Self-Consistent Assignment of Asparagine and Glutamine Amide Rotamers in Protein Crystal Structures

Ways & Means

Christian X. Weichenberger[1] and Manfred J. Sippl[1],*
[1] Center of Applied Molecular Engineering
University of Salzburg
Jakob Haringerstraße 5
5020 Salzburg
Austria

Summary

The current protein structure database contains unfavorable Asn/Gln amide rotamers in the order of 20%. Here, we derive a set of self-consistent potential functions to identify and correct unfavorable rotamers. Potentials of mean force for all heavy atoms are compiled from a database of high-resolution protein crystal structures. Starting from erroneous data, a refinement-correction cycle quickly converges to a self-consistent set of potentials. The refinement is entirely driven by the deposited structure data and does not involve any assumptions on molecular interactions or any artificial constraints. The refined potentials obtained in this way identify unfavorable rotamers with high confidence. Since the state of Asn/Gln rotamers is largely determined by hydrogen bond interactions, the features of the respective potentials are of interest in terms of molecular interactions, protein structure refinement, and prediction. The Asn/Gln rotamer assignment is available as a public web service intended to support protein structure refinement and modeling.

Introduction

The amide groups in the side chains of asparagine (Asn) and glutamine (Gln) are specific examples of functional groups that act simultaneously as hydrogen bond donors and acceptors. These amide groups frequently participate in hydrogen bond networks (Yoder et al., 1993), substrate binding, ligand docking (Faham et al., 1996; Evdokimov et al., 2001; Schafer et al., 2004), protein-protein interactions, and catalysis (Battiste et al., 1996; Coleman et al., 1994; Vernet et al., 1995; Raymond et al., 2004), and segments of Gln and Asn residues have been shown to form the cross-β spine of amyloid-like fibrils (Nelson et al., 2005; Sambashivan et al., 2005). The electron density near the nitrogen and oxygen atoms of Asn and Gln amide groups is compatible with two rotamers that are related by a 2-fold symmetry axis. Therefore, electron density maps obtained from X-ray diffraction experiments of protein crystals yield the positions of the oxygen and nitrogen atoms with high precision but not their identity resulting in the perpetual assignment of wrong rotamers.
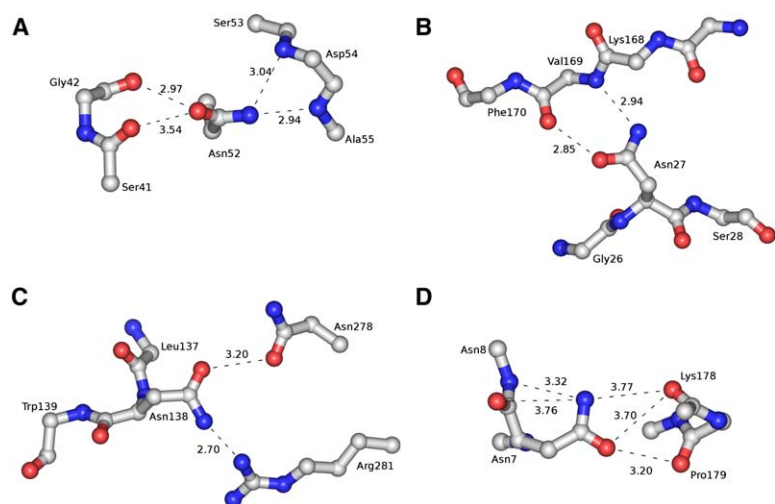
The Asn and Gln amide assignment problem is the subject of repeated investigations, and as a result, the error rate is well documented (Word et al., 1999; McDonald and Thornton, 1995; Hooft et al., 1996). Word et al. (1999)

provide a thorough analysis of Asn and Gln side chain amides found in 100 high-quality protein crystal structures. The authors use a variety of criteria to distinguish correct from incorrect rotamers, including the construction of moveable hydrogen atom positions required for the evaluation of interaction energies. They classify 65% of the 1557 investigated rotamers as clearly correct, 20% as clearly incorrect, and 15% as undecidable. The data set investigated by Word et al. (1999) is restricted to Asn/Gln residues whose B factors are less than 40. The proton locations required for the evaluation of hydrogen bond energies are obtained by using widely accepted standard bond lengths and valence angles between protons and the respective heavy atoms. The frequent cases where hydrogen atoms are rotatable are adjusted by optimizing H-bond geometries and hydrogen bond networks.

Another source of confident rotamer assignments is provided by Higman et al. (2004). These authors compare Asn and Gln amide rotamers of hen egg-white lysozyme determined by residual dipolar couplings in solution with the corresponding rotamers found in 16 high resolution protein crystal structures. The authors report confident assignments for 141 amide rotamers, 26 of which disagree with the rotamers found in the respective crystal structures.

Although the Asn and Gln amide rotamers frequently have indistinguishable electron densities, the interactions with neighboring atoms generally yield large energy differences between the two rotamer states, where the predominating interactions are hydrogen bonds (Figure 1). Hydrogen bonds and hydrogen bond networks are ubiquitous in protein structures. Their importance in protein folding and stability is clear, but their functional form, their molecular mechanism, and their strength are debated. Hydrogen bonds form when a strongly electronegative nitrogen or oxygen atom (the hydrogen bond donor) shares a covalently attached proton with a lone pair of electrons of another oxygen atom (the hydrogen bond acceptor). In proteins, hydrogen bond donors and acceptors are always an integral part of the polypeptide backbone or the amino acid side chains, and since the respective molecular environment strongly affects the free energy, there are many individual types of hydrogen bonded interactions in proteins (Deechongkit et al., 2004). The derivation of the physical properties of these interactions from first principles is a difficult task (Finnis, 2004), but quantitative models can be obtained from the reversible work theorem of statistical thermodynamics (Chandler, 1987; Witten and Pincus, 2004) in conjunction with experimentally determined protein crystal structures (Sippl, 1990). The theorem states that the distribution of atoms observed in molecular systems is a direct consequence of the underlying molecular forces, thus promoting the quantitative analysis of atom pair interactions on the basis of radial distribution functions (Sippl, 1990, 1993). Here, we compile potentials of mean force for the interactions of Asn/Gln side chain amides with all other protein atoms, except hydrogens, and compare the two possible

*Correspondence: sippl@came.sbg.ac.at

Figure 1. Hydrogen Bond Interactions of Asn Side Chain Amides in Protein Crystal Structures

(A)–(D) show the rotamer states R_1 found in the respective PDB files. The large positive values of $\Delta\varepsilon$ of all rotamers shown indicate that they are strongly disfavored. Dashed lines and the associated numbers (in Å) emphasize distances between atoms in close contact. Atoms are colored by atom type: carbon, gray; oxygen, red; nitrogen, blue. (A) Residue Asn-52 of dethiobiotin synthetase, 1dad (resolution 1.6 Å), where the amide oxygen and nitrogen atoms have unfavorable interactions with the main chain oxygen atoms of Ser-41 and Gly-42, and the backbone nitrogen atoms of Asp-54 and Ala-55, respectively ($\Delta\varepsilon = +35.2$). The interactions of Asn-52 are entirely intramolecular, i.e., all residues shown belong to the same chain. (B) Residue Asn-27 of cutinase, 1cus (resolution 1.25 Å) is in close contact with Val-169 and Lys-168 of a neighboring chain ($\Delta\varepsilon = +15.7$). This is one of the rare cases where crystal contacts between neighboring chains alter the preferred rotamer state. In the isolated chain, $\Delta\varepsilon = -8.0$ favors rotamer R_1. (C) Residue Asn-138 of concanavalin B, 1cnv (resolution 1.65 Å). Interactions with Asn-278 and Arg-281 of the neighboring chain are unfavorable ($\Delta\varepsilon = +23.1$). For the isolated chain, $\Delta\varepsilon = +12.4$, and hence the rotamer is disfavored by intra- as well as intermolecular interactions. (D) Residue Asn-7 of aldose reductase, 1ads (resolution 1.6 Å). The interactions of the amide atoms of this residue do not include a clear hydrogen bonding situation. Nevertheless, the interactions among nitrogen and oxygen atoms disfavor rotamer R_1 relative to R_2. Protein structures are illustrated with the program PyMOL (http://pymol.sourceforge.net/).

rotamers in terms of the respective energy difference. We find a close correspondence between the rotamer flips suggested by the energy differences with expert annotated data.

Theory

The link between structure and energy is provided by the relationship (Chandler,1987; Sippl, 1990):

$$E(a,b,r) = -kT\ln[g(a,b,r)],$$

where $E(a,b,r)$ is the generic potential of mean force for the interaction of two particles of types a and b at separation r, $g(a,b,r)$ is the generic radial two particle distribution function, k is Boltzmann's constant, and T is the absolute temperature. The function $E(a,b,r)$ contains annealed contributions specific for the atom pair (a,b) as well as quenched contributions, which are intrinsic to the covalent structure and compact nature of folded protein molecules. Averaging over all atom pairs (a,b) yields the quenched contributions in form of the unspecific radial distribution function $g(r) = \langle g(a,b,r)\rangle_{a,b}$, whereas the annealed contributions are represented by the specific radial distribution functions $g_s(a,b,r) = g(a,b,r)/g(r)$, and the associated specific potential of mean force is

$$\varepsilon(a,b,r) = -kT\ln g(a,b,r) + kT\ln g(r) = -kT\ln\frac{g(a,b,r)}{g(r)}.$$

In terms of the specific potential of mean force, the interaction at r is attractive if $\varepsilon(a,b,r)<0$ and repulsive if $\varepsilon(a,b,r)>0$.

The total interaction energy of the amide N and O atoms is the sum over pairwise interactions with all other atoms in their molecular environment. Denoting the rotamers found in crystal structures of proteins by R_1 and the associated amide nitrogen and oxygen atoms by N_1 and O_1, this energy is

$$\varepsilon(R_1) = \sum_{j\neq h}\varepsilon(a_h,b_j,r_{h,j}) + \sum_{j\neq k}\varepsilon(a_k,b_j,r_{k,j}),$$

where $h = N_1$ and $k = O_1$ and where the summation is over all atoms j of atom type b_j in the environment of the respective atom. In the alternative rotamer, R_2, the N and O atoms swap their positions and the interaction energy, $\varepsilon(R_2)$, is obtained for $h = N_2$ and $k = O_2$. The energy difference, $\Delta\varepsilon = \varepsilon(R_1) - \varepsilon(R_2)$, of the rotamers indicates whether R_1 or R_2 is the preferred rotamer.

Results and Discussion

Potentials of mean force are compiled from a database of 833 highly resolved protein crystal structures containing 7670 Asn and 6513 Gln residues, and $\Delta\varepsilon$ is computed for the corresponding rotamers (see Experimental Procedures). The complete crystal structure is generated from the respective symmetry operations to recover the complete chemical environment of each amino acid residue. An immediate advantage of potentials of mean force is that the modeling of interactions does not require explicit hydrogen atom positions. On the other hand, it is known in advance that the database contains a substantial fraction of unfavorable rotamers, and it is clear that the potentials are affected by these rotamers. To obtain a self-consistent solution, we employ an iterative refinement protocol where all amides with $\Delta\varepsilon>0$ are flipped from rotamer R_1 to R_2 and where radial distribution functions and potentials are recompiled after correction. The iteration essentially converges within the first refinement cycle (see Experimental Procedures).

Figure 2A compares the distributions of R_1 rotamers, i.e., the rotamers found in the crystal structures, as a function of $\Delta\varepsilon$ before and after refinement. The distribution has two peaks which overlap at $\Delta\varepsilon = 0$. The larger peak resides in the region $\Delta\varepsilon<0$, corresponding to amide groups where the R_1 rotamer is favored by interactions
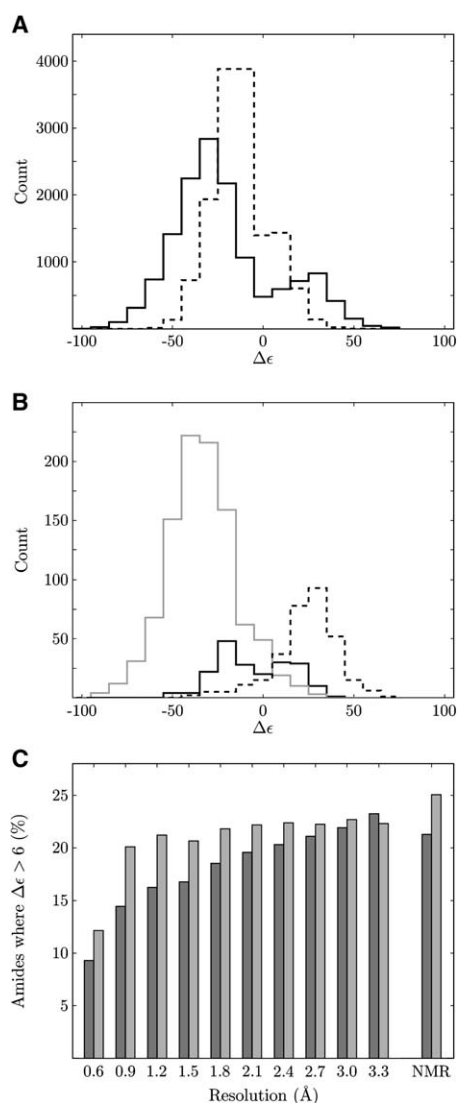
with the chemical environment. The smaller peak lies in the region $\Delta\varepsilon>0$ where R_1 is the disfavored rotamer. Refinement results in a broadening of the distribution due to a general increase of $|\Delta\varepsilon|$ for the individual amides, but most striking is the considerable decrease in rotamer density around $\Delta\varepsilon=0$, corresponding to a stronger discrimination of correct from incorrect rotamers, as compared to the unrefined potentials. The region $-\infty<\Delta\varepsilon<0$ contains 79.0% of the total density, and the respective number for $\Delta\varepsilon>0$ is 21.0%, in close agreement with the error rate of 20% found by Word et al. (1999).

Rotamers defined as correct or incorrect (Word et al., 1999) are in excellent agreement with the respective mean force energy differences $\Delta\varepsilon$, whereas the indistinguishable cases are distributed around $\Delta\varepsilon=0$ with a density distribution resembling the rotamer distribution found in uncorrected protein structures (Figure 2B). We also find perfect agreement between the $\Delta\varepsilon$ values obtained for the 141 lysozyme residues investigated by Higman et al. (2004) and their conclusion that 26 of these residues have incorrect rotamer configurations in the respective crystal structures. If $|\Delta\varepsilon|$ is small or zero, then both rotamers are present in finite amounts, and hence, a classification of rotamers as correct or incorrect has to take into account the occupancies of the rotamers that are obtained from the Boltzmann distribution as $p(R_1)=(1+\exp(\Delta\varepsilon))^{-1}$ and $p(R_2)=1-p(R_1)$. We find that $\Delta\varepsilon\geq6.0$, corresponding to $p(R_1)<0.0024$, is an appropriate conservative threshold for the definition of incorrect rotamers. Similarly $\Delta\varepsilon<-6.0$ defines correct rotamers, and amides that have nonvanishing probabilities for both rotamers are characterized by $-6.0\leq\Delta\varepsilon\leq6.0$.

The amide groups of Asn and Gln side chains on the surface of proteins frequently participate in intermolecular interactions with neighboring molecules. To assess the impact of intermolecular interactions on the preferred rotamer state, we compiled mean force potentials and the respective $\Delta\varepsilon$ values from isolated protein chains completely neglecting interactions with neighboring molecules. Comparing these results with those obtained from the complete crystals, we find only 26 cases of 14,183 (0.2%) where interactions with neighboring molecules cause a clear switch in the preferred rotamer. This surprising result indicates that conformations of amide rotamers are predominantly stabilized by intramolecular interactions. We also find that distribution functions obtained from single molecules and whole crystals are very similar, so that potentials of mean force for intra- and intermolecular interactions are essentially equivalent.

With increasing resolution, the positions of hydrogen atoms emerge from electron density maps, and it is expected that the number of incorrect Asn and Gln rotamers decreases accordingly. In principle it should be possible to correctly assign the proper rotamer at resolutions better than 1.5 Å, but the necessary atomic occupancy refinements and comparisons of alternative

Figure 2. Distributions of Rotamer Energies in Protein Crystal Structures

(A) Distributions of 14,183 Asn and Gln side chain amides found in 833 high-resolution protein crystal structures as a function of $\Delta\varepsilon$ computed from refined (solid line) and unrefined (dashed line) potentials of mean force. The bin size is $|\Delta\varepsilon|=10.0$. Refinement results in a general increase of $|\Delta\varepsilon|$ values and a decrease of the number of rotamers around $\Delta\varepsilon\approx0$. The distribution obtained from the refined potentials has two peaks at $\Delta\varepsilon\approx-30$ and $\Delta\varepsilon\approx+30$.

(B) The distribution of 1557 Asn and Gln side chain rotamers investigated by Word et al. (1999) plotted as a function of $\Delta\varepsilon$, where 1006 residues are classified as correct (gray line), 320 as incorrect (dashed line), and 231 as undecidable (black line). The $\Delta\varepsilon$ values are computed from refined potentials. The distributions of the correct and incorrect rotamers closely match the peaks of the distributions obtained for the 14,183 rotamers in Figure 2A. The distribution of amides corresponding to undecidable cases shows two peaks, resembling the distribution of Figure 2A (solid line). Applying the threshold $|\Delta\varepsilon|>6$ to define correct and incorrect rotamers, as already discussed, and using established terminology (Baldi et al., 2000), we obtain a sensitivity of 92.7%, a specificity of 96.7%, and an overall accuracy of 95.8% relative to the classification of Word et al. (1999). The fraction of amides occupying both rotamer states with significant probability, defined by $|\Delta\varepsilon|\leq6.0$, is 5.8%.

(C) Fraction of incorrect amide rotamers ($\Delta\varepsilon>6.0$) plotted as a function of resolution. Solid bars in dark and light gray correspond to Asn and Gln residues, respectively, where the height of each bar corre-

sponds to the fraction of amides with $\Delta\varepsilon>6.0$ in the respective bin, where the bin size is 0.3 Å. The total number of amides contained in 54,189 protein chains derived from 22,574 protein crystal structures is larger than one million. The bars on the right correspond to the respective fraction of Asn and Gln rotamers found in 3,564 NMR structures.
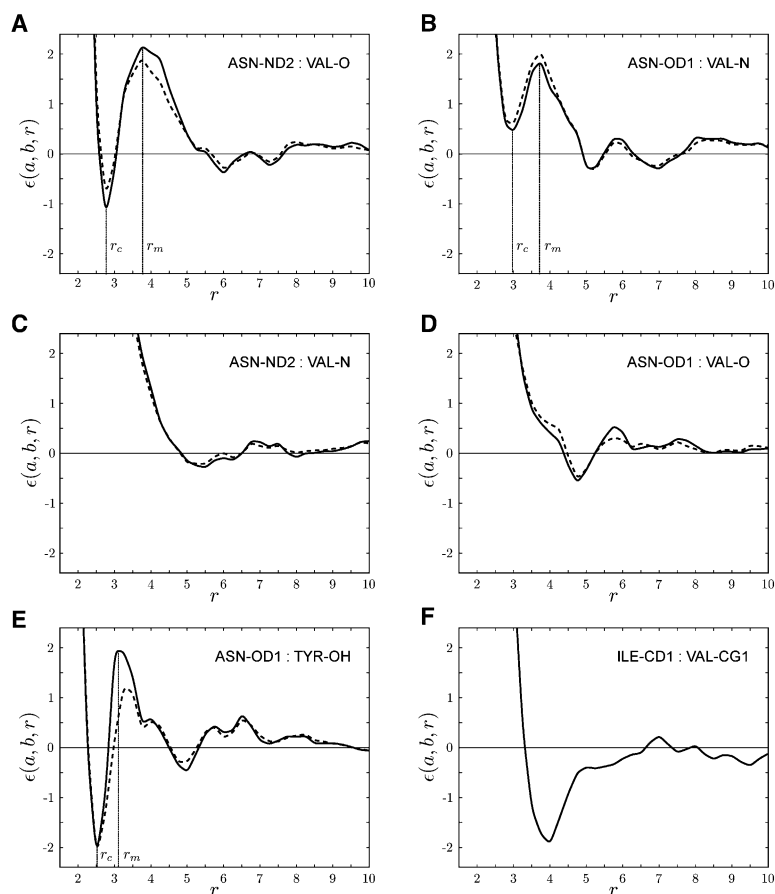
Figure 3. Potentials of Mean Force Compiled from Protein Crystal Structures

Refined (solid lines) and unrefined (dashed lines) potentials of mean force, $\varepsilon(a,b,r)$, between atoms of type a and b are plotted as a function of separation r in the distance range $1.5 \leq r \leq 10.0$ Å. All potentials converge to zero at separations larger than $r \geq 15$ Å. Hydrogen bond contact distances, r_c, and the location of the maximum of the hydrogen bond barrier, r_m, are shown for the hydrogen bond interactions (A, B, and E). The atom type codes correspond to Asn side chain amide oxygen (ASN-OD1), Asn side chain amide nitrogen (ASN-ND2), valine (Val) backbone oxygen (VAL-O), Val backbone nitrogen (VAL-N), tyrosine (Tyr) side chain hydroxyl oxygen (TYR-OH), isoleucine (Ile) carbon atom (ILE-CD1), and Val carbon atom (VAL-CG1). (A) Interaction of the hydrogen bond donor ASN-ND2 and hydrogen bond acceptor VAL-O. (B) Interaction of the hydrogen bond acceptor ASN-OD1 and hydrogen bond donor VAL-N. (C) Interaction of the hydrogen bond donor ASN-ND2 and the hydrogen bond donor VAL-N. (D) Interaction of the hydrogen bond acceptor ASN-OD1 and the hydrogen bond acceptor VAL-O. (E) Interaction of the hydrogen bond acceptor ASN-OD1 and TYR-OH, the hydroxyl oxygen of tyrosine, which acts simultaneously as a hydrogen bond donor and acceptor. (F) Interaction of the aliphatic carbon atom ILE-CD1 and the aliphatic carbon atom VAL-CG1. All hydrogen bond interactions (A, B, and E) have a large energy barrier and a narrow energy well in the range $2.7 \leq r \leq 3.0$, corresponding to hydrogen bond contact, but the depth of the minima vary considerably. In particular, the interaction of ASN-OD1 with VAL-N (B) has a minimum where $\varepsilon(a,b,r) > 0$, indicating that the formation of this bond consumes free energy. The shallow minima near 5 Å for the repulsive interactions (C and D) originate from the covalent structure of the side chain amide. For example, whenever a hydrogen bond contact is formed between ASN-ND2 and VAL-O (A), the neighboring oxygen atom ASN-OD1 necessarily is in the distance range 5.0–6.0 Å from the VAL-O atom (C). The attractive hydrophobic interaction (F) has no barrier.

conformations are generally not contained in standard crystallographic refinement protocols. The dependency of rotamer error rate on resolution is revealed by the $\Delta\varepsilon$ values computed for more than a million Asn and Gln residues found in 54,189 protein chains contained in 22,574 protein crystal structures. We exclude Asn and Gln side chains whose amides are closer than 8 Å to nonprotein atoms since potentials of mean force for these interactions are currently not available. As expected, the number of incorrect rotamers decreases with increasing resolution, but even at resolutions better than 1 Å, the average error rate is above 10% (Figure 2C).

The amide assignment problem originates from X-ray analysis of protein crystal structures, but protein structures determined by nuclear magnetic resonance (NMR) techniques have similar error rates of 21% for Asn and 25% for Gln residues (Figure 2C). Protein crystal structures and protein NMR structures are generally derived from experimental constraints where refinement protocols are used to remove errors or to supply information that cannot be obtained from the respective experiment. These refinement protocols frequently employ scoring functions that are based on various models of molecular interactions like Lennard-Jones and Coulomb potentials. Potentials of mean force are based on the principles of statistical thermodynamics. Their functional form and variation is entirely determined by the underlying experimental data.

The specific potentials of mean force for the interactions of Asn and Gln amide nitrogen and oxygen atoms with other hydrogen bond donors and acceptors have the form of a molecular lock or shutter, characterized by an energy barrier with a maximum at r_m separating a narrow energy well at hydrogen bond contact, r_c, from large distances, r_∞ (Figure 3). The molecular lock is a general model for spatially precise and kinetically stable atom-pair interactions where precision is determined by the tightness of the energy minimum at bond contact, r_c, and stability is mediated by $\varepsilon_s = \varepsilon(r_m) - \varepsilon(r_c)$, the difference between the height of the energy barrier at r_m and the depth of the energy well at r_c. A consequence of the energy barrier is that stable bonds can be formed even if work is expended in the formation of bond contacts (Figure 3B). The barrier creating the lock is a specific feature of hydrogen bonds and other highly polar interactions (Sippl, 1996). Other potentials have comparatively broad attractive basins, and they lack barriers. Typical examples are hydrophobic interactions (Chandler, 2005) between aliphatic carbon atoms (Figure 3F). Such interactions are less precise since the range of preferred

distances is considerably larger as compared to hydrogen bond interactions, and their stability is solely determined by $\varepsilon(r)$.

Contrasting potentials of mean force before and after rotamer correction, we find that the major consequence of refinement is a sharpening of the minima at r_c and an increase of energy barriers at r_m (Figure 3). Compared to the rate of 20% of unfavorable rotamers in the original crystal structures, the effect on the potentials is rather small, but these small changes add constructively so that the refinement results in an overall doubling of the absolute values of the energy differences $|\Delta\varepsilon|$ between rotamers (Figure 2). The resulting potential functions can be used to validate and refine Asn/Gln rotamers in protein crystal structures, which is available as a public web service (http://flipper.services.came.sbg.ac.at/).

Experimental Procedures

Compilation of Potentials of Mean Force
Radial distribution functions are obtained from the number densities $n(a,b,r)$ found in protein crystal structures, where a and b are atom types and r is the distance between atoms (Sippl et al., 1996; Sippl, 1996). Distances are sampled in bin sizes of 0.25 Å with sparse data corrections for low counts (Sippl, 1990). The number densities are compiled from a library of 833 protein chains, containing 7,670 asparagine and 6,513 glutamine residues, i.e., a total of 14,183 side chain amides. The resolution of these structures is better than 1.6 Å, and their R factor is smaller than 0.25. Any two protein chains in this library have less than 20% sequence identity (Wang and Dunbrack, 2003). The radial distribution functions are defined as $g(a,b,r)=n(a,b,r)/n(a,b)$ where $n(a,b)=\sum_r n(a,b,r)$. The quenched distribution function is defined as $g(r)=\sum_{a,b} n(a,b,r)/\sum_{a,b,r} n(a,b,r)$. The specific potentials of mean force are obtained from the radial distribution functions as $\varepsilon(a,b,r)=-kT\ln[g(a,b,r)/g(r)]$ (k, Boltzmann's constant; T, temperature). Energies are reported in units of $kT=1$. The atom types a and b correspond to the 167 individual atom types of standard amino acids found in PDB files (Berman et al., 2000), excluding hydrogen atoms. In particular, the backbone atoms N, C$^\alpha$, C$'$, and O of individual amino acids are distinct atom types. The potentials are symmetric, $\varepsilon(a,b,r)\equiv\varepsilon(b,a,r)$, so that the total number of distinct pairwise atomic interactions is $167\times166/2=14,028$.

Refinement of Potentials of Mean Force
Potentials are compiled from the protein crystal structures, and the energy differences $\Delta\varepsilon$ of the two rotamers of all Asn and Gln residues are computed. All rotamers with $\Delta\varepsilon>0$ are flipped to the alternative rotamer, and mean force potentials are recompiled from the altered protein crystal structures. The compilation of potentials and the consecutive flipping of high-energy rotamers ($\Delta\varepsilon>0$) constitutes a single refinement cycle. The refinement process is repeated as long as rotamer flips are observed. The first refinement cycle flips 2,802 of the 14,183 Asn and Gln residues. This number drops to 129 in the second cycle. Very few rotamers are flipped in subsequent iterations, and the whole set of rotamers is stable after 14 refinement cycles.

Implementation of Asn/Gln Rotamer Correction
The Asn and Gln rotamer correction is available as a web service, accessible at http://flipper.services.came.sbg.ac.at/. The service, implemented in Python and C, has a response time in the order of seconds and accepts PDB codes or protein structure files in PDB format.

Acknowledgments

References

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A., and Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16, 412–424.

Battiste, J.L., Mao, H.Y., Rao, N.S., Tan, R.Y., Muhandiram, D.R., Kay, L.E., Frankel, A.D., and Williamson, J.R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide RRE RNA complex. Science 273, 1547–1551.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. 28, 235–242.

Chandler, D. (1987). Introduction to Modern Statistical Mechanics (Oxford, UK: Oxford University Press).

Chandler, D. (2005). Interfaces and the driving force of hydrophobic assembly. Nature 437, 640–647.

Coleman, D.E., Berghuis, A.M., Lee, E., Linder, M.E., Gilman, A.G., and Sprang, S.R. (1994). Structures of active conformations of Gi alpha 1 and the mechanism of GTP hydrolysis. Science 265, 1405–1412.

Deechongkit, S., Nguyen, H., Powers, E.T., Dawson, P.E., Gruebele, M., and Kelly, J.W. (2004). Context-dependent contributions of backbone hydrogen bonding to beta-sheet folding energetics. Nature 430, 101–105.

Evdokimov, A.G., Anderson, D.E., Routzahn, K.M., and Waugh, D.S. (2001). Structural basis for oligosaccharide recognition by Pyrococcus furiosus maltodextrinbinding protein. J. Mol. Biol. 305, 891–904.

Faham, S., Hileman, R.E., Fromm, J.R., Linhardt, R.J., and Rees, D.C. (1996). Heparin structure and interactions with basic fibroblast growth factor. Science 271, 1116–1120.

Finnis, M. (2004). Interatomic Forces in Condensed Matter (Oxford, UK: Oxford University Press).

Higman, V.A., Boyd, J., Smith, L.J., and Redfield, C. (2004). Asparagine and glutamine side-chain conformation in solution and crystal: a comparison for hen egg-white lysozyme using residual dipolar couplings. J. Biomol. NMR 30, 327–346.

Hooft, R.W., Sander, C., and Vriend, G. (1996). Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. Proteins 26, 363–376.

McDonald, I.K., and Thornton, J.M. (1995). The application of hydrogen bonding analysis in X-ray crystallography to help orientate asparagine, glutamine and histidine side chains. Protein Eng. 8, 217–224.

Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A., Riekel, C., Grothe, R., and Eisenberg, D. (2005). Structure of the cross-beta spine of amyloid-like fibrils. Nature 435, 773–778.

Raymond, A.C., Rideout, M.C., Staker, B., Hjerrild, K., and Burgin, A.B. (2004). Analysis of human tyrosyl-DNA phosphodiesterase I catalytic residues. J. Mol. Biol. 338, 895–906.

Sambashivan, S., Liu, Y., Sawaya, M.R., Gingery, M., and Eisenberg, D. (2005). Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. Nature 437, 266–269.

Schafer, K., Magnusson, U., Scheffel, F., Schiefner, A., Sandgren, M.O., Diederichs, K., Welte, W., Hulsmann, A., Schneider, E., and Mowbray, S.L. (2004). X-ray structures of the maltose-maltodextrin-binding protein of the thermoacidophilic bacterium alicyclobacillus acidocaldarius provide insight into acid stability of proteins. J. Mol. Biol. 335, 261–274.

Sippl, M.J. (1990). Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. J. Mol. Biol. 213, 859–883.

Sippl, M.J. (1993). Recognition of errors in three-dimensional structures of proteins. Proteins 17, 355–362.

Sippl, M.J. (1996). Helmholtz free energy of peptide hydrogen bonds in proteins. J. Mol. Biol. *260*, 644–648.

Sippl, M.J., Ortner, M., Jaritz, M., Lackner, P., and Flockner, H. (1996). Helmholtz free energies of atom pair interactions in proteins. Fold. Des. *1*, 289–298.

Vernet, T., Tessier, D.C., Chatellier, J., Plouffe, C., Lee, T.S., Thomas, D.Y., Storer, A.C., and Menard, R. (1995). Structural and functional roles of asparagine 175 in the cysteine protease papain. J. Biol. Chem. *270*, 16645–16652.

Wang, G., and Dunbrack, R. (2003). PISCES: a protein sequence culling server. Bioinformatics *19*, 1589–1591.

Witten, T., and Pincus, P. (2004). Structured Fluids: Polymers, Colloids, Surfactants (Oxford, UK: Oxford University Press).

Word, J.M., Lovell, S.C., Richardson, J.S., and Richardson, D.C. (1999). Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J. Mol. Biol. *285*, 1735–1747.

Yoder, M.D., Keen, N.T., and Jurnak, F. (1993). New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. Science *260*, 1503–1507.