Published in final edited form as:
Comput Biol Chem. 2010 June ; 34(3): 172–183. doi:10.1016/j.compbiolchem.2010.06.002.

Fine Grained Sampling of Residue Characteristics Using Molecular Dynamics Simulation

Hyun Joo[*],
Chemistry Department, University of the Pacific, 3601 Pacific Avenue, Stockton, CA 95211

Xiaotao Qu[*],
Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612

Rosemarie Swanson,
Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843

C. Michael McCallum, and
Chemistry Department, University of the Pacific, 3601 Pacific Avenue, Stockton, CA 95211

Jerry Tsai
Chemistry Department, University of the Pacific, 3601 Pacific Avenue, Stockton, CA 95211

Hyun Joo: hjoo@pacific.edu; Xiaotao Qu: xiaotao.qu@moffitt.org; Rosemarie Swanson: rosmar@tamu.edu; C. Michael McCallum: mmccallum@pacific.edu; Jerry Tsai: jtsai@pacific.edu

Abstract

In a fine-grained computational analysis of protein structure, we investigated the relationships between a residue's backbone conformations and its side-chain packing as well as conformations. To produce continuous distributions in high resolution, we ran molecular dynamics simulations over a set of protein folds (dynameome). In effect, the dynameome data set samples not only the states well represented in the PDB but also the known states that are not well represented in the structural database. In our analysis, we characterized the mutual influence among the backbone, φ, ψ angles with the first side-chain torsion angles (χ_1) and the volumes occupied by the side chains. The dependencies of these relationships on side-chain environment and amino acids are further explored. We found that residue volumes exhibit dependency on backbone 2° structure conformation: side-chains pack more densely in extended β-sheet than in α-helical structures. As expected, residue volumes on the protein surface were larger than those in the interior. The first side-chain torsion angles are found to be dependent on the backbone conformations in agreement with previous studies, but the dynameome data set provides higher resolution of rotamer preferences based on the backbone conformation. All three *gauche*$^-$, *gauche*$^+$, and *trans* rotamers show different patterns of φ, ψ dependency, and variations in χ_1 value are skewed from their canonical values to relieve the steric strains. By demonstrating the utility of dynameomic modeling on the native state ensemble, this study reveals details of the interplay among backbone conformations, residue volumes and side-chain conformations.

Correspondence to: Jerry Tsai, jtsai@pacific.edu.

[*]These authors contributed equally to this work

Keywords

molecular dynamics simulation; residue volume; side-chain packing; dynameome; backbone conformation; rotamer; Ramachandran plot

1 Introduction

During the past two decades, the progress in protein structure prediction from amino acid sequence has been made with simple representations of side chains, for example, as 3 rotamers according to the first torsion angles (χ_1) of side-chains (Moult et al. 2007; Moult et al. 2009; Moult et al. 2005; Moult et al. 2001; Moult et al. 2003; Moult et al. 1997; Moult et al. 1999). It has been proposed that the local backbone conformation has the greatest influence in determining side-chain conformation (Samudrala and Moult 1998). Given the native backbone conformation, accurate packing of side chains can be achieved (Chung and Subbiah 1995). In contrast, the side-chain conformation also affects the backbone conformation (Chakrabarti and Pal 1998). As a step to improve prediction accuracy, we pursue a higher resolution description of the relationship between backbone and side-chain conformations. In particular, our study seeks to understand the determinants of this relationship better.

The most widely used description relating backbone and side-chain conformation are rotamer libraries (Chakrabarti and Pal 2001; Chung and Subbiah 1996; Dunbrack 2002; Shapovalov and Dunbrack 2007). A rotamer library clusters the observed conformations of side-chains into groups, from which Bayesian distributions can be derived (Dunbrack and Cohen 1997; Shapovalov and Dunbrack 2007). Populated rotamers are thought to reflect local minima on a potential energy surface or to represent an average conformation over some region of dihedral angle space (Dunbrack 2002). Even though recent rotamer libraries have benefited from the increased number of structures (especially high-resolution structures) in the PDB (Krivov et al. 2009; Wang and Dunbrack 2003), these libraries' coverage of conformational space is still limited by both sampling the conformational space from the PDB and the fact that the structures in the PDB are closely clustered around the minimum-energy X-ray crystal structures only. Furthermore, some structures deposited in the PDB are not able to distinguish one conformation from the others, particularly when the resolution is low (MacArthur and Thornton 1999). On one hand, broad distributions of side-chain dihedral angles are often observed(Dunbrack 2002). On the other hand, many rotamer conformations that can be accommodated by residues, such as those on the protein surface, are highly under-represented in crystallographic structures (MacArthur and Thornton 1999; West and Smith 1998; Zhao et al. 2001). Thus, sampling side-chain conformations from a continuous conformational space would provide higher accuracy.

Secondly, to reduce system complexity as well as to complete the conformational sampling, rotamer libraries bin side-chain conformations based on the three most populated rotamer conformations around the Cα atom: *gauche*$^+$, *gauche*$^-$ and *trans*. In addition, the backbone conformation is also binned to discrete areas of secondary structure space. By defining side-chain conformations in this way, rotamer libraries decrease the combinatorial complexity of packing/placing side chains in protein structure prediction. The result of this approximation is that rotamer libraries decrease the description resolution of the relationship between backbone and side-chain conformation. Suggested library improvements include adding extra information, such as a side-chain-orientation-dependent term (Skolnick 2006) or the addition of solvated rotamers, in which several water molecules accompany the rotamer (Jiang et al. 2005). Moreover, a refined rotamer library, in which only high resolution non-clashed side-chains are included with smaller and more continuous bins, has greatly

improved the accuracy over other rotamer libraries (Lovell et al. 2000) However, current approaches using rotamer libraries are reaching their limits (Levitt et al. 1997a; Shapovalov and Dunbrack 2007) in predicting high-resolution structures.

We have undertaken a study that provides a more detailed description of the relationship between a residue's backbone conformation and its side chain. To produce a more complete view of the native state conformation space, we follow the approach of previous work (Day and Daggett 2003) and generate a dataset of molecular dynamics (MD) simulations over a set of protein folds (dynameome) (Beck et al. 2008b). These dynameomic approaches have been shown to accurately sample the structures near the native conformation across different protein folds and reproduce the ensemble properties of the native state environment (Benson and Daggett 2008; Jonsson et al. 2009; Kehl et al. 2008; Rueda et al. 2007a; Rueda et al. 2007b; Simms et al. 2008). Therefore, the purpose of the dynameome data set is to model a more continuous set of native state conformations, as opposed to the classic use of molecular dynamics simulations for time dependent information like kinetics. Containing over a million structures, this dynameome data set allows a more refined view of protein structure. Specifically, we investigate the mutual dependence of backbone conformation (φ, ψ), the volume occupied by residues and the first side-chain torsion angle (χ_1). Our analysis finds that side-chain volumes exhibit a somewhat counterintuitive dependence on secondary structure. Increasing the resolution over previous analyses (Dunbrack 2002; Lovell et al. 2000), we detail the backbone's influence on each of the 3 χ_1 rotamer angles. We also investigate the effect that χ_1 has upon residue volume. For each of these analyses, we discuss the physical basis responsible for their structural propensities

2 Methods

2.1 Dataset

A total of 2 sets of protein structures were used in this analysis: one for MD simulations and one to calculate residue volumes from available protein structures. For the MD simulations, a dataset of 77 protein folds was generated in the following manner (See Table 1 for a detailed list of PDB codes.). An initial list of structures was obtained using the PISCES server (Wang and Dunbrack 2003) on the May 2003 release of the Protein Data Bank or PDB (Berman et al. 2002) and the following criteria: sequence identity less than 20%, resolution smaller than 1.6, and R factor smaller than 0.25. This initial list was winnowed down based on continuous chains. While structures with missing N or C termini were included, the selected structures were required to have all heavy atoms from the N terminus to the C terminus. Those structures with missing internal residues or even missing side chains were excluded. Lastly, due to ENCAD memory restrictions primarily with the number of solvating waters (Levitt et al. 1995), the structures were limited to around 150 residues. This resulted in the final 77 protein folds listed in Table 1. Because we wanted to focus our analysis on a study of the distributions and fluctuations of residues in tertiary structure, we only used the monomeric form or single domain form of the protein in this study. For those proteins that are part of a multimeric group, we extracted only 1 chain as indicated by the capital letter at the 5th position of the PDB code in Table 1. Although the 2 Å CαRMSD cutoff limits the total data size, it ensures sampling of only near-native behavior. A second set of protein structures was selected to demonstrate the sampling of the PDB in Figure 1 and is named the PDB set. Again, we used the PISCES server on the May 2003 PDB using criteria to ensure we sampled different folds: sequence identity less than 5%, resolution less than 1, and R factor smaller than 0.25. To make the comparison as similar as possible, only monomeric structure solved by X-ray crystallography were considered. This resulted in a final set of 689 structures, a list of which is available upon request.

2.2 MD simulation

We ran 5 independent 10 nanosecond (ns) MD simulation on each protein using the ENCAD program (Levitt et al. 1995) and the F3C explicit water model (Levitt et al. 1997b). The ENCAD program and the associate force-field provide a useful means to approach this problem, as it does not suffer from some of the problems that the CHARMM and AMBER force-fields exhibited and which have since been corrected.[63;64] The ENCAD suite has been used successfully and recently in many applications including folding/unfolding studies[65–68] and replica-exchange studies.[69] In addition, some comparisons have been made between different force-fields.[68]

For each simulation, the coordinates of each structure were placed in a box of water and then energy-minimized. Each box of water was trimmed so that the edges were at least 8 Å away from the closest protein atom. All waters within 1.67 Å of the protein were removed, and the box sides were corrected to match the density of water (0.997 g/ml) at 298 K. Sodium or chloride ions replaced water molecules at random positions to yield an electrically neutral system. Conjugate gradient energy minimization was performed in the following order: The protein was fixed while the water molecules were minimized over 1,000 steps. The protein was then minimized in the next 1,000 steps, holding the water molecules fixed. Finally, the entire system was minimized over 1,000 steps. To begin each of the simulations from a unique starting point, the system was equilibrated to 298 K using a different random-seed number to assign initial velocities. During the calculations, the coordinates of the structure were updated at two femtosecond intervals and sampled every picosecond (500 steps), such that each 10 ns simulation generated 10,000 steps. All simulations are summarized in Table 2. The largest simulation has a water box of 60.5 Å by 55.1 Å by 50.0 Å in size and 4590 water molecules around a 122-residue protein 1QTO (Kawano et al. 2000) while the smallest simulation has a water box of 40.5 Å by 28.0 Å by 34.4 Å in size and 1211 water molecules around a 21-residue protein 1G7A (Smith et al. 2001).

2.3 Data analysis

Because the initial steps in the simulation equilibrate the system to 298 K, we decided to disregard the first 1 ns of the simulation and performed analysis using only the last 9 ns of the simulation (1–10 ns). This will ensure the removal of any possible experimental anomalies such as artificially extended side chains or backbone strain. Programs written in C and PERL were created to analyze the native ensemble of structures. Coordinates were viewed using PyMol (DeLano 2002). Figures were generated using the R statistical package (Becker et al. 1988). To enforce consistency of sampling only around the native conformation, the folds are further selected based on the criteria that the averaged C RMSD is below 2 Å for each protein.

2.3.1 Secondary Structure Assignment—For each structure output from the simulation, the secondary structure of the protein was defined using PROMOTIF (Hutchinson and Thornton 1996) and categorized in the following manner. Residues without any assignment were assigned to the random coil (C) class. Both β-turns and G-turns were combined as turn (T). All the helices were classified as (H). Strand and β-bulges were combined as extended strand (E). Because PROMOTIF defines secondary structure primarily on hydrogen bond patterns instead of torsion angles (Kabsch and Sander 1983), secondary structure classes have a broad definition. As a result, a small number of residues erroneously assigned secondary structure classes that do not match their backbone conformations, and for this reason were left out of the analysis.

2.3.2 Volume Calculations—The volumes were calculated using the Voronoi Polyhedra method (Voronoi 1908) for heavy atoms, which is explained in more detail in a previous

study (Gerstein et al. 1995). For each residue, the total volume of this residue is the sum of each atom's volume that is in contact with atoms belonging to another residues. Only contacts with surface area larger than 1 $Å^2$ were considered. The residue directly contacting the water solvent is defined as an exposed residue, while buried residues were those that only contacted protein. For each residue, the total volume is the sum of each atom's volume.

To compare volumes between residues in the dynameomic data set, we normalized all 20 residue volumes to a common scale: percentage of mean volume or vol%. First, mean volumes, <*vol*>, for each 20 amino acids were calculated over the whole dynameome data set. As shown in (1), the vol% is derived by dividing the volume, *vol*, of a residue in a particular structure and at a particular time step by the respective residue's <*vol*>.

$$vol\% = \frac{vol}{<vol>} \times 100 \quad (1)$$

The residue volume plots in Figures 1 and 2 show mean <vol%> averaged over vol% values at a particular backbone conformation (φ, ψ) over certain sets of residues and/or conditions like secondary structure and exposure to solvent. Only values above 300 counts cutoff were plotted.

For the volume calculations of the experimental PDB dataset, each structure was placed in the center of a water box, since the Voronoi procedure cannot calculate the volumes of the exposed residues with undefined neighbors. This water box was taken from a MD simulation of pure water using the same parameters as in the protein simulations. Duplication of water box is applied if necessary to generate large enough box for protein. Any water atom within a distance of 1.8Å of protein atoms was removed. Volumes, torsion angles are calculated using the same method for simulated structures as described above. The same approach was used to plot PDB data as the dynameomic data. Mean volumes were calculated for each type of residue and the average percentage of mean volume was plotted as it is done in Figure 1b. A count cutoff of 250 was used and a different backbone bin size was used due to the sparcity of the data (see below).

2.3.3 Calculation of Torsion angles—The, φ, ψ, and χ_1 values for each residue in every structure were calculated using PROMOTIF and values were rounded up to the next a whole number. In effect, we used 1° bins for the dynameomic data and a 5° bins for the data from the PDB. The χ_1 values (except those from PRO, ALA and GLY) were classified using similar nomenclature to a previous study(Lovell et al. 2000). As a simplification, **M**, **P** and **T** are used to refer the 3 χ_1 rotamers. **M** stands for *gauche minus* where $-120° < \chi_1 < 0°$, **P** stands for *gauche plus* conformation where $0° < \chi_1 < 120°$, and **T** stands for *trans* conformation where $120° < \chi_1 < 240°$. In the case of conformationally restricted PRO, we used the convention that **P** was any $\chi_1 < 0°$ and M was any $\chi_1 > 0$ (Dunbrack and Cohen 1997). For ILE and THR, since χ_1 is defined differently than other residues, the calculated χ_1 values were translated to reflect corresponding Cγ atoms in other residues by subtracting 120°. These translated values were then evaluated as **M**, **P**, or **T** as defined above. The plots in Figure 4 were made using certain criteria. For the χ_1 rotamer population plots, a count cutoff of 300 was used. At each backbone conformation, the values over the 3 χ_1 rotamers add up to 1 or 100%. For example, at the φ, ψ value of −60°, −40° in the α-helical region, **M** population is 29%, **P** is 1% and **T** is 70%, which adds up to 1. The distribution of vol% versus χ_1 angle required that the counts be on a log scale. The count cutoff was 100. While the bin size for χ_1 values was 1° (as explained above), the bin size for vol% is 0.5%.

3 Results and Discussion

3.1 The dynameome dataset

The purpose of the dynameome dataset is to provide a more complete sampling of native protein conformational space (instead of the usual kinetic properties measured in MD simulations). As a first step, a set of structures was chosen to broadly represent all protein folds using the PISCES server (Wang and Dunbrack 2003). Using the SCOP (Murzin et al. 1995) classification (Table 1), the set of 77 structures consists of 25 α-helical proteins, 15 β-sheet; 27 are mixed α/β, and 10 belong to the "other" classification. The largest structure (1AKR (O'Farrell et al. 1998)) is an α/β protein with 147 residues, while the smallest one (1G7A (Smith et al. 2001)) has 21 residues and is classified as a small protein in SCOP (Murzin et al. 1995). The average size is 93 residues. To insure that the MD simulations sampled near native conformations, structures with an averaged CαRMSD below 2Å from their starting structures were used (see Methods). This cutoff reduces artifacts from non-native conformations but ensures the high resolution sampling. The dynameome drifts on average 1.8Å CαRMSD from the native structure with a standard deviation 0.1Å per fold. Such a small deviation demonstrates that our dynameome dataset samples conformational space close to native structures only. Within this RMSD, we were able to sample about a million structures for the analysis of the ensemble averaged properties of the native state. Specific statistics for each protein fold are summarized in Table 2. Our analysis focused on the interdependence between the backbone state, residue packing, and side-chain conformation.

3.2 Average side-chain residue volumes of the 20 amino acids

In Table 3, the average residue volumes were measured over the dynameome for each of the 20 amino acids and compared to the residue volumes calculated from the ProtOr standard set of protein atom volumes(Tsai et al. 1999). As expected, the average residue volumes calculated from the dynameome are larger than the ProtOr set on average by about 3%. It has been shown that residues are more regularly packed when they are buried deeper in the protein, which results in smaller volumes, as opposed to the heterogeneous packing at the protein/water surface, which results in larger volumes (Tsai and Gerstein 2002;Tsai et al. 1999). As contrasted in the middle columns of Table 3, residue volumes are smaller by about 4% on average when buried than when exposed. Closer examination of buried residue volumes of dynameome set shows a close match to the ProtOr volumes, but there are some notable differences between the two sets. The CYS, TRP, PHE, and MET residues are significantly larger, whereas the charged ASP, GLU, SER, and THR are smaller. The largest volume difference comes from CYS volume. The volume of CYS from the ProtOr set is 16 $Å^3$ smaller than that from dynameome, corresponding to 13% of its average volume. The primary factor for this difference is that the CYS residues used to define the ProtOr set were mostly disulfide bonded (Tsai et al. 1999), which significantly reduces a CYS residue's volume. To sample wide range of side-chain torsion angles, all disulfide bonds are reduced to –SH in this data set and results in larger CYS volumes than those derived from crystal structures. MET volumes are 8% larger in dynameomic result reflecting the dynamic effect of high degree of freedom around sulfur atom of the long aliphatic side chain. The dynameome volumes for negatively charged ASP and GLU are 13% and 11%, smaller, respectively, than the ProtOr volumes. Since these are all buried, we find that they form strong hydrogen bonds which contribute to packing efficiency of these negatively charged residues(Kuntz 1972;Schell et al. 2006). Along with negatively charged residues, SER and THR show smaller dynameome mean volumes than ProtOr volumes. All have partially charged or highly electronegative oxygen atoms which can participate in shorter hydrogen bonding due to the way ENCAD force field has been parameterized. This explains smaller residue volumes of residues with hydroxyl group and carboxyl groups in dynameome set.

However, the remaining residues deviate by an average of less than 3% from the ProtOr values. Therefore, the dynameome's buried volumes are generally consistent with the ProtOr volumes calculated from crystal structure data. This result supports the idea that our dataset is a good approximation of near native conformation.

The average volumes of four types of secondary structure are also presented in Table 3. When comparing an individual residue's volume across secondary structure, two interesting features are observed. First, there is no significant difference between the volumes associated with different secondary structures. It is commonly assumed that residues in α-helices (H) and β-strands (E) pack well; in turns (T) moderately well; and in coils (C) more loosely. However, Table 3 shows a maximum residue volume variation within secondary structure of only about 8% for CYS, and the average difference between secondary structures is only 1%. Such small volume differences suggest that packing is not optimized for helices or sheets over other secondary structures. The average standard deviation is about 7% of the mean volume, indicating that the dynamics of side-chain motion is not affected much by backbone conformation. The second feature is that these small differences show a different order to how well secondary structure packs residues. Even though the volume differences reported in Table 3 between secondary structures are generally small and are within the variation, the size of our dataset strongly supports that even these minor differences are meaningful. There is a general trend that residues in strands exhibit the smallest volumes followed by coils/turns and those attached to helices are usually the largest. On average, comparing strand with helix, a residue in a strand occupies only 98% as much volume as the same residue in a helix. If we assume that smaller volumes indicate denser packing while the larger volumes looser, then our results demonstrate that sheets pack best, followed by turns/coils and lastly by helices. This ordering is somewhat counterintuitive since it is generally accepted that the helical and coil backbones pack the tightest, whereas sheet and turn backbones less well. Yet, when including the full residue's side chain, it makes sense that regular sheet structures allow tighter residue packing than helical cylinder with side-chains extended spirally.

3.3 Volume variation with backbone conformation (φ, ψ)

The residue volume dependence on backbone φ, ψ torsion angles is plotted in Figure 1. Residue volumes were "normalized" for comparisons by expressing them as percent of the corresponding amino acid's mean volume (vol%; see Methods for details). Using a grey scale, darker color indicates larger than average volumes (looser packing) and lighter color indicates smaller than average volumes (tighter packing). Figure 1a plots the vol% versus, from 408 experimentally determined structures selected from the PDB (Wang and Dunbrack 2003) (see Methods for details). This distribution from the PDB data does not show continuous distribution even with the interpolation performed by the R statistical package (Becker et al. 1988).

The dynameome dataset (Figure 1b) was able to reproduce this distribution exhibited by the experimental PDB data with much high resolution. Consistent with other studies(Feig 2008; Griffiths-Jones et al. 1998), an increased population in α_L region was observed. The residues in this region are lacking regular secondary structure and mostly exposed to solvent, but these residues are believed to be critical for β-sheet structure (Minor and Kim 1994). Furthermore, the dynameome data is consistent with other Ramachandran analysis (Ho et al. 2003; Mandel et al. 1977; Ramachandran et al. 1963; Ramachandran and Sasisekharan 1968) by showing no values in strongly disallowed regions such as the blank region around φ = 0 (Ho et al. 2003). The plot in Figure 1b clearly depicts the smooth dependency of residue volumes on backbone conformation. The connected region between right handed helical and sheet region is also seen. This ability to sample over many possible conformations in the native ensemble allows us to study the characteristics of native

structure in greater detail, supporting the idea that our dynameome dataset is an approximation of the near native ensembles. In contrast to the sparse sampling from experimental structures, the dynameome's broad sampling of protein structures, as seen in recent publications using MD simulation to study backbone conformation propensities (Beck et al. 2008a; Feig 2008), produces a far smoother distribution. The range of residue volumes from the PDB is broader than from the dynameome, since the approximation of a static water box for solvent (see Methods) would produce larger volumes for surface residues of the PDB.

3.4 Volume variations with different packing environment

The volumes of buried and exposed residues on backbone conformations are shown in Figures 2a and 2b, respectively. The buried residues generally occupy smaller space and pack tighter than exposed residues (Figure 2a and 2b). For buried residues, Figure 2a shows very limited sampling of φ, ψ space, populating only the regions of well-defined secondary structure near the center of the sheet region and helical regions. Furthermore, the buried sheet region packs even tighter than exposed sheet region. On the other hand, exposed residues in Figure 2b exhibit the same range of sampling as seen over the dynameome in Figure 1b. Figure 2a and 2b shows the non-local environment influences residue volumes and packing. In addition to the tighter interior packing, the very restricted conformational space sampled by buried residues in this study suggests that theoretical studies of protein folding as well as structural verification procedures could benefit from both crystal structures and the solution-like structures.

As mentioned earlier, the residue volumes are related to backbone conformation. In other words, different φ, ψ regions foster different packing environments. To further investigate this relationship, we split up Figure 1b into the 4 classes of secondary structure in Figure 2c to 2f. The pattern of residue volumes over φ, ψ space is consistent across the secondary structure classifications. Therefore, we can discuss the plots in terms of the dependence of residue volumes on backbone conformation. Overall, the plots confirm that proteins pack more loosely within right-handed helical region than in the sheet region. The helical region shows a saddle-like pattern, where residues pack more loosely toward the saddle's edges, in the H, C, and T classes of secondary structures. For an α-helix, this less dense packing corresponds to the conformational requirement for the side-chain to point radially away from the cylinder formed by the backbone. In contrast, the sheet region in the upper-left corner defined by $-180 < \varphi < -125°$ and $125° < \psi < 180°$ exhibits tighter packing. Structurally, this region corresponds to an alignment of the CO and HN dipoles between two strands (Ho et al. 2003), indicating that main-chain hydrogen bonds promote/permit tighter packing. For example, β-hairpin formation of main-chain hydrogen bonds favors tight packing of the side-chains between the two strands. Interestingly, these plots show dependency on ψ angles. A "belt" of conformations including the left handed helical conformations ($-90° < \psi < 90°$) corresponds to looser packing, while outside that region tighter packing is observed.

3.5 Volume dependence of 20 amino acids on backbone conformations

Figure 3 shows the variation of vol% with respect to φ, ψ for the individual amino acids. We will discuss them in terms of their distribution of vol% and population. In general, the vol% patterns shows that residues occupy more volume when its backbone conformation falls into the right-handed helical region than in the sheet region. In the right-handed helical region, the saddle is generally seen, where packing is less dense towards the edges. In the sheet region, generally all amino acids pack a little more densely than average. The amino acids HIS, GLY, MET, PHE, SER, THR, TRP, TYR, CYS produce the smallest volumes or tightest packing in the φ, ψ region toward −180°, 180°. ASP shows an interesting spur of

larger than average volumes at φ of −45 to −90 and ψ of −45 to −90. Closer inspection of these conformations reveals that these larger than average ASP volumes occur in turn conformations in contact with the water solvent. For all residues, bridging areas between right-handed helical and sheet regions are packed less densely.

For sampling of Ramachandran space shown in Figure 3, the 20 residues exhibit the expected distributions, where GLY samples the most conformational space and PRO is most restricted. Surprisingly, GLY does not populated extensively in the sheet region probably due to its lack of side chain interactions. In general, the remaining 18 residues, regardless of shape and size, follow similar distribution patterns to what is seen over the entire dynameome in Figure 1b. If we assume that the more φ, ψ space a residue can populate, the easier it can replace other residues or be replaced, the clear difference among the amino acids in their populated regions may give clues as to which amino acids are least responsible for maintaining the folded state of a protein: namely, GLY, ALA, SER, THR, and ASP. In contrast, TRP, CYS and MET show quite restricted conformational possibilities (as does of course, PRO). HIS, PHE, and TYR also have relatively limited backbone conformational freedom. These results are in good agreement with BLOSUM62 matrix (Henikoff and Henikoff 1992), which represents how well amino acids are conserved during evolution as well as the likelihood that each will substitute for another, and with an in-depth statistical analysis of Ramachandran distributions of the 20 amino acids (Dahl et al. 2008).

3.6 Backbone Dependency of Side-chain Conformation

Figure 4 plots the population (Figures 4a–b) and value (Figures 4d–f) of the first side-chain torsion angle χ_1 based on the 3 rotamer classes of *gauche*$^-$ (**M**), *gauche*$^+$ (**P**) and *trans* (**T**) (Lovell et al. 2000) against backbone torsion angles. Numerous similar studies have been done using limited PDB data (Benedetti et al. 1983;Chandrasekaran and Ramachandran 1970;Dunbrack 2002;Dunbrack and Cohen 1997;Dunbrack and Karplus 1993;Lovell et al. 2000;Ponder and Richards 1987). Such libraries are usually studied by clustering observed conformations or by dividing torsion angle space into bins and determining the average conformation in each bin (Dunbrack 2002). However, our dynameome dataset exhibits a continuous sampling of the near-native conformational space that allows us to point out unique features of the native state ensemble that are less clear when data size is limited. The three top panels of Figure 4(a–c) show the population of side-chains found in each of these three χ_1 rotamers (**M**, **P**, and **T**, respectively) as a function of φ and ψ (residues PRO, ALA and GLY are excluded). At any given, φ ψ angle, the population percentages from each of the three rotamers sums to 100%. Figure 4a shows that the **M** rotamer is highly populated in the sheet region where $-135° < \varphi < -90°$ and $\psi > 135°$ and the fringe of the two helical regions. In Figure 4b, the **P** rotamer only populates limited regions due to its nudged conformation and mostly where both **T** and **M** rotamers are not favored ($-180° < \varphi < -150°$, $150° < \psi < 180°$). On the contrary, in the region where $\psi > 150°$, the **T** rotamer is scarce due to the clash between N_i and Cγ group, but becomes the preferred rotamer in the sheet region where $\varphi < -135°$ and $90° < \psi < 135°$. The **T** rotamer is preferred when ψ is around −45° regardless φ angle.

The bottom 3 panels of Figure 4(d–f) plot the value χ_1 in the 3 rotameric states as a function of φ, ψ. While consistent with previous work (Dunbrack 2002; Dunbrack and Cohen 1997; Dunbrack and Karplus 1994; Lovell et al. 2000), our dynamoemic data set provides the fine details of χ_1 rotamer dependence on backbone conformation. Figure 4d clearly shows that the **M** rotamer is dependent more on ψ than φ. Its optimal value of −60° is shown in its area of highest population. However, the most widely populated value for the **M** rotamer is lower at −70°. The range of values for the **M** rotamer reflects this with a distribution from −55° to −85°. The **P** rotamer shown in Figure 4e is dependent on both φ, ψ with larger χ_1 values towards 70° centered around φ, ψ values of −125°,145° that decrease radially outward.

Again, the **P** rotamer occupies its optimal χ_1 value of 60° in its most populated area towards φ, ψ values of −180°,180°. However, the χ_1 distribution of the **P** rotamer ranges asymmetrically from 40° to 70°. The **T** rotamer (Figure 4f) displays the strongest dependency on than the other rotamers. The **T** rotamer also exhibits dependence on φ in the left-handed helical region. Optimal χ_1 values for the **T** rotamer occur in bands where ψ is between −90° to −45° and between 90° to 135°. The χ_1 distribution for the **T** rotamer is also skewed and ranges from 175° to 205°.

3.7 Volume dependency on χ_1 rotamer conformation

For completeness, we plotted the percent volume against χ_1 rotamer values in Figure 5. While there is no strong correlation between volumes and χ_1, these distributions result from the flexibility of different χ_1 rotamers. Consistent with Figure 4, all the **M**, **P**, and **T** rotamers exhibit peak χ_1 values of −70°, 65° and 180°, respectively at a vol% of 100. Essentially, most residue volumes are at their mean. The plot does not include data from the amino acid PRO, since the residue also restricts the χ_1 value in the P rotamer around 0°. As χ_1 moves away from its mean in the 3 rotamers, residue volumes still peak around their mean volume, but with a drop off in population. Also, we see that transition between the **M** and **T** rotamer distributions, but not with the **P** rotamer. The **P** rotamer is limited on both sides by the N-Cα and the Cα-C bonds that create an energy barrier for Cα-Cβ bond rotation. No such barrier is posed by hydrogen attached to the Cα as a barrier to the rotation between the **M** and **T** rotamers. The primary difference between the three χ_1 rotamers is their range of volumes. With vol% extending up to 132%, the **T** rotamer samples less dense environments than both the **M** and **P** rotamers, because the **T** rotamer is the least sterically hindered conformation. The **P** rotamer samples the least amount of residue volumes, which is consistent with the fact that **P** is only favored in limited backbone conformations (Figure 4b). For the **T** rotamer, especially in right-handed helical and sheet regions where φ values are negative, the backbone bends away from the side-chain and allows residues in the **T** rotamer about 10% more volume. Altogether, Figure 5 indicates that rotamer conformations are not restrained to specific packing environments. Therefore, side-chain packing and χ_1 rotamers are independent of each other and excluded volume is not sufficient enough to define explicit rotamer conformations (Kussell et al. 2001).

3.8 Rationalization of the interdependence between χ_1 and φ, ψ

The relationship between the χ_1 angle and the backbone conformation can be summed up in detail using steric interactions as diagrammed in Figure 6. Similar explanations have been made using butane and syn-pentane interaction (Dunbrack and Cohen 1997;Dunbrack and Karplus 1993) as well as similar Newman projections (Dunbrack 2002;Dunbrack and Karplus 1993;Dunbrack and Karplus 1994). Because our dynameome provides high-resolution sampling of torsion angles, the interdependence of the backbone and χ_1 angle can be more clearly visualized, such as the steric repulsions between a residue's Cγ atom with its main-chain N-Cα or Cα-C bonds. To simplify the discussion, the subscripts for atoms on the reference residue i are omitted, but used to refer to atoms preceding or adjacent to the reference residue.

In Figure 6, the χ_1 angle of three rotamers is indicated by the Cγ position where the "dial indicator" on the χ_1 dial is the Cβ-Cγ bond, and the positions of "**T**" (crossing the magenta arc), "**M**" (yellow) and "**P**" (cyan) conformation are indicated with dashed-outline Newman-projection-style bonds. The allowable φ, ψ torsion angles are colored in grey in the dial indicator. The position of the χ_1 rotamer relative to the φ or ψ angles helps to explain its dependence. The **T** rotamer is influenced by the next residue i+1, the **M** rotamer is influenced by the previous residue i−1, and the **P** rotamer is influenced by both residues i−1 and i+1. However, all 3 rotamers show more dependency on ψ than φ. The ψ torsion angle

involves the atoms N, Cα, C and N_{i+1} atoms and can occupy almost any angle. This flexibility can bring two heavy atoms N_{i+1} or O to pack against the Cγ atoms depending upon ψ angle. For the T rotamers, this pushes the Cγ atom towards the Hα atom, so that χ_1 angle in T rotamers becomes smaller than 180°. For the **P** rotamer, these interactions move the Cγ towards the N atom and lower values of χ_1. Since the φ torsion angle involves the atoms C_{i-1}, N, Cα and C, the atoms O_{i-1} and Cγ can form syn-pentane interactions (Dunbrack and Karplus 1994). When φ angle is between −180 and −150, the **M** rotamer is affected by the clash of O_{i-1} with Cγ. Specifically the clash between O_{i-1} and H on Cγ causes a very low population and larger negative χ_1 angle value (See Figure 4). Because the φ angle dependency of χ_1 is also noticeable in the left-handed helical region where two oxygen atoms facing each other create a sterically crowd environment, the **P** rotamer has very low population in this region as an extreme case.

Besides the above general effects, each rotamer has certain unique properties to their χ_1 dependence on backbone conformation. With the Cγ facing toward the N atom, the **M** rotamer is expected to depend on φ and is influenced by the O_{i-1} clashes with the Cγ and substituents on it. Even the Cγ atom facing away from the C atom shows some dependence on ψ. While this was suggested as the influence of other rotamers (Dunbrack and Cohen 1997), we find the **M** rotamer's dependence on ψ angle is due to the packing of the Cβ with the O atom and the clashes of H_N with hydrogens on Cβ as ψ changes. For the **P** rotamer, due to the clashes of Cγ atoms "pinched" between N-Cα and C-O bonds, it rarely populates in the left-handed helical region. The **P** rotamer shows clear dependency on both φ and ψ. The φ dependence is due to the packing of the Cγ group with the H_N and O_{i-1} atoms. The ψ dependency is due to the O and N_{i+1} atoms. Also worth noticing, the lack of the **P** rotamer in the right handed helical region can be attributed to the requirement of forming hydrogen bonds in helices, which keeps the Cγ from taking this conformation. As expected, the **T** rotamer is highly ψ dependent in the right-handed helical and sheet regions. The **T** rotamer stays in its optimal conformation around ψ = 120° and ψ = −60° where atoms O and N_{i+1} both have the least interaction with a residue's Cγ atom. As the Cα–C bond rotates, either atom N_{i+1} or atom O approaches atom Cγ, and the Cγ is pushed towards the Hα, increasing the χ_1 value.

4 Conclusion

In this study, we took advantage of MD simulations to generate about a million structures that sample the ensemble of near-native conformations. In contrast to the sparse data provided by the PDB, we were able to sample from a continuous conformational space and to better characterize the dependency of side-chain packing and conformation upon backbone conformation. We were able to determine the contribution of the local environment (backbone conformation) and non-local environment (solvent exposure) on residue volume with implications describing side-chain packing. A comparison between buried and exposed residues shows that buried residues (protein core) prefer tight packing and are found only in a rather well defined conformational space (Figure 2a). We also found that packing is different for different secondary structures, where strands promote tighter packing while helices promote looser packing (Table 3 & Figure 1 and 2). In addition, the packing has a strong dependency on backbone conformation regardless of different side-chain conformations (Figures 5). Because the dynameome dataset allows more fine-grained analysis, we were also able to define more precisely the relationship between the first side-chain rotamter χ_1 and the backbone. First, all rotamers show dependence on the ψ torsion angle due to clashes of the O and N_{i+1} atoms with the Cγ, while the influence of the φ torsion angle is less so due to weaker interactions of the N_i and O_{i-1} with the Cγ. Second, the variance of all 3 χ_1 rotamers from their canonical conformation are skewed to one side due to syn-pentane interactions. Third, "non-local" interactions, such as hydrogen bonds

from i-4 residues in α helices play an important role in side-chain conformation. These results help to define the exact role that backbone conformation plays on the determination of protein folding. Although we have couched our discussion in terms of the dependence of side-chain characteristics on main-chain conformation, in fact it is a two-way street. While the backbone conformation sets the placement for the side chain, the packing of side chains determines the position of the backbone atoms. While we find that there are other factors involved, our study using the dynameome to model the native state ensemble clearly characterizes the interplay of this relationship.

Acknowledgments

This work was supported by NIH/NIGMS grant R0GM81631. We would also like to thank Michael Levitt for helpful discussions.

References

Beck DA, Alonso DO, Inoyama D, Daggett V. The intrinsic conformational propensities of the 20 naturally occurring amino acids and reflection of these propensities in proteins. Proc Natl Acad Sci U S A. 2008a; 105:12259–64. [PubMed: 18713857]

Beck DA, Jonsson AL, Schaeffer RD, Scott KA, Day R, Toofanny RD, Alonso DO, Daggett V. Dynameomics: mass annotation of protein dynamics and unfolding in water by high-throughput atomistic molecular dynamics simulations. Protein Eng Des Sel. 2008b; 21:353–68. [PubMed: 18411224]

Becker, RA.; Chambers, JM.; Wilks, AR. The New S Language. Chapman & Hall; New York: 1988.

Benedetti E, Morelli G, Nemethy G, Scheraga HA. Statistical and energetic analysis of side-chain conformations in oligopeptides. Int J Pept Protein Res. 1983; 22:1–15. [PubMed: 6885244]

Benson NC, Daggett V. Dynameomics: large-scale assessment of native protein flexibility. Protein Sci. 2008; 17:2038–50. [PubMed: 18796694]

Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. Acta Crystallogr D Biol Crystallogr. 2002; 58:899–907. [PubMed: 12037327]

Chakrabarti P, Pal D. Main-chain conformational features at different conformations of the side-chains in proteins. Protein Eng. 1998; 11:631–47. [PubMed: 9749916]

Chakrabarti P, Pal D. The interrelationships of side-chain and main-chain conformations in proteins. Prog Biophys Mol Biol. 2001; 76:1–102. [PubMed: 11389934]

Chandrasekaran R, Ramachandran GN. Studies on the conformation of amino acids. XI. Analysis of the observed side group conformation in proteins. Int J Protein Res. 1970; 2:223–33. [PubMed: 5538390]

Chung SY, Subbiah S. The use of side-chain packing methods in modeling bacteriophage repressor and cro proteins. Protein Sci. 1995; 4:2300–9. [PubMed: 8563626]

Chung SY, Subbiah S. How similar must a template protein be for homology modeling by side-chain packing methods? Pac Symp Biocomput. 1996:126–41. [PubMed: 9390228]

Dahl DB, Bohannan Z, Mo Q, Vannucci M, Tsai J. Assessing side-chain perturbations of the protein backbone: a knowledge-based classification of residue Ramachandran space. J Mol Biol. 2008; 378:749–58. [PubMed: 18377931]

Day R, Daggett V. All-atom simulations of protein folding and unfolding. Adv Protein Chem. 2003; 66:373–403. [PubMed: 14631823]

DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA, USA: 2002.

Dunbrack RL Jr. Rotamer libraries in the 21st century. Curr Opin Struct Biol. 2002; 12:431–40. [PubMed: 12163064]

Dunbrack RL Jr, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci. 1997; 6:1661–81. [PubMed: 9260279]

Dunbrack RL Jr, Karplus M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. J Mol Biol. 1993; 230:543–74. [PubMed: 8464064]

Dunbrack RL Jr, Karplus M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. Nat Struct Biol. 1994; 1:334–40. [PubMed: 7664040]

Feig M. Is Alanine Dipeptide a Good Model for Representing the Torsional Preferences of Protein Backbones? Journal of Chemical Theory and Computation. 2008; 4:1555–1564.

Gerstein M, Tsai J, Levitt M. The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra. J Mol Biol. 1995; 249:955–66. [PubMed: 7540695]

Griffiths-Jones SR, Sharman GJ, Maynard AJ, Searle MS. Modulation of intrinsic phi, psi propensities of amino acids by neighbouring residues in the coil regions of protein structures: NMR analysis and dissection of a beta-hairpin peptide. J Mol Biol. 1998; 284:1597–609. [PubMed: 9878373]

Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A. 1992; 89:10915–9. [PubMed: 1438297]

Ho BK, Thomas A, Brasseur R. Revisiting the Ramachandran plot: hard-sphere repulsion, electrostatics, and H-bonding in the alpha-helix. Protein Sci. 2003; 12:2508–22. [PubMed: 14573863]

Hutchinson EG, Thornton JM. PROMOTIF--a program to identify and analyze structural motifs in proteins. Protein Sci. 1996; 5:212–20. [PubMed: 8745398]

Jiang L, Kuhlman B, Kortemme T, Baker D. A “solvated rotamer” approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. Proteins. 2005; 58:893–904. [PubMed: 15651050]

Jonsson AL, Scott KA, Daggett V. Dynameomics: a consensus view of the protein unfolding/folding transition state ensemble across a diverse set of protein folds. Biophys J. 2009; 97:2958–66. [PubMed: 19948125]

Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983; 22:2577–637. [PubMed: 6667333]

Kawano Y, Kumagai T, Muta K, Matoba Y, Davies J, Sugiyama M. The 1.5 A crystal structure of a bleomycin resistance determinant from bleomycin-producing Streptomyces verticillus. J Mol Biol. 2000; 295:915–25. [PubMed: 10656800]

Kehl C, Simms AM, Toofanny RD, Daggett V. Dynameomics: a multi-dimensional analysis-optimized database for dynamic protein data. Protein Eng Des Sel. 2008; 21:379–86. [PubMed: 18411222]

Krivov GG, Shapovalov MV, Dunbrack RL Jr. Improved prediction of protein side-chain conformations with SCWRL4. Proteins. 2009; 77:778–95. [PubMed: 19603484]

Kuntz ID. Tertiary structure in carboxypeptidase. J Am Chem Soc. 1972; 94:8568–72. [PubMed: 4638988]

Kussell E, Shimada J, Shakhnovich EI. Excluded volume in protein side-chain packing. J Mol Biol. 2001; 311:183–93. [PubMed: 11469867]

Levitt M, Gerstein M, Huang E, Subbiah S, Tsai J. Protein folding: the endgame. Annu Rev Biochem. 1997a; 66:549–79. [PubMed: 9242917]

Levitt M, Hirshberg M, Sharon R, Daggett V. Potential-Energy Function and Parameters for Simulations of the Molecular-Dynamics of Proteins and Nucleic-Acids in Solution. Computer Physics Communications. 1995; 91:215–231.

Levitt M, Hirshberg M, Sharon R, Laidig KE, Daggett V. Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. Journal of Physical Chemistry B. 1997b; 101:5051–5061.

Lovell SC, Word JM, Richardson JS, Richardson DC. The penultimate rotamer library. Proteins. 2000; 40:389–408. [PubMed: 10861930]

MacArthur MW, Thornton JM. Protein side-chain conformation: a systematic variation of chi 1 mean values with resolution - a consequence of multiple rotameric states? Acta Crystallogr D Biol Crystallogr. 1999; 55:994–1004. [PubMed: 10216296]

Mandel N, Mandel G, Trus BL, Rosenberg J, Carlson G, Dickerson RE. Tuna cytochrome c at 2.0 A resolution. III. Coordinate optimization and comparison of structures. J Biol Chem. 1977; 252:4619–36. [PubMed: 194885]

Minor DL Jr, Kim PS. Context is a major determinant of beta-sheet propensity. Nature. 1994; 371:264–7. [PubMed: 8078589]

Moult J, Fidelis K, Kryshtafovych A, Rost B, Hubbard T, Tramontano A. Critical assessment of methods of protein structure prediction-Round VII. Proteins. 2007; 69(Suppl 8):3–9. [PubMed: 17918729]

Moult J, Fidelis K, Kryshtafovych A, Rost B, Tramontano A. Critical assessment of methods of protein structure prediction - Round VIII. Proteins. 2009; 77(Suppl 9):1–4. [PubMed: 19774620]

Moult J, Fidelis K, Rost B, Hubbard T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)--round 6. Proteins. 2005; 61(Suppl 7):3–7. [PubMed: 16187341]

Moult J, Fidelis K, Zemla A, Hubbard T. Critical assessment of methods of protein structure prediction (CASP): round IV. Proteins Suppl. 2001; 5:2–7.

Moult J, Fidelis K, Zemla A, Hubbard T. Critical assessment of methods of protein structure prediction (CASP)-round V. Proteins. 2003; 53(Suppl 6):334–9. [PubMed: 14579322]

Moult J, Hubbard T, Bryant SH, Fidelis K, Pedersen JT. Critical assessment of methods of protein structure prediction (CASP): round II. Proteins Suppl. 1997; 1:2–6.

Moult J, Hubbard T, Fidelis K, Pedersen JT. Critical assessment of methods of protein structure prediction (CASP): round III. Proteins Suppl. 1999; 3:2–6.

Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol. 1995; 247:536–40. [PubMed: 7723011]

O'Farrell PA, Walsh MA, McCarthy AA, Higgins TM, Voordouw G, Mayhew SG. Modulation of the redox potentials of FMN in Desulfovibrio vulgaris flavodoxin: thermodynamic properties and crystal structures of glycine-61 mutants. Biochemistry. 1998; 37:8405–16. [PubMed: 9622492]

Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol. 1987; 193:775–91. [PubMed: 2441069]

Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol. 1963; 7:95–9. [PubMed: 13990617]

Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. Adv Protein Chem. 1968; 23:283–438. [PubMed: 4882249]

Rueda M, Chacon P, Orozco M. Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. Structure. 2007a; 15:565–75. [PubMed: 17502102]

Rueda M, Ferrer-Costa C, Meyer T, Perez A, Camps J, Hospital A, Gelpi JL, Orozco M. A consensus view of protein dynamics. Proc Natl Acad Sci U S A. 2007b; 104:796–801. [PubMed: 17215349]

Samudrala R, Moult J. Determinants of side chain conformational preferences in protein structures. Protein Eng. 1998; 11:991–7. [PubMed: 9876919]

Schell D, Tsai J, Scholtz JM, Pace CN. Hydrogen bonding increases packing density in the protein interior. Proteins. 2006; 63:278–82. [PubMed: 16353166]

Shapovalov MV, Dunbrack RL Jr. Statistical and conformational analysis of the electron density of protein side chains. Proteins. 2007; 66:279–303. [PubMed: 17080462]

Simms AM, Toofanny RD, Kehl C, Benson NC, Daggett V. Dynameomics: design of a computational lab workflow and scientific data repository for protein simulations. Protein Eng Des Sel. 2008; 21:369–77. [PubMed: 18411223]

Skolnick J. In quest of an empirical potential for protein structure prediction. Curr Opin Struct Biol. 2006; 16:166–71. [PubMed: 16524716]

Smith GD, Pangborn WA, Blessing RH. Phase changes in T(3)R(3)(f) human insulin: temperature or pressure induced? Acta Crystallogr D Biol Crystallogr. 2001; 57:1091–100. [PubMed: 11468392]

Tsai J, Gerstein M. Calculations of protein volumes: sensitivity analysis and parameter database. Bioinformatics. 2002; 18:985–95. [PubMed: 12117797]

Tsai J, Taylor R, Chothia C, Gerstein M. The packing density in proteins: standard radii and volumes. J Mol Biol. 1999; 290:253–66. [PubMed: 10388571]

Voronoi G. New parametric applications concerning the theory of quadratic forms - Second announcement. Journal Fur Die Reine Und Angewandte Mathematik. 1908; 134:198–287.

Wang G, Dunbrack RL Jr. PISCES: a protein sequence culling server. Bioinformatics. 2003; 19:1589–91. [PubMed: 12912846]

West NJ, Smith LJ. Side-chains in native and random coil protein conformations. Analysis of NMR coupling constants and chi1 torsion angle preferences. J Mol Biol. 1998; 280:867–77. [PubMed: 9671556]

Zhao S, Goodsell DS, Olson AJ. Analysis of a data set of paired uncomplexed protein structures: new metrics for side-chain flexibility and model evaluation. Proteins. 2001; 43:271–9. [PubMed: 11288177]

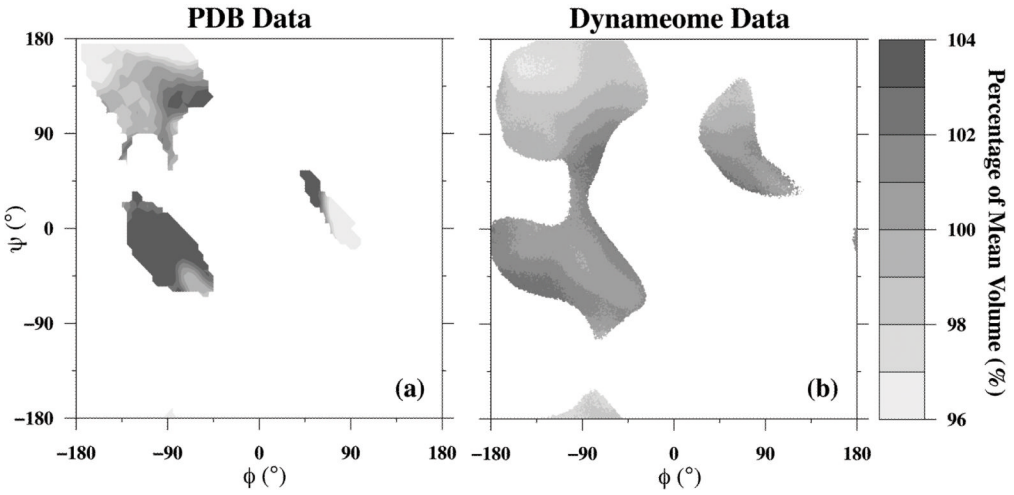

Figure 1.

Contour plots of percentage of average volumes in backbone torsion angle φ, ψ spaces using the data from **(a)** 689 PDB structures with a 5° resolution for φ, ψ value, **(b)** dynameome dataset with 1° resolution. A grey scale is used, where the darker color indicates smaller volumes and the lighter color indicates larger volumes.

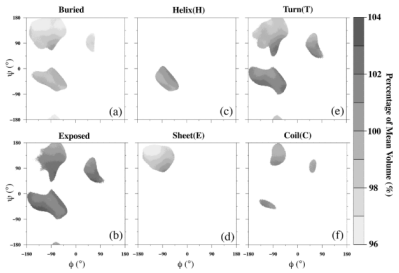

Figure 2.

The percentage of the mean volume in different packing environment is shown as contour plots in backbone torsion angle spaces (φ, ψ) from the dynameome. **(a)** Buried residues, **(b)** Exposed residues, **(c)** Residues classified as in α-helix (H) conformation, **(d)** Residues classified as in β-sheet (E) conformation, **(e)** Residues classified as in Turn (T) conformation, **(f)** Residues classified as in Coil (C) conformation. A grey scale is used, where lighter color indicates smaller volumes and darker color indicates larger volumes.

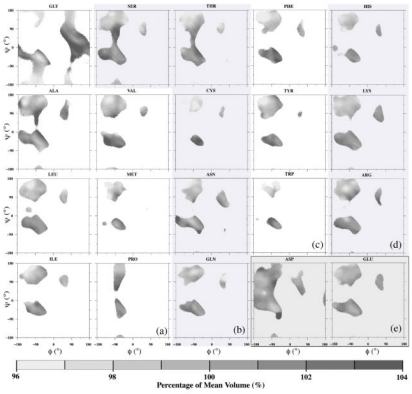

Figure 3.

The contour plots of vol% is shown for the 20 amino acid side-chains on backbone conformations. A grey background box was used to distinguish different physical properties of amino acid's R group. From left to right, top to bottom, area **(a)** consists amino acids with nonpolar, aliphatic R groups, area **(b)** consists amino acids with polar but uncharged R groups, area **(c)** includes amino acids with aromatic R groups, and area **(d)** and **(e)** are amino acid with positively and negatively charged R groups respectively.

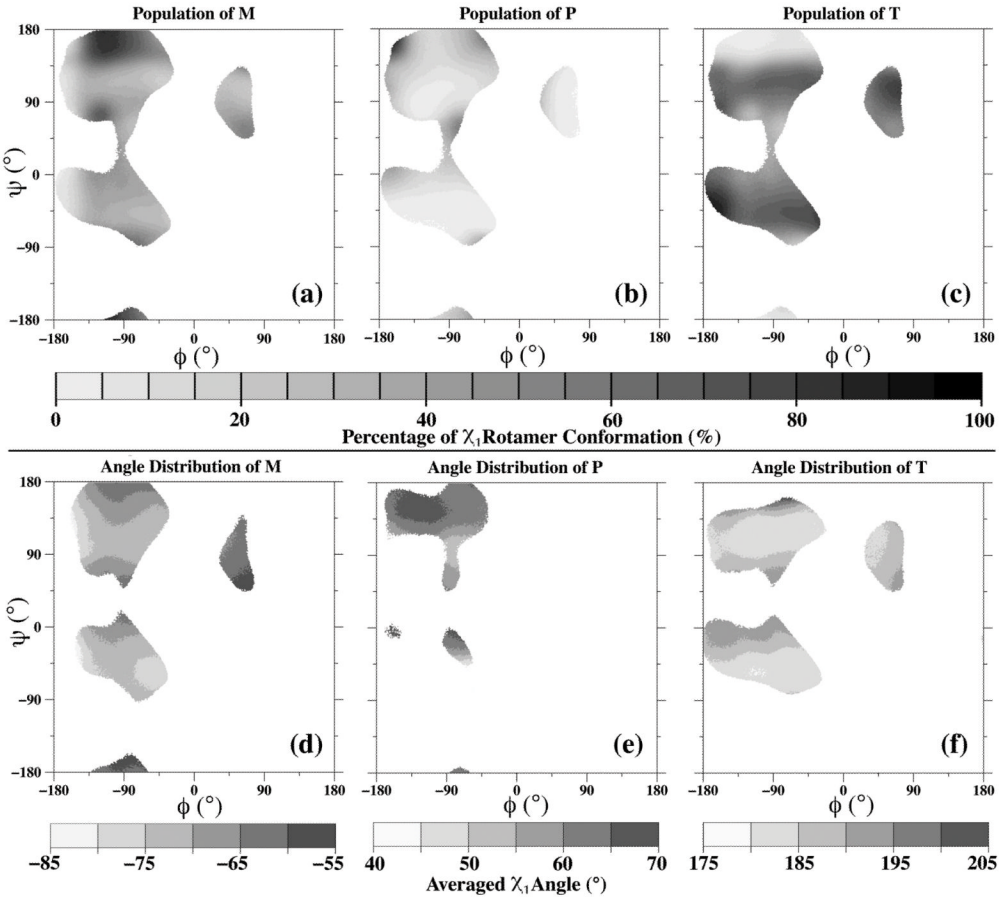

Figure 4.

Population and angle distribution of 3 χ_1 rotamers versus backbone conformation. χ_1 rotamer populations in percentages are plotted against backbone, φ ψ torsion angles for the **(a) M**, **(b) P**, and **(c) T** rotamers. At any given, φ ψ angle, the percentage of the population in each of the 3 rotamers sums to 100%. A grey scale is used from black (higher occupancy) to white (lower occupancy). The χ_1 rotamer angles are plotted against backbone, φ ψ torsion angles for the **(a) M**, **(b) P**, and **(c) T** rotamers.

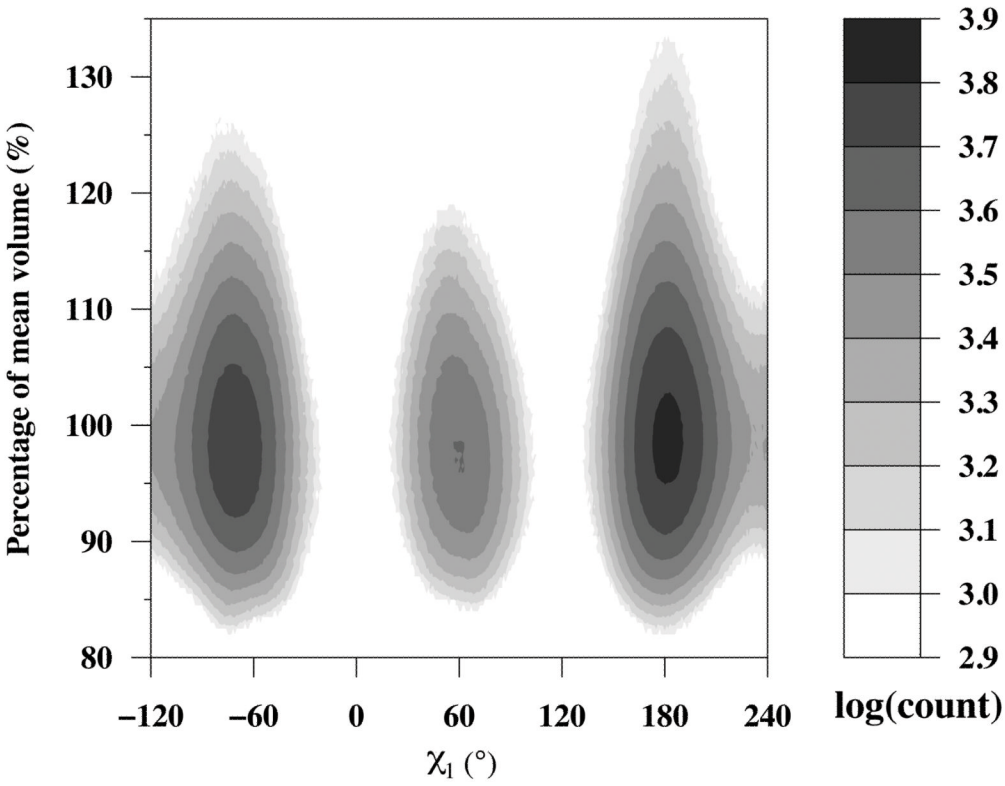

Figure 5.
Distribution of vol% plotted against their χ_1 value. This was done for all residues except GLY, ALA, and PRO. The counts are shown on a log scale, where darker indicates more observations.

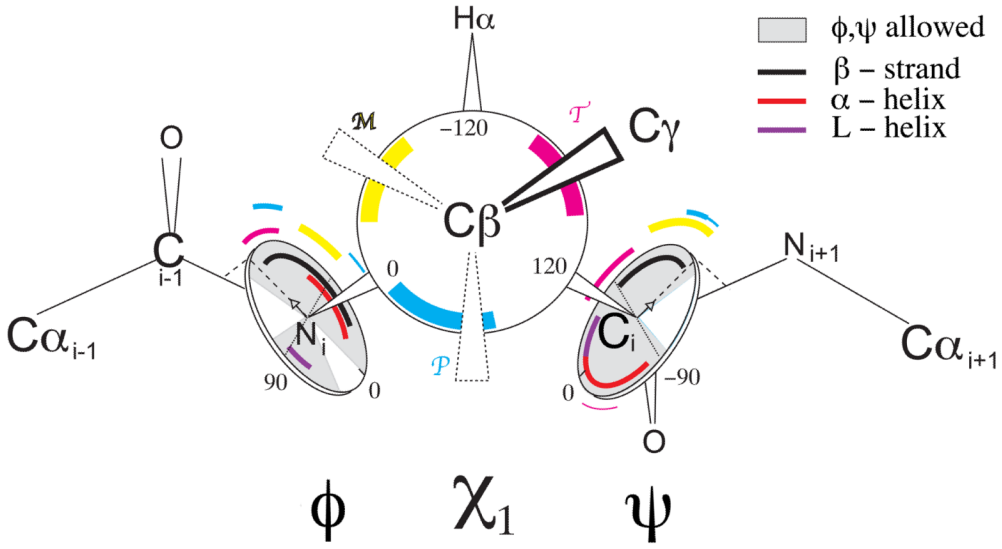

Figure 6.

Schematic representation of allowed χ_1 torsion angle depending on φ, ψ backbone torsion angles. A portion of a peptide chain (from $C\alpha_{i-1}$ to $C\alpha_{i+1}$) is shown in extended conformation. Three dials represent, φ, ψ and χ_1 torsion angles. On the φ, ψ dials, the allowed regions are indicated by the light gray shading, and forbidden regions are white. The region favored by β-stand is marked by a heavy black arc, by α-helix is marked by a heavy red arc, and by the left-handed helix is marked by a short purple arc on the dials. The preferred regions for side-chain rotamers found in this work are indicated on χ_1 dial perpendicular to Cβ–Cγ bond, on which three rotamers **T**, **M**, and **P** are in magenta, yellow, and cyan arc, respectively. The preferred rotamers on the backbone conformations are indicated just outside of φ, ψ dials in the same colors.

Table 1

SCOP Classification of Simulated folds in the dynameome dataset

SCOP Class	Number of folds	Smallest[a]	Largest[a]	Member[b]
alpha	25	31	131	1aie, 2erl, 2cpgA, 1utg, 1i27A, 1dp7P, 1g8qA, 1cy5A, 1fk5A, 1lriA, 1psrA, 3caoA, 1jr8A, 256bA, 1bkrA, 1kr7A, 1i8oA, 1dlwA, 1elwA, 2a0b, 2mhr, 1fazA, 1ijyA, 1e85A, 1c52
beta	15	64	135	1c8cA, 1c9oA, 1c4qA, 1c5eA, 3vub, 3chbD, 1qauA, 2mcm, 1f86A, 1flmA, 1whi, 2cuaA, 1bfg, 1rie, 1c11A
alpha/beta	6	87	147	1aba, 1thx, 1jf8A, 1ccwA, 1i5gA, 1akr
alpha+beta	21	61	138	2igd, 1cseI, 1b3aA, 1cc8A, 1vcc, 1euvB, 1fm0D, 1iqzA, 1opd, 1cyo, 1rgeA, 1tldA, 4ubpA, 1lkkA, 1e w4A, 1bkf, 1kafA, 1kpf, 1qtoA, 1c7kA, 1gmuA
other[c]	10	21	83	1g7aA, 1isuA, 1f94A, 1i71A, 1jekB, 1jekA, 1et1A, 1ppt, 1wfbA, 1g6uA

[a] Number of residues in each class

[b] Members are arranged in order of increasing the number of residues. The 5th character in the PDB id denotes the chain that was used if the fold was part of a multimeric interaction in the PDB file.

[c] Includes SCOP classification other than all alpha, all beta, alpha/beta, alpha+beta, which contains classification of small proteins: 1g7aA, 1isuA, 1f94A, and 1i71A; coiled coil proteins: 1jekB and 1jekA; peptides: 1et1A, 1ppt and 1wfbA; designed proteins:1g6uA

Table 2

Summary of parameters for the MD simulations for each protein in the dynameomic dataset: PDB ID, number of residues, number of water molecules, size of water box and average RMSD for all MD generated structures.

PDB ID	Num of Residue	Num of Water	Box Size (Å^3)* 1000	RMSD (Å)		PDB ID	Num of Residue	Num of Water	Box Size (Å^3)* 1000	RMSD (A)	
				Mean	STD					Mean	STD
1aba	87	3437	115	1.6	0.2	1i8oA	113	3554	121	1.9	0.1
1aie	31	2256	72	1.7	0.2	1ijyA	122	3990	137	1.9	0.1
1akr	147	4017	139	1.9	0.1	1iqzA	81	2740	93	1.8	0.2
1b3aA	67	3286	108	1.9	0.1	1isuA	62	2314	78	1.9	0.1
1bfg	126	3848	133	1.7	0.2	1jekA	40	2420	88	1.7	0.2
1bkf	107	3744	127	1.8	0.2	1jekB	34	2188	71	1.4	0.4
1bkrA	108	3561	122	1.8	0.2	1jf8A	130	3984	137	1.9	0.1
1cllA	135	3919	136	2.0	0.0	1jr8A	105	3663	125	1.8	0.1
1c4qA	69	2284	78	1.9	0.1	1kafA	108	3656	125	1.8	0.2
1c52	131	3883	134	1.9	0.1	1kpf	111	4053	139	1.9	0.1
1c5eA	95	3323	112	1.7	0.2	1kr7A	110	3350	115	1.7	0.2
1c7kA	131	3857	133	2.0	0.0	1lkkA	105	3864	130	1.9	0.1
1c8cA	64	2823	93	1.9	0.1	1lriA	98	3190	108	1.8	0.1
1c9oA	66	2717	90	1.8	0.1	1opd	85	2829	96	1.8	0.1
1cc8A	72	2374	81	1.9	0.1	1ppt	36	2035	66	1.7	0.2
1ccwA	137	4048	139	1.9	0.1	1psrA	100	4173	142	1.9	0.1
1csel	63	2389	81	1.3	0.3	1qauA	112	4352	145	1.8	0.1
1cy5A	92	2937	101	1.7	0.2	1qtoA	122	4590	167	1.9	0.1
1cyo	88	3576	121	1.8	0.2	1rgeA	96	3226	110	1.9	0.1
1dlwA	116	3489	119	1.8	0.1	1rie	127	3997	137	1.8	0.1
1dp7P	76	3772	126	1.4	0.2	1tldA	100	3636	124	1.9	0.1
1e85A	124	3791	130	1.8	0.2	1thx	108	3559	121	1.9	0.1
1elwA	117	3493	127	1.6	0.2	1utg	70	3536	116	1.6	0.3
1etlA	34	1901	62	1.6	0.3	1vcc	77	3150	105	1.8	0.2
1euvB	79	3347	111	1.9	0.1	1wfbA	37	2126	68	1.3	0.4
1ew4A	106	3507	120	1.9	0.1	1whi	122	4127	140	1.7	0.2
1f86A	115	3562	122	1.8	0.1	256bA	106	3560	121	1.7	0.2

PDB ID	Num of Residue	Num of Water	Box Size (Å^3)* 1000	RMSD (Å)		PDB ID	Num of Residue	Num of Water	Box Size (Å^3)* 1000	RMSD (A)	
				Mean	STD					Mean	STD
1f94A	63	2533	85	1.9	0.1	2a0b	118	3859	132	1.6	0.2
1fazA	122	4008	137	1.8	0.2	2cpgA	43	2490	81	1.9	0.1
1fk5A	93	2823	96	1.9	0.1	2cuaA	122	3855	132	1.9	0.1
1flmA	122	4116	139	1.8	0.1	2erl	40	1900	62	1.9	0.1
1fm0D	81	2706	92	1.9	0.1	2igd	61	2598	86	1.9	0.1
1g6uA	47	2534	89	1.5	0.3	2mcm	112	3234	110	2.0	0.0
1g7aA	21	1211	39	1.9	0.1	2mhr	118	3943	135	1.8	0.1
1g8qA	90	3879	128	1.8	0.2	3caoA	102	4297	143	1.9	0.1
1gmuA	138	4295	147	1.8	0.2	3chbD	103	3851	130	1.8	0.1
1i27A	73	2956	99	1.8	0.2	3vub	101	3908	131	1.7	0.2
1i5gA	144	4174	145	1.8	0.1	4ubpA	100	3671	131	1.8	0.1
1i71A	83	3497	117	1.8	0.2	**AVG**[a]	**93**	**3342**	**114**	**1.8**	**0.1**

[a] average over all MD simulations used in the analysis.

Table 3

Average volumes of the 20 amino acids from entire dynameome dataset (**Ave**) with standard deviation (**Std**), crystal structures (**ProtOr**), exposed resides (**Exposed**), buried residues (**Buried**), and secondary structures, sheets (**E**), helices (**H**), coils (**C**), and turns (**T**) in $Å^3$

	Ave	Std	ProtOr[a]	Exposed	Buried	Secondary Structure			
						E	H	C	T
GLY	65	5	65	67	63	63	65	65	66
ALA	91	8	90	93	90	91	91	92	91
VAL	142	10	139	147	141	140	144	143	143
LEU	171	13	164	176	170	168	172	173	172
ILE	169	12	164	173	168	167	171	169	169
PHE	202	15	192	207	201	196	207	202	201
TYR	203	13	197	211	201	199	205	206	204
TRP	240	15	228	246	238	232	243	239	242
MET	181	14	167	185	179	180	182	182	180
PRO	128	9	123	131	125	126	129	128	129
SER	94	6	95	96	92	94	94	95	95
THR	122	9	126	124	119	121	123	122	122
CYS	120	14	103	125	119	114	123	119	118
ASN	127	8	125	130	124	127	127	128	128
GLN	156	9	149	158	152	154	156	156	155
HIS	163	11	160	165	161	159	164	164	163
LYS	174	10	167	176	170	172	175	175	175
ARG	200	11	194	203	195	198	201	201	201
ASP	105	5	117	106	104	107	104	106	106
GLU	133	7	142	134	130	132	132	134	133

[a] from reference (Tsai et al. 1999).