-6	y	-175	-6	y
173	GLU	-	2	43	179	135	n	179	135	n	179	135	n
173	GLU	-	3	-147	-85	-117	n	-85	-117	n	-85	-117	n
174	ASN	-	1	-71	-68	3	y	-68	3	y	-68	3	y
174	ASN	-	2	170	-26	-16	y	-26	-16	y	-26	-16	y
176	LYS	-	1	71	-168	121	n	-168	121	n	-168	121	n
176	LYS	-	2	-176	-177	-2	y	-177	-2	y	-177	-2	y
176	LYS	-	3	150	-174	36	y	-174	36	y	-174	36	y
176	LYS	-	4	-173	179	-8	y	179	-8	y	179	-8	y
177	GLU	-	1	-99	-51	48	n	-51	48	n	-51	48	n
177	GLU	-	2	45	-71	-117	n	-71	-117	n	-71	-117	n
177	GLU	-	3	75	62	-13	y	62	-13	y	62	-13	y
178	THR	-	1	46	51	4	y	51	4	y	51	4	y
179	LEU	-	1	-66	-63	3	y	-63	3	y	-63	3	y
179	LEU	-	2	-59	-63	-5	y	-63	-5	y	-63	-5	y
180	GLN	-	1	-62	-68	-6	y	-68	-6	y	-68	-6	y
180	GLN	-	2	-167	-170	-3	y	-170	-3	y	-170	-3	y
180	GLN	-	3	-130	63	13	y	63	13	y	63	13	y
181	ARG	-	1	-176	-93	83	n	-93	83	n	-93	83	n
181	ARG	-	2	-165	69	-125	n	69	-125	n	69	-125	n
181	ARG	-	3	85	180	95	n	180	95	n	180	95	n
181	ARG	-	4	-160	151	-49	n	151	-49	n	151	-49	n
182	ALA	THR	1	49	-60	-109	n	-60	-109	n	-60	-109	n
183	ASP	-	1	-63	-141	-78	n	-141	-78	n	-141	-78	n
183	ASP	-	2	-176	-12	-16	y	-12	-16	y	-12	-16	y
185	PRO	-	1	22	32	11	y	32	11	y	32	11	y
185	PRO	-	2	-30	-33	-3	y	-33	-3	y	-33	-3	y
186	LYS	-	1	-64	-74	-11	y	-74	-11	y	-74	-11	y
186	LYS	-	2	-174	179	-6	y	179	-6	y	179	-6	y
186	LYS	-	3	-174	-114	59	n	-114	59	n	-114	59	n
186	LYS	-	4	174	163	-11	y	163	-11	y	163	-11	y
187	THR	-	1	62	55	-7	y	55	-7	y	55	-7	y
188	HSD	-	1	71	78	6	y	78	6	y	78	6	y
188	HSD	-	2	-80	-80	0	y	-80	0	y	-80	0	y
189	VAL	MET	1	-160	-60	100	n	-124	35	y	-124	35	y
189	VAL	MET	2	-169	-180	-11	y	108	-83	n	108	-83	n
189	VAL	MET	3	-72	-180	-108	n	-62	9	y	-62	9	y
190	THR	-	1	59	55	-5	y	55	-5	y	55	-5	y
191	HSD	-	1	-180	173	-7	y	173	-7	y	173	-7	y
191	HSD	-	2	-66	-88	-21	y	-88	-21	y	-88	-21	y
192	HSD	-	1	-56	-56	0	y	-56	0	y	-56	0	y
192	HSD	-	2	-78	-73	5	y	-73	5	y	-73	5	y
194	ILE	VAL	1	173	168	-5	y	168	-5	y	168	-5	y
195	SER	-	1	68	59	-9	y	59	-9	y	59	-9	y
196	ASP	-	1	-178	-73	104	n	-73	104	n	-73	104	n
196	ASP	-	2	-32	-42	-11	y	-42	-11	y	-42	-11	y
197	HSD	-	1	71	-62	-133	n	-62	-133	n	-62	-133	n
197	HSD	-	2	103	94	-9	y	94	-9	y	94	-9	y
198	GLU	-	1	-90	-38	52	n	-38	52	n	-38	52	n
198	GLU	-	2	108	-180	72	n	-180	72	n	-180	72	n
198	GLU	-	3	-138	-78	-120	n	-78	-120	n	-78	-120	n
200	THR	-	1	-62	-48	13	y	-48	13	y	-48	13	y
201	LEU	-	1	-56	-69	-13	y	-69	-13	y	-69	-13	y
201	LEU	-	2	-65	-74	-9	y	-74	-9	y	-74	-9	y

202	ARG	-	1	-179	-176	3	y	-176	3	y	-176	3	y
202	ARG	-	2	161	165	4	y	165	4	y	165	4	y
202	ARG	-	3	-72	-80	-8	y	-80	-8	y	-80	-8	y
202	ARG	-	4	115	123	8	y	123	8	y	123	8	y
203	CYS	-	1	-175	-170	4	y	-179	-5	y	-180	-5	y
204	TRP	-	1	-67	-59	8	y	-59	8	y	-59	8	y
204	TRP	-	2	-96	-104	-8	y	-104	-8	y	-104	-8	y
206	LEU	-	1	-62	-57	5	y	-57	5	y	-57	5	y
206	LEU	-	2	-72	-69	3	y	-69	3	y	-69	3	y
207	GLY	SER	1	-150	-60	90	n	-60	90	n	-60	90	n
208	PHE	-	1	53	55	1	y	55	1	y	55	1	y
208	PHE	-	2	-87	94	1	y	94	1	y	94	1	y
209	TYR	-	1	-175	-179	-4	y	-179	-4	y	-179	-4	y
209	TYR	-	2	-84	83	-14	y	83	-14	y	83	-14	y
210	PRO	-	1	36	37	1	y	37	1	y	37	1	y
210	PRO	-	2	-45	-41	4	y	-41	4	y	-41	4	y
212	GLU	-	1	177	-178	5	y	-178	5	y	-178	5	y
212	GLU	-	2	50	45	-4	y	45	-4	y	45	-4	y
212	GLU	-	3	32	20	-11	y	20	-11	y	20	-11	y
213	ILE	-	1	-162	-161	1	y	-161	1	y	-161	1	y
213	ILE	-	2	158	162	4	y	162	4	y	162	4	y
214	THR	-	1	-58	-66	-8	y	-66	-8	y	-66	-8	y
215	LEU	-	1	-112	-67	45	n	-67	45	n	-83	29	y
215	LEU	-	2	169	-61	130	n	-61	130	n	-146	45	n
216	THR	-	1	55	-11	-66	n	-11	-66	n	-11	-66	n
217	TRP	-	1	-81	-95	-13	y	-95	-13	y	-88	-7	y
217	TRP	-	2	-72	-67	5	y	-67	5	y	-74	-2	y
218	GLN	-	1	-75	-74	1	y	-74	1	y	-74	1	y
218	GLN	-	2	177	178	1	y	178	1	y	178	1	y
218	GLN	-	3	20	-6	-26	y	-6	-26	y	-6	-26	y
219	ARG	-	1	-168	-177	-10	y	-177	-10	y	-177	-10	y
219	ARG	-	2	-86	176	-98	n	176	-98	n	176	-98	n
219	ARG	-	3	-88	43	131	n	43	131	n	43	131	n
219	ARG	-	4	-179	-161	18	y	-161	18	y	-161	18	y
220	ASP	-	1	-71	-64	7	y	-64	7	y	-64	7	y
220	ASP	-	2	152	-41	-13	y	-41	-13	y	-41	-13	y
222	GLU	-	1	170	-166	24	y	-166	24	y	-166	24	y
222	GLU	-	2	156	106	-50	n	106	-50	n	106	-50	n
222	GLU	-	3	43	-58	79	n	-58	79	n	-58	79	n
223	ASP	-	1	-83	-85	-2	y	-85	-2	y	-85	-2	y
223	ASP	-	2	-178	-7	-9	y	-7	-9	y	-7	-9	y
224	GLN	-	1	177	-65	118	n	-65	118	n	-65	118	n
224	GLN	-	2	168	-51	141	n	-51	141	n	-51	141	n
224	GLN	-	3	-34	-60	-26	y	-60	-26	y	-60	-26	y
225	THR	-	1	74	-39	-112	n	-39	-112	n	-39	-112	n
226	GLN	-	1	-72	-49	23	y	-49	23	y	-49	23	y
226	GLN	-	2	-74	-58	16	y	-58	16	y	-58	16	y
226	GLN	-	3	96	12	97	n	12	97	n	12	97	n
227	ASP	-	1	-54	-63	-10	y	-63	-10	y	-63	-10	y
227	ASP	-	2	-65	-68	-4	y	-68	-4	y	-68	-4	y
228	THR	-	1	61	-54	-116	n	-54	-116	n	48	-13	y
229	GLU	-	1	-179	170	-10	y	170	-10	y	170	-10	y
229	GLU	-	2	174	-176	10	y	-176	10	y	-176	10	y
229	GLU	-	3	-175	37	32	y	37	32	y	37	32	y
230	LEU	-	1	-177	-178	-1	y	-178	-1	y	-180	-3	y
230	LEU	-	2	-170	-163	7	y	-163	7	y	-97	73	n
231	VAL	-	1	-64	-69	-5	y	-69	-5	y	-69	-5	y
232	GLU	-	1	-170	171	-20	y	171	-20	y	171	-20	y
232	GLU	-	2	-176	-166	11	y	-166	11	y	-166	11	y
232	GLU	-	3	162	-1	17	y	-1	17	y	-1	17	y
233	THR	-	1	-54	-50	4	y	-50	4	y	-50	4	y
234	ARG	-	1	52	60	8	y	60	8	y	60	8	y
234	ARG	-	2	-155	-167	-13	y	-167	-13	y	-167	-13	y
234	ARG	-	3	-165	-163	2	y	-163	2	y	-163	2	y

234	ARG	-	4	172	-174	14	y	-174	14	y	-174	14	y
235	PRO	-	1	29	27	-2	y	27	-2	y	27	-2	y
235	PRO	-	2	-39	-41	-2	y	-41	-2	y	-41	-2	y
238	ASP	-	1	59	54	-6	y	54	-6	y	54	-6	y
238	ASP	-	2	-176	8	4	y	8	4	y	8	4	y
240	THR	-	1	73	61	-12	y	61	-12	y	61	-12	y
241	PHE	-	1	-59	-63	-4	y	-63	-4	y	-63	-4	y
241	PHE	-	2	-95	-89	5	y	-89	5	y	-89	5	y
242	GLN	-	1	-47	-61	-14	y	-61	-14	y	-61	-14	y
242	GLN	-	2	180	175	-5	y	175	-5	y	175	-5	y
242	GLN	-	3	-16	-2	14	y	-2	14	y	-2	14	y
243	LYS	-	1	-154	-157	-3	y	-157	-3	y	-157	-3	y
243	LYS	-	2	176	-173	11	y	-173	11	y	-173	11	y
243	LYS	-	3	-178	-177	1	y	-177	1	y	-177	1	y
243	LYS	-	4	75	68	-8	y	68	-8	y	68	-8	y
244	TRP	-	1	57	57	0	y	57	0	y	57	0	y
244	TRP	-	2	77	76	-1	y	76	-1	y	76	-1	y
245	ALA	VAL	1	66	-180	114	n	180	114	n	-62	-128	n
247	VAL	-	1	60	65	5	y	65	5	y	65	5	y
248	VAL	-	1	-178	174	-8	y	174	-8	y	174	-8	y
249	VAL	-	1	143	-49	168	n	-49	168	n	-49	168	n
250	PRO	-	1	-17	-25	-8	y	-25	-8	y	-25	-8	y
250	PRO	-	2	36	37	0	y	37	0	y	37	0	y
251	SER	-	1	-52	-173	-120	n	-173	-120	n	-173	-120	n
253	GLU	GLN	1	-59	-78	-19	y	-78	-19	y	-78	-19	y
253	GLU	GLN	2	-56	-174	-118	n	-174	-118	n	-174	-118	n
253	GLU	GLN	3	-56	56	-68	n	56	-68	n	56	-68	n
254	GLU	-	1	-62	-78	-17	y	-78	-17	y	-78	-17	y
254	GLU	-	2	80	88	9	y	88	9	y	88	9	y
254	GLU	-	3	-173	-23	-30	y	-23	-30	y	-23	-30	y
255	GLN	-	1	-53	-62	-9	y	-62	-9	y	-62	-9	y
255	GLN	-	2	-52	-69	-17	y	-69	-17	y	-69	-17	y
255	GLN	-	3	82	-28	69	n	-28	69	n	-28	69	n
256	ARG	-	1	-60	-51	9	y	-51	9	y	-51	9	y
256	ARG	-	2	-176	-166	11	y	-166	11	y	-166	11	y
256	ARG	-	3	-176	-172	5	y	-172	5	y	-172	5	y
256	ARG	-	4	-96	-176	-80	n	-176	-80	n	-176	-80	n
257	TYR	-	1	-58	-78	-19	y	-78	-19	y	-78	-19	y
257	TYR	-	2	-98	-95	4	y	-95	4	y	-95	4	y
258	THR	-	1	50	54	3	y	54	3	y	54	3	y
259	CYS	-	1	164	-179	17	y	-178	18	y	-178	18	y
260	HSD	-	1	-72	-62	10	y	-62	10	y	-62	10	y
260	HSD	-	2	-86	-79	7	y	-79	7	y	-79	7	y
261	VAL	-	1	-176	-172	4	y	-172	4	y	-172	4	y
262	GLN	-	1	-70	-64	7	y	-64	7	y	-64	7	y
262	GLN	-	2	179	-177	4	y	-177	4	y	-177	4	y
262	GLN	-	3	-85	79	-16	y	79	-16	y	79	-16	y
263	HSD	-	1	-167	-173	-6	y	-173	-6	y	-173	-6	y
263	HSD	-	2	81	66	-15	y	66	-15	y	66	-15	y
264	GLU	-	1	-159	69	-132	n	69	-132	n	69	-132	n
264	GLU	-	2	178	-180	2	y	-180	2	y	-180	2	y
264	GLU	-	3	123	84	-39	y	84	-39	y	84	-39	y
266	LEU	-	1	-66	-62	4	y	-62	4	y	-62	4	y
266	LEU	-	2	-64	-64	0	y	-64	0	y	-64	0	y
267	PRO	-	1	-44	22	67	n	22	67	n	22	67	n
267	PRO	-	2	40	-39	-79	n	-39	-79	n	-39	-79	n
268	LYS	-	1	-62	-61	0	y	-61	0	y	-61	0	y
268	LYS	-	2	176	172	-4	y	172	-4	y	172	-4	y
268	LYS	-	3	-177	-180	-3	y	-180	-3	y	-180	-3	y
268	LYS	-	4	-66	-173	-106	n	-173	-106	n	-173	-106	n
269	PRO	-	1	34	20	-14	y	20	-14	y	20	-14	y
269	PRO	-	2	-43	-34	9	y	-34	9	y	-34	9	y
270	LEU	-	1	-174	-75	98	n	-75	98	n	-75	98	n
270	LEU	-	2	-146	-66	79	n	-66	79	n	-66	79	n

	Structure 0			Structure 1			Structure 3		
χ	# y	# n	Frac. y	# y	# n	Frac. y	# y	# n	Frac. y
1	187	40	0.82	185	42	0.81	185	42	0.81
2	136	35	0.80	136	35	0.80	132	39	0.77
3	41	35	0.54	43	33	0.57	43	33	0.57
4	16	16	0.50	16	16	0.50	18	14	0.56
1&2	167	60	0.74	166	61	0.73	163	64	0.72

Table 2. χ angles of model structures.

		HLA-B27		HLA-A68		HLA-B35		HLA-B53		HLA-Cw4		HLA-Cw6		HLA-Cw7	
χ_i	Res#	Type	χ	Type	χ	Type	χ	Type	χ	Type	χ	Type	χ	Type	χ
1	1	GLY		---		---		---		---		CYS	-60	---	
1	**5**	**MET**	**-171**	**---**	**-171**	**---**	**-171**	**---**	**-171**	**---**	**-171**	**---**	**-171**	**---**	**-171**
2	**5**	**MET**	**172**	**---**	**172**	**---**	**172**	**---**	**172**	**---**	**172**	**---**	**172**	**---**	**172**
3	**5**	**MET**	**58**	**---**	**58**	**---**	**58**	**---**	**58**	**---**	**58**	**---**	**58**	**---**	**58**
1	**7**	**TYR**	**-77**	**---**	**-62**	**---**	**-54**	**---**	**-54**	**---**	**-77**	**---**	**-77**	**---**	**-77**
2	**7**	**TYR**	**103**	**---**	**94**	**---**	**105**	**---**	**105**	**---**	**103**	**---**	**103**	**---**	**103**
1	**9**	**HSD**	**-42**	**TYR**	**-59**	**TYR**	**-59**	**TYR**	**-59**	**SER**	**-42**	**ASP**	**-70**	**ASP**	**-70**
2	**9**	**HSD**	**81**	**TYR**	**-81**	**TYR**	**96**	**TYR**	**96**			**ASP**	**-142**	**ASP**	**-143**
1	11	SER	-72	---	-73	ALA		ALA		---	-73	ALA		ALA	
1	12	VAL	-175	---	-175	MET	-180	MET	-180	---	-175	---	-175	---	-175
2	12					MET	-180	MET	-180						
3	12					MET	180	MET	180						
1	13	SER	46	---	46	---	46	---	46	---	46	---	46	---	46
1	14	ARG	-58	---	-58	---	-58	---	-58	TRP	-75	---	-58	---	-58
2	14	ARG	-171	---	-171	---	-171	---	-171	TRP	-75	---	-171	---	-171
3	14	ARG	-68	---	-68	---	-68	---	-68			---	-68	---	-68
4	14	ARG	-174	---	-174	---	-174	---	-174			---	-174	---	-174
1	17	ARG	-70	---	-70	---	-70	---	-70	---	173	---	-70	---	-70
2	17	ARG	-171	---	-171	---	-171	---	-171	---	163	---	-171	---	-171
3	17	ARG	-59	---	-59	---	-59	---	-59	---	77	---	-59	---	-59
4	17	ARG	117	---	117	---	117	---	117	---	-170	---	117	---	117
1	**22**	**PHE**	**176**	**---**	**175**	**---**	**176**	**---**	**176**	**---**	**169**	**---**	**176**	**---**	**176**
2	**22**	**PHE**	**75**	**---**	**97**	**---**	**-86**	**---**	**-86**	**---**	**77**	**---**	**75**	**---**	**75**
1	**24**	**THR**	**180**	**ALA**		**ALA**		**ALA**		**ALA**		**SER**	**-180**	**SER**	**-180**
1	25	VAL	-63	---	-63	---	-63	---	-63	---	-63	---	-63	---	-63
1	32	LEU	176	GLN	176	---	176	---	176	GLN	176	GLN	176	GLN	176
2	32	LEU	-159	GLN	180	---	-159	---	-159	GLN	180	GLN	180	GLN	180
3	32			GLN	180					GLN	180	GLN	180	GLN	180
1	34	VAL	-75	---	-75	---	-75	---	-75	---	-75	---	-75	---	-75
1	36	PHE	175	---	-170	---	175	---	175	---	175	---	175	---	175
2	36	PHE	96	---	150	---	96	---	96	---	96	---	96	---	96
1	43	PRO	32	GLN	69	---	32	---	32	---	32	---	32	---	32
2	43	PRO	-43	GLN	-58	---	-43	---	-43	---	-43	---	-43	---	-43
3	43			GLN	116										
1	45	GLU	180	MET	-168	THR	-60	THR	-60	GLY		GLY		GLY	
2	45	GLU	160	MET	178										
3	45	GLU	-86	MET	75										
1	49	ALA		---		---		---		GLU	-104	---		---	
2	49									GLU	-77				
3	49									GLU	-73				
1	52	ILE	-168	---	-168	---	-168	---	-168	VAL	72	VAL	72	VAL	72
2	52	ILE	69	---	69	---	69	---	69						
1	**59**	**TYR**	**176**	**---**	**176**	**---**	**176**	**---**	**176**	**---**	**176**	**---**	**176**	**---**	**176**
2	**59**	**TYR**	**90**	**---**	**81**	**---**	**81**	**---**	**81**	**---**	**90**	**---**	**90**	**---**	**90**
1	62	ARG	179	---	179	---	179	---	179	---	178	---	177	---	178
2	62	ARG	-163	---	-163	---	-163	---	-163	---	-173	---	-173	---	-173
3	62	ARG	-177	---	-177	---	-177	---	-177	---	85	---	86	---	85
4	62	ARG	-97	---	-97	---	-97	---	-97	---	178	---	175	---	177
1	**63**	**GLU**	**-59**	**ASN**	**-55**	**ASN**	**-77**	**ASN**	**-77**	**---**	**-59**	**---**	**-59**	**---**	**-59**
2	**63**	**GLU**	**-43**	**ASN**	**-42**	**ASN**	**-92**	**ASN**	**-92**	**---**	**-43**	**---**	**-43**	**---**	**-43**
3	**63**	**GLU**	**-74**							**---**	**-74**	**---**	**-74**	**---**	**-74**
1	65	GLN	-48	ARG	-167	---	-48	---	-48	---	-48	---	-48	---	-48
2	65	GLN	173	ARG	-62	---	173	---	173	---	173	---	173	---	173

3	65	GLN	110	ARG	-50	---	110	---	110	---	110	---	110	---	110
4	65			ARG	-168										
1	**66**	**ILE**	**-71**	**ASN**	**-60**	**---**	**-71**	**---**	**-71**	**LYS**	**104**	**LYS**	**104**	**LYS**	**104**
2	**66**	**ILE**	**164**	**ASN**	**0**	**---**	**164**	**---**	**164**	**LYS**	**-169**	**LYS**	**-168**	**LYS**	**-169**
3	**66**									**LYS**	**174**	**LYS**	**175**	**LYS**	**172**
4	**66**									**LYS**	**-175**	**LYS**	**-176**	**LYS**	**-175**
1	**67**	**CYS**	**-68**	**VAL**	**-180**	**PHE**	**-71**	**PHE**	**-71**	**TYR**	**-72**	**TYR**	**-73**	**TYR**	**-73**
2	**67**					**PHE**	**-126**	**PHE**	**-126**	**TYR**	**126**	**TYR**	**127**	**TYR**	**128**
1	69	ALA		ALA		THR	-60	THR	-60	ARG	-67	ARG	-67	ARG	-67
2	69									ARG	168	ARG	168	ARG	168
3	69									ARG	-174	ARG	-174	ARG	-174
4	69									ARG	154	ARG	154	ARG	154
1	**70**	**LYS**	**174**	**GLN**	**174**	**ASN**	**-93**	**ASN**	**-94**	**GLN**	**174**	**GLN**	**174**	**GLN**	**174**
2	**70**	**LYS**	**96**	**GLN**	**96**	**ASN**	**-80**	**ASN**	**-82**	**GLN**	**96**	**GLN**	**96**	**GLN**	**96**
3	**70**	**LYS**	**170**	**GLN**	**180**					**GLN**	**180**	**GLN**	**180**	**GLN**	**180**
4	**70**	**LYS**	**82**												
1	71	ALA		SER	74	THR	-60	THR	-60	---		---		---	
1	**73**	**THR**	**-51**	**---**	**-51**	**---**	**-51**	**---**	**-51**	**ALA**		**ALA**		**ALA**	
1	**74**	**ASP**	**-80**	**---**	**-80**	**TYR**	**-78**	**TYR**	**-76**	**---**	**-80**	**---**	**-80**	**---**	**-80**
2	**74**	**ASP**	**-1**	**---**	**-1**	**TYR**	**11**	**TYR**	**16**	**---**	**-1**	**---**	**-1**	**---**	**-1**
1	76	GLU	-56	VAL	180	---	-56	---	-56	VAL	180	VAL	180	VAL	180
2	76	GLU	170			---	170	---	170						
3	76	GLU	-12			---	-12	---	-12						
1	**77**	**ASP**	**-74**	**---**	**-74**	**SER**	**-74**	**ASN**	**-74**	**ASN**	**-74**	**ASN**	**-74**	**SER**	**-74**
2	**77**	**ASP**	**-16**	**---**	**-16**			**ASN**	**-16**	**ASN**	**-16**	**ASN**	**-16**		
1	79	ARG	-73	GLY		---	-73	---	-73	---	-73	---	-73	---	-73
2	79	ARG	-179			---	-179	---	-179	---	-179	---	-179	---	-179
3	79	ARG	-83			---	-83	---	-83	---	-83	---	-83	---	-83
4	79	ARG	-108			---	-108	---	-108	---	-108	---	-108	---	-108
1	**80**	**THR**	**-58**	**---**	**-58**	**ASN**	**-60**	**ILE**	**-58**	**LYS**	**-62**	**LYS**	**-62**	**ASN**	**-60**
2	**80**					**ASN**	**0**	**ILE**	**-180**	**LYS**	**174**	**LYS**	**174**	**ASN**	**0**
3	**80**									**LYS**	**-173**	**LYS**	**-172**		
4	**80**									**LYS**	**160**	**LYS**	**160**		
1	**81**	**LEU**	**-98**	**---**	**-98**	**---**	**-98**	**ALA**		**---**	**-98**	**---**	**-98**	**---**	**-98**
2	**81**	**LEU**	**164**	**---**	**164**	**---**	**164**			**---**	**164**	**---**	**164**	**---**	**164**
1	82	LEU	-67	ARG	-67	ARG	-67	---	-67	ARG	-67	ARG	-67	ARG	-67
2	82	LEU	-55	ARG	-180	ARG	-180	---	-55	ARG	-180	ARG	-180	ARG	-180
3	82			ARG	-180	ARG	-180			ARG	-180	ARG	-180	ARG	-180
4	82			ARG	-180	ARG	-180			ARG	-180	ARG	-180	ARG	-180
1	83	ARG	175	GLY		GLY		---	175	GLY		GLY		GLY	
2	83	ARG	-178					---	-178						
3	83	ARG	73					---	73						
4	83	ARG	-110					---	-110						
1	**84**	**TYR**	**-68**	**---**	**-68**	**---**	**-68**	**---**	**-68**	**---**	**-63**	**---**	**-63**	**---**	**-68**
2	**84**	**TYR**	**179**	**---**	**179**	**---**	**179**	**---**	**179**	**---**	**179**	**---**	**179**	**---**	**179**
1	90	ALA		---		---		---		ASP	-60	ASP	-60	ASP	-60
2	90									ASP	0	ASP	0	ASP	0
1	94	THR	-56	---	-56	ILE	-56	ILE	-56	---	-56	---	-56	---	-56
2	94					ILE	180	ILE	180						
1	**95**	**LEU**	**-107**	**ILE**	**-60**	**ILE**	**-60**	**ILE**	**-60**	**---**	**-107**	**---**	**-107**	**---**	**-107**
2	**95**	**LEU**	**-53**	**ILE**	**-180**	**ILE**	**-180**	**ILE**	**-180**	**---**	**-53**	**---**	**-53**	**---**	**-53**
1	**97**	**ASN**	**168**	**MET**	**168**	**ARG**	**-180**	**ARG**	**174**	**ARG**	**168**	**TRP**	**-160**	**ARG**	**168**
2	**97**	**ASN**	**-91**	**MET**	**180**	**ARG**	**169**	**ARG**	**165**	**ARG**	**168**	**TRP**	**149**	**ARG**	**180**
3	**97**			**MET**	**-180**	**ARG**	**168**	**ARG**	**172**	**ARG**	**-165**			**ARG**	**-180**
4	**97**					**ARG**	**-130**	**ARG**	**-122**	**ARG**	**124**			**ARG**	**180**
1	**99**	**TYR**	**66**	**---**	**66**	**---**	**66**	**---**	**66**	**PHE**	**66**	**---**	**64**	**SER**	**66**
2	**99**	**TYR**	**-99**	**---**	**-99**	**---**	**-99**	**---**	**-99**	**PHE**	**-99**	**---**	**-51**	**SER**	
1	103	VAL	-53	---	-53	LEU	57	LEU	57	LEU	57	LEU	57	LEU	57
2	103					LEU	-167	LEU	-167	LEU	-167	LEU	-167	LEU	-167
1	105	PRO	26	SER	60	---	26	---	26	---	26	---	26	---	26
2	105	PRO	-35			---	-35	---	-35	---	-35	---	-35	---	-35
1	109	LEU	-174	PHE	44	---	-174	---	-174	---	-174	---	-174	---	-174

2	109	LEU	-179	PHE	-90	---	-179	---	-179	---	-179	---	-179	---	-179
1	110	LEU	-116	---	-116	---	-116	---	-116	---	-116	---	-73	---	-116
2	110	LEU	-62	---	-62	---	-62	---	-62	---	-62	---	-60	---	-62
1	113	TYR	-62	---	-62	HSD	-62	HSD	-62	---	-62	---	-62	---	-62
2	113	TYR	95	---	95	HSD	95	HSD	95	---	95	---	95	---	95
1	**114**	**HSD**	**175**	**ARG**	**175**	**ASP**	**179**	**ASP**	**179**	**ASN**	**-108**	**ASP**	**-110**	**ASP**	**-152**
2	**114**	**HSD**	**-76**	**ARG**	**-180**	**ASP**	**-164**	**ASP**	**-166**	**ASN**	**-87**	**ASP**	**58**	**ASP**	**134**
3	**114**			**ARG**	**-180**										
4	**114**			**ARG**	**-180**										
1	**116**	**ASP**	**-71**	**---**	**-71**	**SER**	**-71**	**SER**	**-71**	**PHE**	**-69**	**SER**	**-160**	**SER**	**-71**
2	**116**	**ASP**	**86**	**---**	**86**					**PHE**	**61**				
1	123	TYR	-174	---	-174	---	-174	---	-174	---	-174	---	-174	---	-174
2	123	TYR	-103	---	-103	---	-103	---	-103	---	-103	---	-103	---	-103
1	127	ASN	-65	LYS	-155	---	-65	---	-65	---	-65	---	-65	---	-65
2	127	ASN	-22	LYS	-173	---	-22	---	-22	---	-22	---	-22	---	-22
3	127			LYS	82										
4	127			LYS	-132										
1	131	SER	-59	ARG	172	---	-59	---	-59	ARG	172	ARG	172	ARG	173
2	131			ARG	-79					ARG	-79	ARG	-79	ARG	-79
3	131			ARG	-48					ARG	-49	ARG	-49	ARG	-49
4	131			ARG	-120					ARG	-136	ARG	-136	ARG	-138
1	132	SER	55	---	52	---	55	---	55	---	55	---	55	---	55
1	133	TRP	-63	---	-63	---	-63	---	-63	---	-46	---	-63	---	-63
2	133	TRP	-84	---	-84	---	-84	---	-84	---	-80	---	-84	---	-84
1	138	THR	53	MET	-60	---	53	---	53	---	53	---	53	---	53
2	138			MET	180										
3	138			MET	180										
1	141	GLN	-80	---	-80	---	-80	---	-80	---	-80	---	-80	---	-80
2	141	GLN	171	---	171	---	171	---	171	---	171	---	171	---	171
3	141	GLN	-38	---	-38	---	-38	---	-38	---	-38	---	-38	---	-38
1	**143**	**THR**	**-61**	**---**	**-61**	**---**	**-61**	**---**	**-61**	**---**	**-61**	**---**	**-61**	**---**	**-61**
1	144	GLN	-173	LYS	-173	---	-173	---	-173	---	-173	---	-173	---	-173
2	144	GLN	177	LYS	177	---	177	---	177	---	177	---	177	---	177
3	144	GLN	-24	LYS	-180	---	-24	---	-24	---	-24	---	-24	---	-24
4	144			LYS	-180										
1	**146**	**LYS**	**-67**	**---**	**-67**	**---**	**-67**	**---**	**-67**	**---**	**-67**	**---**	**-67**	**---**	**-67**
2	**146**	**LYS**	**-176**	**---**	**-176**	**---**	**-176**	**---**	**-176**	**---**	**-176**	**---**	**-176**	**---**	**-176**
3	**146**	**LYS**	**177**	**---**	**177**	**---**	**177**	**---**	**177**	**---**	**177**	**---**	**177**	**---**	**177**
4	**146**	**LYS**	**-179**	**---**	**-179**	**---**	**-179**	**---**	**-179**	**---**	**-179**	**---**	**-179**	**---**	**-179**
1	**147**	**TRP**	**-77**	**---**	**-77**	**---**	**-77**	**---**	**-77**	**---**	**-64**	**---**	**-64**	**LEU**	**-77**
2	**147**	**TRP**	**-171**	**---**	**-171**	**---**	**-171**	**---**	**-171**	**---**	**-155**	**---**	**-154**	**LEU**	**-60**
1	148	GLU	-61	---	-73	---	-74	---	-74	---	-74	---	-74	---	-74
2	148	GLU	177	---	171	---	169	---	169	---	169	---	169	---	169
3	148	GLU	-35	---	-5	---	-8	---	-8	---	-7	---	-7	---	-8
1	151	ARG	-47	HSD	-47	---	-47	---	-47	---	-47	---	-47	---	-47
2	151	ARG	158	HSD	90	---	158	---	158	---	158	---	158	---	158
3	151	ARG	65			---	65	---	65	---	65	---	65	---	65
4	151	ARG	121			---	121	---	121	---	121	---	121	---	121
1	**152**	**VAL**	**172**	**---**	**172**	**---**	**172**	**---**	**172**	**GLU**	**-71**	**GLU**	**-135**	**ALA**	
2	**152**									**GLU**	**-88**	**GLU**	**-170**		
3	**152**									**GLU**	**-4**	**GLU**	**73**		
1	154	GLU	-79	---	-79	---	-79	---	-79	---	-79	---	-79	ASP	-68
2	154	GLU	174	---	174	---	174	---	174	---	174	---	174	ASP	171
3	154	GLU	-107	---	-107	---	-107	---	-107	---	-107	---	-107		
1	**155**	**GLN**	**-70**	**---**	**-70**	**---**	**-70**	**---**	**-70**	**---**	**-70**	**---**	**-70**	**GLU**	**-70**
2	**155**	**GLN**	**174**	**---**	**174**	**---**	**174**	**---**	**174**	**---**	**174**	**---**	**174**	**GLU**	**174**
3	**155**	**GLN**	**-83**	**---**	**-83**	**---**	**-83**	**---**	**-83**	**---**	**-83**	**---**	**-83**	**GLU**	**-83**
1	**156**	**LEU**	**-57**	**TRP**	**-132**	**---**	**-60**	**---**	**-60**	**ARG**	**-59**	**TRP**	**-65**	**---**	**-60**
2	**156**	**LEU**	**-62**	**TRP**	**57**	**---**	**-78**	**---**	**-78**	**ARG**	**-180**	**TRP**	**-71**	**---**	**-80**
3	**156**									**ARG**	**-177**				
4	**156**									**ARG**	**178**				
1	157	ARG	-179	---	-179	---	-179	---	-179	---	-179	---	-179	---	176

2	157	ARG	161	---	161	---	161	---	161	---	161	---	161	---	173
3	157	ARG	76	---	76	---	76	---	76	---	76	---	76	---	169
4	157	ARG	-105	---	-105	---	-105	---	-105	---	-105	---	-105	---	-106
1	**159**	**TYR**	**167**	**---**	**167**	**---**	**167**	**---**	**167**	**---**	**167**	**---**	**167**	**---**	**167**
2	**159**	**TYR**	**74**	**---**	**74**	**---**	**74**	**---**	**74**	**---**	**74**	**---**	**74**	**---**	**74**
1	160	LEU	-65	---	-65	---	-65	---	-65	---	-65	---	-65	---	-65
2	160	LEU	-73	---	-73	---	-73	---	-73	---	-73	---	-73	---	-73
1	**163**	**GLU**	**-110**	**THR**	**-60**	**LEU**	**-110**	**LEU**	**-110**	**THR**	**-60**	**THR**	**-60**	**THR**	**-60**
2	**163**	**GLU**	**-174**			**LEU**	**-60**	**LEU**	**-60**						
3	**163**	**GLU**	**-75**												
1	165	VAL	-176	---	-176	---	-176	---	-176	---	-176	---	-176	---	-176
1	**167**	**TRP**	**-80**	**---**	**-80**	**---**	**-80**	**---**	**-80**	**---**	**-80**	**---**	**-80**	**---**	**-80**
2	**167**	**TRP**	**-98**	**---**	**-98**	**---**	**-98**	**---**	**-98**	**---**	**-98**	**---**	**-98**	**---**	**-98**
1	**171**	**TYR**	**-64**	**---**	**-64**	**---**	**-64**	**---**	**-64**	**---**	**-64**	**---**	**-64**	**---**	**-64**
2	**171**	**TYR**	**-27**	**---**	**-27**	**---**	**-27**	**---**	**-27**	**---**	**-27**	**---**	**-27**	**---**	**-27**
1	178	THR	51	---	51	---	51	---	51	---	51	SER	180	---	51

Table 3. Eluted peptides from HLA alleles

HLA-A2[a]

Pooled eluted peptides

1	2	3	4	5	6	7	8	9	
	L							**V**	anchor
	M		E		V		K		strong
			K						strong
I		A	G	I	I	A	E	Lwweak	
L		Y	P	K	L	Y	S		weak
F		F	D	Y	T	H			weak
K		P	T	N					weak
M		M		G					weak
Y		S		F					weak
V		R		V					weak

Eluted peptides

1	2	3	4	5	6	7	8	9
L	**L**	D	V	P	T	A	A	**V**
S	**L**	L	P	A	I	V	E	L
Y	**L**	L	P	A	I	V	H	I
T	**L**	W	V	D	P	Y	E	**V**

Viral peptides

1	2	3	4	5	6	7	8	9
I	**L**	K	E	P	V	H	G	**V**
G	I	L	G	F	V	F	T	L
I	**L**	G	F	V	F	T	L	T
F	**L**	Q	S	R	P	E	P	T
A	M	Q	M	L	K	E		
P	I	A	P	G	Q	M	R	E
Q	M	K	D	C	T	E	R	Q

HLA-A68[b]

Eluted peptides

1	2	3	4	5	6	7	8	9	10	11	12
E	**V**	A	P	P	E	Y	H	**R**			
A	**V**	A	A	V	A	A	R	**R**			
D	**V**	F	R	D	P	A	L	**K**			
E	**V**	A	P	P	E	Y	H	R	**K**		
E	**V**	I	L	I	D	P	F	H	**K**		
T	**V**	F	D	A	K	R	L	I	G	**R**	
K	**T**	G	G	P	I	Y	K	**R**			

HLA-B27[c]

Eluted peptides

1	2	3	4	5	6	7	8	9			
R	**R**	Y	Q	K	S	T	E	L			
R	**R**	I	K	E	I	V	K	**K**			
R	**R**	V	K	E	V	V	K	**K**			
R	**R**	W	L	P	A	G	d	a			
R	**R**	S	K	E	I	T	V	**R**			
G	**R**	I	D	K	P	I	L	**K**			
F	**R**	Y	N	G	L	i	H	**r**			
K	**R**	F	E	G	L	T	Q	**R**			
R	**R**	F	T	R	P	E	H	–			
R	**R**	I	S	G	V	D	R	Y			
A	**R**	L	F	G	I	R	A	**K**			
S	**R**	Y	W	A	I	R	T	**R**			
K	**R**	W	I	I	L	G	L	N	K	I	Y
G	**R**	A	F	V	T	I	G	**K**			

HLA-B35[d]

Pooled eluted peptides

1	2	3	4	5	6	7	8	9	
	P							**Y**	anchor
		F							strong
F			E	E					weak
Y			I	K					weak
L			N						weak
M			Q						weak
			Y						weak

P. falciparum peptides

1	2	3	4	5	6	7	8	9
K	**P**	K	D	E	L	D	**Y**	
K	S	K	D	E	L	D	**Y**	
K	**P**	N	D	K	S	L	**Y**	

HLA-B53[d]

Pooled eluted peptides

1	2	3	4	5	6	7	8	9	
	P								anchor
			E	I					strong
S		F	I	L	Y				weak
Y		K	L						weak
F		N	Q						weak
M		Q							weak
									weak

P. falciparum peptides

1	2	3	4	5	6	7	8	9
M	**P**	N	D	P	N	R	N	V
M	**P**	N	Y	P	N	R	N	V
M	**P**	N	N	P	N	R	N	V
I	**P**	S	L	A	L	M	L	I
M	**P**	L	E	T	Q	L	A	I
K	**P**	I	V	Q	Y	D	N	F
E	**P**	A	P	F	D	E	T	L
H	**P**	S	D	G	K	C	N	L

HLA-Cw4[e]

Pooled eluted peptides

1	2	3	4	5	6	7	8	9	
	Y							**L**	anchor
	F							**M**	anchor
								F	anchor
								i	anchor
				A	V				auxiliary anchor
				M	L				auxiliary anchor
				F	I				auxiliary anchor
					F				auxiliary anchor
Q		D	D		D	Q	K		reproducibly detected
F		E	N			A			reproducibly detected
		G	E			E			reproducibly detected
		N	K						reproducibly detected
			P						reproducibly detected

HLA-Cw6[e]

Pooled eluted peptides

1	2	3	4	5	6	7	8	9	
								L	anchor
								I	anchor
								V	anchor
								m	anchor
				F	V				auxiliary anchor
				M	I				auxiliary anchor
				L	T				auxiliary anchor
				T					auxiliary anchor
				V					auxiliary anchor
I	P	Y	P	K		R	K		reproducibly detected
	R	F	E			K	E		reproducibly detected
		P	D			Q	F		reproducibly detected
		D	K			N	Y		reproducibly detected
		K	N			A			reproducibly detected
		I	G			E			reproducibly detected
		N							reproducibly detected

HLA-Cw7[e]

Pooled eluted peptides

1	2	3	4	5	6	7	8	9	
								Y	anchor
								L	anchor
								f	anchor
								m	anchor
	Y			F	V				auxiliary anchor
				M	I				auxiliary anchor
				Y	L				auxiliary anchor
				I	A				auxiliary anchor
	P	P	D	K		y	K		reproducibly detected
	r	F	E			v	E		reproducibly detected
		G	P			D	D		reproducibly detected
		N							reproducibly detected
		A							reproducibly detected

---

[a] Pooled eluted peptides and compilation of viral peptides from Falk et al. (Ref. 16). Eluted peptides from Hunt et al. (Ref. 17). Residues in bold type are "anchor residues."

[b] Guo et al. (Ref. 6). Residues in bold type are part of peptide binding motif.

[c] Jardetzky et al. (Ref. 5). Residues in lower-case indicate less than high confidence. Last three sequences are from viral peptides.

[d] Hill et al. (Ref. 1). Sequence information, designated as anchor, strong, and weak by Hill et al., is from pooled peptides. Sequences following are peptides from P. falciparum which bind to the HLA allele (B35 or B53).

[e] Rötzschke et al. (Ref. 2). Anchor and auxiliary anchor residues in bold. Lower-case letters indicate less confidence in residue identification.

Table 4. Pocket residues

Pocket B

	9	24	45	67	99	Anchor residue at P2
a2	F	A	M	V	Y	L, M
a68	Y	A	M	M	Y	V, T
b27	H	T	E	C	Y	R
b35	Y	A	T	F	Y	P
b53	Y	A	T	F	Y	P
cw4	S	A	G	Y	F	Y, F
cw6	D	S	G	Y	Y	
cw7	D	S	G	Y	S	Y

Pocket C

	9	97	99	Anchor residue at P5/P6
a2	F	R	Y	
a68	Y	M	Y	
b27	H	N	Y	
b35	Y	R	Y	
b53	Y	R	Y	I
cw4	S	R	F	A, F, M
cw6	D	W	Y	F, M, L, T, V
cw7	D	R	S	F, M, Y, I

Pocket F

	81	95	116	Anchor residue at PC
a2	L	V	Y	V
a68	L	I	D	K, R
b27	L	L	D	K, R
b35	L	I	S	Y
b53	A	I	S	—
cw4	L	L	F	L, F, M, I
cw6	L	L	S	L, V, I, M
cw7	L	L	S	Y, L, F, M

Table 5. Interaction energies after energy minimization of peptide model in the antigen binding site[a]

			Total Energy			van der Waals Energy		Bonded Energy
HLA Allele	P	Type	P/HLA	P/Pep.	P/HLA+ P/pep	P/HLA	P/Pep	P/Pep
B27	2	ALA	-2.1	-5.4	-7.4	-2.2	-0.5	0.1
	2	ARG	-88.4	-44.0	-132.4	-4.6	2.7	27.8
	2	ASN	-11.5	9.7	-1.9	-6.3	-0.0	42.9
	2	GLU	-9.3	-8.4	-17.8	-9.3	0.1	1.8
	2	ILE	-4.8	114.5	109.7	-3.7	1.0	111.6
	2	LEU	-8.0	30.0	22.0	-8.3	-1.1	38.2
	2	PRO	-3.8	32.4	28.6	7.1	1.1	26.7
	2	SER	-5.5	6.8	1.3	-3.2	-0.2	2.1
	2	TRP	-25.9	181.9	156.0	-13.3	12.0	171.6
	2	TYR	-0.1	78.1	77.9	-5.5	12.6	73.5
B27	9	ALA	-3.8	-4.5	-8.4	-3.5	-0.4	0.0
	9	ARG	-65.3	7.6	-57.7	-13.3	0.9	87.0
	9	ASN	-22.7	-21.6	-44.3	-4.6	-0.6	5.8
	9	GLU	-13.4	-6.5	-19.9	-10.2	-0.1	1.1
	9	ILE	-11.1	8.3	-2.8	-10.4	0.6	7.0
	9	LEU	-10.5	-3.0	-13.5	-9.9	-0.3	4.4
	9	PRO	-7.9	11.7	3.8	-5.2	-1.2	12.4
	9	SER	-5.2	6.5	1.3	-5.1	-0.3	1.7
	9	TRP	-15.6	80.6	65.0	-16.1	12.7	68.7
	9	TYR	-38.8	30.3	-8.5	-5.9	9.6	27.9
A68	2	ALA	-1.9	-5.3	-7.2	-2.3	-0.5	0.1
	2	ARG	-20.7	-50.0	-70.8	-16.9	0.1	22.2
	2	ASN	-5.2	-17.6	-22.8	-7.6	0.4	12.5
	2	GLU	-29.8	-9.4	-39.2	-9.9	0.4	1.2
	2	ILE	-6.6	34.2	27.6	-7.3	0.1	33.1
	2	LEU	-8.8	8.5	-0.3	-10.2	0.8	15.5
	2	PRO	-4.4	14.9	10.4	-6.8	1.5	8.5
	2	SER	-2.9	5.4	2.4	-3.6	-0.3	1.9
	2	TRP	-17.4	103.6	86.2	-18.1	15.6	89.5
	2	TYR	-14.1	45.1	31.0	-14.7	15.2	38.6
A68	9	ALA	-3.8	-4.5	-8.4	-3.5	-0.4	0.1
	9	ARG	-71.2	-12.4	-83.6	-14.1	0.8	71.9
	9	ASN	-25.8	-21.5	-47.4	-2.6	-0.5	6.3
	9	GLU	-8.5	-6.4	-15.0	-9.6	0.0	1.1
	9	ILE	-11.0	9.1	-1.9	-9.8	0.3	8.0
	9	LEU	-10.0	-1.4	-11.4	-9.0	-0.5	6.1
	9	PRO	-9.1	13.1	4.1	-5.4	-1.2	13.8
	9	SER	-5.9	7.2	1.2	-5.0	-0.3	2.4
	9	TRP	-21.8	56.9	35.1	-15.8	5.0	52.8
	9	TYR	-37.5	30.7	-6.8	5.2	10.0	28.1
B35	2	ALA	-2.1	-5.3	-7.4	-2.1	-0.5	0.1
	2	ARG	-15.4	154.8	139.4	-8.1	4.1	228.8
	2	ASN	-13.6	16.4	2.8	-4.8	0.3	45.3
	2	GLU	-19.8	66.1	46.3	-7.0	1.8	77.9
	2	ILE	-8.0	103.8	95.8	-7.7	15.9	85.9
	2	LEU	8.3	108.5	116.8	5.4	8.0	107.7
	2	PRO	-6.0	18.2	12.1	-6.2	1.9	11.4
	2	SER	-5.9	6.7	0.8	-3.3	0.0	2.7

	2	TRP	-12.5	76.9	64.4	-13.8	9.4	69.0
	2	TYR	-16.7	55.8	39.2	-15.5	20.1	43.1
B35	9	ALA	-3.5	-4.5	-8.0	-3.3	-0.4	0.0
	9	ARG	-27.5	-48.8	-76.3	-12.5	0.6	31.7
	9	ASN	-11.4	-24.7	-36.1	-6.0	-0.4	2.5
	9	GLU	-18.0	-6.4	-24.5	-8.8	-0.2	1.4
	9	ILE	-9.4	6.2	-3.1	-9.4	0.4	5.1
	9	LEU	-8.4	-3.5	-11.9	-8.5	-0.4	3.9
	9	PRO	-5.0	8.8	3.8	-5.1	-1.1	9.5
	9	SER	-4.3	5.7	1.4	-4.7	-0.3	1.1
	9	TRP	-13.7	125.7	112.0	-14.0	10.0	116.1
	9	TYR	-16.5	9.3	-7.3	-12.2	10.1	7.0
B53	2	ALA	-2.1	-5.3	-7.4	-2.1	-0.5	0.1
	2	ARG	-25.2	112.5	87.3	-16.0	5.4	173.1
	2	ASN	-13.4	15.8	2.4	-5.2	0.3	44.7
	2	GLU	-20.0	64.1	44.1	-7.0	1.9	75.9
	2	ILE	-8.1	102.3	94.2	-7.8	15.7	84.5
	2	LEU	7.5	108.8	116.4	4.6	8.4	107.5
	2	PRO	-6.1	18.1	12.0	-6.2	1.9	11.4
	2	SER	-5.7	6.7	1.0	-3.3	0.0	2.7
	2	TRP	-12.5	77.0	64.6	-13.9	9.3	69.2
	2	TYR	-16.7	47.5	30.8	-15.4	19.6	35.3
B53	9	ALA	-3.3	-4.5	-7.8	-3.2	-0.4	0.0
	9	ARG	-25.7	3.2	-22.4	-14.9	1.4	85.3
	9	ASN	-7.6	-24.8	-32.3	-6.0	-0.5	2.6
	9	GLU	-21.3	-6.4	-27.7	-9.4	-0.1	1.4
	9	ILE	-9.7	5.0	-4.7	-9.7	0.0	4.1
	9	LEU	-7.8	-3.6	-11.5	-7.9	-0.4	3.7
	9	PRO	-1.6	10.8	9.2	-4.8	-1.2	11.6
	9	SER	-3.9	5.6	1.7	-4.6	-0.3	1.0
	9	TRP	-10.1	101.0	90.9	-11.0	15.5	86.6
	9	TYR	-11.7	11.9	0.2	-9.5	11.0	8.6
Cw4	2	ALA	-1.6	79.5	77.8	-2.0	-0.1	82.9
	2	ARG	-14.1	185.5	171.5	-5.1	18.7	241.6
	2	ASN	11.1	137.2	148.3	6.2	10.2	156.8
	2	GLU	23.9	261.7	285.6	52.1	7.1	262.4
	2	ILE	65.5	361.7	427.2	73.5	23.0	337.4
	2	LEU	1632.2	4426.5	6058.7	1627.2	-0.1	4429.3
	2	PRO	0.9	204.8	205.7	3.1	1.5	198.5
	2	SER	-2.4	117.7	115.4	-1.6	0.4	112.3
	2	TRP	1607.4	591.8	2199.2	1623.8	19.0	573.4
	2	TYR	-16.2	191.6	175.4	-12.3	22.1	176.7
Cw4	9	ALA	-3.7	-4.5	-8.2	-3.6	-0.4	0.0
	9	ARG	-5.9	110.9	105.1	-5.1	2.3	187.3
	9	ASN	-9.9	-23.4	-33.3	-7.7	-0.7	4.0
	9	GLU	-28.4	-5.6	-33.9	-10.2	0.7	1.1
	9	ILE	-9.8	8.3	-1.5	-9.7	0.3	7.3
	9	LEU	-9.5	0.4	-9.1	-9.6	-0.4	7.7
	9	PRO	-3.2	12.8	9.6	-5.4	-1.2	13.5
	9	SER	-4.0	5.6	1.6	-5.3	-0.3	1.0
	9	TRP	-12.4	191.3	178.9	-13.6	12.5	180.4
	9	TYR	-11.1	69.3	58.2	-11.5	16.7	60.2
Cw6	2	ALA	-1.8	99.9	98.2	-2.7	-0.3	103.6
	2	ARG	-54.1	156.0	101.9	-6.0	8.8	228.6
	2	ASN	8.0	143.8	151.8	2.8	8.3	165.2
	2	GLU	-15.7	122.8	107.1	9.4	3.6	128.6
	2	ILE	81.1	397.2	478.3	87.8	28.9	366.0
	2	LEU	1.0	179.1	180.1	2.8	3.4	182.6

	2	PRO	-3.3	233.6	230.3	-0.2	1.7	227.0
	2	SER	-6.1	108.7	102.6	-1.4	0.9	102.8
	2	TRP	-15.9	226.5	210.6	-12.4	17.5	209.1
	2	TYR	-16.8	223.3	206.6	-12.5	23.2	207.7
Cw6	9	ALA	-3.7	-4.5	-8.2	-3.6	-0.4	0.0
	9	ARG	-18.8	-23.4	-42.2	-8.2	2.0	53.7
	9	ASN	-7.8	-24.2	-32.0	-7.1	-0.7	3.4
	9	GLU	-21.2	-5.9	-27.1	-10.3	0.3	1.3
	9	ILE	-9.8	10.5	0.6	-9.4	0.2	9.5
	9	LEU	-9.4	1.3	-8.1	-9.2	-0.3	8.7
	9	PRO	-3.8	13.8	10.0	-5.4	-1.2	14.5
	9	SER	-3.1	5.7	2.5	-5.2	-0.3	1.1
	9	TRP	-7.5	177.3	169.8	-7.8	15.4	162.8
	9	TYR	-9.9	18.7	8.8	-12.6	13.1	13.4
Cw7	2	ALA	-2.4	164.7	162.3	-2.4	-0.2	168.0
	2	ARG	-38.9	254.5	215.5	2.2	19.7	321.2
	2	ASN	-2.6	144.2	141.5	-2.8	2.3	172.2
	2	GLU	-17.6	193.5	175.9	1.0	6.8	195.7
	2	ILE	58.4	408.7	467.1	67.7	17.9	389.5
	2	LEU	11.3	280.6	291.9	13.2	2.0	285.8
	2	PRO	-3.6	284.2	280.6	-0.9	7.4	270.6
	2	SER	-3.8	121.2	117.4	-3.6	0.6	115.3
	2	TRP	-20.1	260.6	240.5	-16.4	20.5	240.5
	2	TYR	-10.8	269.3	258.5	-6.2	28.3	248.5
Cw7	9	ALA	-3.3	-4.5	-7.8	-3.0	-0.4	0.0
	9	ARG	-29.6	-4.3	-33.9	-12.4	0.3	85.4
	9	ASN	-9.6	-23.4	-33.1	-1.9	-0.4	3.7
	9	GLU	-22.7	-6.2	-28.9	-8.2	-0.2	1.6
	9	ILE	-7.5	11.8	4.3	-7.4	0.4	10.7
	9	LEU	-6.7	-2.6	-9.3	-6.9	-0.6	5.0
	9	PRO	-2.1	10.6	8.5	-5.0	-1.2	11.3
	9	SER	-3.9	5.7	1.8	-4.1	-0.3	1.1
	9	TRP	-13.3	47.6	34.3	-13.7	8.4	39.6
	9	TYR	-10.8	16.2	5.4	-9.8	9.9	14.1

[a] Abbreviations: P=Peptide residue position; P/HLA=Interaction energy between peptide residue and HLA protein; P/Pep=Interaction energy between peptide residue and whole of peptide

Figure 1. Comparison of HLA-A, B, and C sequences