Statistical analysis of amino acid side chain flexibility for 1:n Protein-Protein docking

Dissertation zur Erlangung des Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)
der Technischen Fakultät
der Universität Bielefeld

vorgelegt von

Kerstin Koch

Bielefeld, den 18. Juni 2003

Dipl.-Biol. Kerstin Koch
AG Angewandte Informatik
Technische Fakultät
Universität Bielefeld
e-mail: kerstin@techfak.uni-bielefeld.de

Abdruck der genehmigten Dissertation
zur Erlangung des akademischen Grades Doktor-Ingenieur (Dr.-Ing.).
Von Kerstin Koch am 18. Juni 2003
der Technischen Fakultät der Universität Bielefeld vorgelegt.
Verteidigt und genehmigt am 17. Juli 2003.

Prüfungsausschuss:

Prof. Dr. Gerhard Sagerer
Prof. Dr. Hans-Peter Lenhof
Prof. Dr. Jens Stoye
Dr. Elke Möllmann

Gedruckt auf alterungsbestaendigem Papier nach DIN ISO 9706

Contents

Abstract

In the following thesis, the conformations of amino acid side chains and their flexibilities upon complex formation are investigated. For conformation prediction and side chain demangling tasks, new rotamer libraries optimised for the docking problem are compiled with the data of unbound and complexed proteins. The flexibilities of side chains and their preferred directions of movement are investigated according to the environment of the side chain like Secondary Structure, rotamericity and solvent accessible surface (SAS) area. The preferred conformations and flexibilities of different residues can be integrated in docking algorithms to allow semi-flexible docking of two unbound structures. During semi-flexible docking, side chain flexibiliby is taken into account. Steric clashes which are caused by flexible side chains have not to be penalised that much as steric clashes shown by side chains which are inflexible because flexible side chains are thought to move away if unfavourable interactions occur. In side chain demangling, clashing side chains can be moved in their preferred direction of movement or according to the probability of the rotamer combinations.

1. Introduction

In the post-genome area with many genomes known, the understanding of the genome and its interpretation in terms of structure and function of the coded proteins is the new challenge in the field of bioinformatics. Nowadays, there are thousands of enzyme structures in the PDB (20868, may 6th, 2003), but because of the genome projects, there are a lot of more sequences available (125744 annotated SwissProt entries on April 30th, 2003, 861417 TrEMBL entries on may 9th , 2003). The biological function of a protein is mediated by its structure, but it is a difficult and time consuming task to solve the structures of macromolecules. Many Nobel Prizes for chemistry are assigned in this field (see http://almaz.com/nobel/). Therefore, the interpretation of the genome in terms of structure and biological function will be an important problem in the field of bioinformatics for the next years.

To close the sequence-function gap, one task for protein-bioinformatics is the prediction of function from the protein sequence, another important challenge is prediction of the 3D structure of proteins given the amino acid sequence. In the 1960, Christian Anfinsen discovered that the sequence of a protein determines its 3D structure. For structure prediction, statistic investigations about amino acids, their occurance in Secondary Structure elements and their preferred 3D conformations are performed. The information about preferred conformations is given in rotamer libraries, which consists of discretised angle ranges and the associated probabilities. To model the side chain conformation of unknown protein structures, the conformation with the highest probability according to the rotamer library which does not penetrate atoms of other residues is chosen.

Other lines of action in protein bioinformatics are the prediction of protein complexes and target identification for drug development. Proteins are involved in all metabolic pathways and are important for pathogenicity of organisms. In many diseases, proteins and their regulation are affected, so that knowledge about proteins, their regulation and the pathways they are involved in is important to fight diseases. To discover new targets for drugs against pathogenic organisms, one ideally wants to find proteins which have important functions in these organisms, but not in humans or animals. If a protein is involved in an disease, one wants to find substances which bind to the protein and inhibit its action, ideally without having an effect on other important proteins of the metabolism, which would cause side effects. This task is addressed in the field of molecular docking, where the investigations are aimed to weather two molecules may bind and in which orientation they may do so. If the two molecules are proteins (e.g. an enzyme and its inhibitor or a protein and antibodies), their binding is called protein-protein docking. If only one is a protein and the other is a small organic molecule, the docking is called protein-ligand docking.

In 1890, the German chemist Emil Fischer proposed the lock-and-key model. The two binding substances are said to have a complementary geometrical surface and therefore fit together

like a key fits its lock. In this model no conformational changes upon complex formation are taken into account. In 1958, Daniel E. Koshland jr. postulated that the active site of many enzymes undergo conformational changes during docking [27]. According to his model, the final conformation of the two partners is reached only if a substrate or inhibitor is bound to the active site of the enzyme. This conformational change upon complex formation is called induced fit. Therefore the structure of a protein in the unbound form differs from the complexed structure of the same protein when a second molecule is bound. Flexibility occurring during the induced fit has to be integrated in docking algorithms if unbound molecules are docked. There are two kinds of flexibility which can occur during docking: the domain movement, which includes hinge or shear motions of many atoms including the backbone, and the more local side chain flexibility which is a movement around rotatable bonds within the side chain of a protein [15].

In this thesis, side chain placement and flexibility upon complex formation are investigated for 1:n protein-protein docking. At this, one protein is docked against many other proteins using a database backend. The time needed per protein should be kept short. Therefore no time consuming energy calculation can be done for the scoring of hypotheses. The scoring function in our approach [34] is based on the geometrical fit of the two proteins. Steric clashes are penalised, and the hypotheses with less steric clashes will get better scores. With the flexibility parameters from the thesis, the penalisation of steric clashes can be adopted, because if flexible amino acids are involved in steric clashes, they tend to move away avoiding the clash and therefore the penalty can be smaller compared to the penalty for inflexible residues.

Another possibility for the avoidance of steric clashes is side chain demangling, where clashing side chains are moved away. Knowledge about the probabilities for different side chain rotamers are important. Clashing side chains can be put in the most favourable rotamer without steric clashes. If the probability for this rotamer is low, the penalty for the steric clash can be raised again because the side chain has to be in an unfavourable rotamer combination. The dynamic assessment of steric clashes can on the one hand integrate the desired flexibility in the scoring function, on the other hand prevent too much flexibility because the rotamer probabilities and preferred direction of movement for a side chain are taken into account. If too much flexibility was allowed, many false positive results would be obtained. The flexibility can be assigned to the proteins in the preprocessing phase of the docking, so that the time for a docking run can be kept short.

As integration platform, a MYSQL database system is used. With this relational database, the information on the proteins from the test set are stored which thus being easily accessible for the docking algorithm and for statistical investigations. The flexibility and conformation information can be stored easily and again be accessed in the preprocessing phase of the algorithm.

The thesis is arranged as follows: after an introduction to biology in chapter 2 and protein docking (chapter 3) containing the state of the art in the field of protein flexibility and rotamer libraries, the probabilities for different conformations of amino acid side chains in different environments are investigated in chapter 4. Dependent probabilities for χ angle pairs and the χ_1 conformation given the backbone are investigated. At the end of the section, the probability for the whole set of side chain angles is shown in different rotamer libraries which are compiled according to different environments. The flexibility of side chains upon complex formation are given in chapter 5. Not only the probability for a rotamer change of single angles, but also the probability for concerted rotamer changes within one side chain are investigated. In the last part of the flexibility section, the flexibility is calculated depending on the χ_1 rotamer, the whole side chain conformation and the backbone. In chapter 6, the test set and examples

of MySQL tables can be seen. The evaluation of the rotamer libraries by calculating their performance in pruning the search tree compared to full search can be seen in chapter 7. In the last section, a short summary and outlook are given.

2. Biological Background

Proteins are build from a set of 5 atoms: Carbon (C), Hydrogen (H), Oxygen (O) Nitrogen (N) and Sulfur (S). These atoms can form bonds with different distances. A C–C bond has the length $0.154nm$ (1.54 Å). The distance between a C and an O in a C=O double bond is 1.24 Å, a C–O single bond has a length of 1.34 Å.

There are 20 proteinogenic amino acids which can be found in proteins. All residues have an ${}_2HN - CH - COOH$ or peptide group in common. The different chemical properties of the amino acids (aa) are due to the variable part of the protein. Aa can be polar, charged, apolar and of different size. These properties of the aa are used for the building of 3D structures, because aa with different features are used in different places of the protein, e.g. hydrophobic residues in the core. The different aa can be seen in figures A.1, A.2, A.3 in the appendix.

The first and largest group of aa are the hydrophobic ones. The simplest of these is called Glycine (GLY, G). It has a hydrogen atom as side chain. The side chain of Alanine (ALA, A) consists of a methyl group. Amino acids with longer carbon hydrogen side chains are Valine (VAL, V), Leucine (LEU, L) and Isoleucine (ILE, I). They are hydrophobic and important for the stabilisation of the 3D structure of proteins in a water environment, because these residues tend to be in the inner (hydrophobic) part of the protein. Proline (PRO, P) is a special amino acid because the nitrogen atom of the side chain is bond to the C_α atom of the backbone which leads to a cyclic structure. Proline is often found in the bends of folded proteins and is only mildly hydrophobic. There are two hydrophobic residues with aromatic side chains: Phenylalanine (PHE, F) with a phenylringsystem in its side chain and Tryptophane (TRP, W) with an indolring as part of its side chain. PHE and TRP are very hydrophobic amino acids. Methionine (MET, M) has a sulphur atom in its side chain. Free SH groups are reactive, but the one in the side chain of MET is bond to a methyl group and therefore non-reactive.

The second group of aa have uncharged, but polar side chains with reactive groups like OH, NH_2 or SH. Cysteine (CYS,C) has a non-protected SH group at the end of the side chain. This makes the side chain highly reactive, so that disulphid bridges can be build which are important for stabilisation of protein structures. The side chain of Tyrosine (TYR, Y) consists of a ring with a hydroxyl group. Because of the reactive hydroxyl group, TYR is not that hydrophobic compared to the other ring systems. There are two other residues with hydroxylgroups in the side chains: Serine (SER, S) and Threonine (THR, T). Because of the OH group, they are more reactive than other hydrophobic residues. Asparagine (ASN, N) and Glutamine (GLN, Q) have a NH_2 group in their side chain. The NH_2 group is in contrast to the $COOH$ group polar, but not charged.

There are five residues which are charged: Lysine (LYS, K) and Arginine (ARG, R) have a side chain with a NH_2 group which is positively charged at a neutral pH value. The NH_2 in the Histidine (HIS, H) side chain can be charged or not according to its environment. This is very

important for building and loosing bonds in the active site during catalysis. The side chains of Aspartate (ASP, D) and Glutamate (GLU, E) are charged as well. The Carboxyl group is negatively charged at physiological pH values. The charged residues are very important for biochemical reactions because electrostatic interactions can be established by charges.

To form a peptide or protein, the backbone part of the different aa have to be linked together. An example for a dipeptide can be seen in figure 2.1. The Carboxyl group of one amino acid can react with the α amino group of another residue to form a peptide bond (the yellow highlighted bond in figure 2.1). Water is separated during the reaction. The direction of a polypeptide chain is from the amino terminus with the free NH_3^+ group and to the Carboxy group with the free COO^- group.

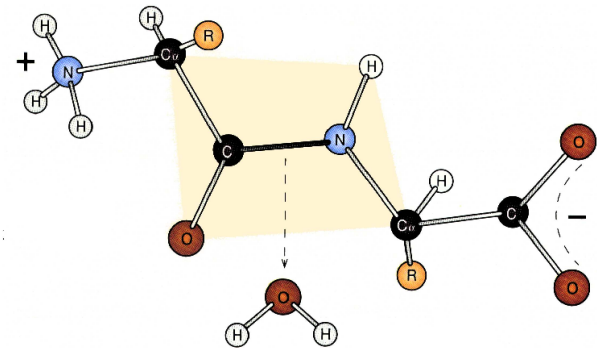


Figure 2.1.: A dipeptide, from [49]

A polypeptide chain consisting of at least 50 amino acids is called a protein. The repetitive units of the polypeptide chain are called the backbone, the variable rest of the amino acids are called side chains.

The sequence of the residues within the polypeptide chain is called primary structure. In 1953, Frederic Sanger was the first to sequence a protein and to show that each protein has a precisely determined primary structure. The relationship between primary structure and 3D structure of a protein was investigated by Christian Anfinsen on ribonuclease. He unfolded a protein by reduction with β-mercaptoethanol and urea until no catalytic activity was seen. After freeing the enzyme from β-mercaptoethanol and urea, the catalytic activity was regained. His experiments proved that the 3D structure of a protein which mediates its function is fully defined by the primary structure.

The primary structure is genetically determined and therefore links the genetic information of the cell and the function of the proteins in the metabolism. The genetic sequence is translated to proteins within two processes: transcription and translation.

During transcription, the information of the Desoxyribonucleic acid (DNA) is translated to messenger Ribonucleicacid (m-RNA). Three nucleotides of DNA or m-RNA (called a codon) code for one aa. The m-RNA nucleotides form hydrogen bonds with the complementary base pairs of the DNA strand and are connected by an enzyme called polymerase. During translation, the genetic code of the m-RNA is translated to polypeptide chains via transfer-RNA (t-RNA). This process takes place at the ribosomes, which are cell organells. The t-RNA has two important regions: the first region called the anti-codon region is complementary to the codon region of the m-RNA. In the second region, the aa for which the t-RNA molecule codes

is bond. At special places at the ribosomes, the bond residue is connected to other aa forming a peptide. After all residues coded by the m-RNA strand are bond together, the protein is released from the ribosome. A few modifications of the proteins may take place after the transcription, e.g. hydroxylation of PRO, carboxylation of GLU and phosphorylation of SER and TYR residues.

The Secondary Structure of a piece of polypeptide chain is a local spatial arrangement of the backbone atoms without regarding the side chains [20]. The ϕ and ψ angle ranges in Secondary Structures are fixed and repetitive. There are three common Secondary Structures: Helices, β-sheets and turns. The rest of the proteins which forms no Secondary Structures is classified as coiled coils (random coils). The Secondary Structures are stabilised by hydrogen bonds.

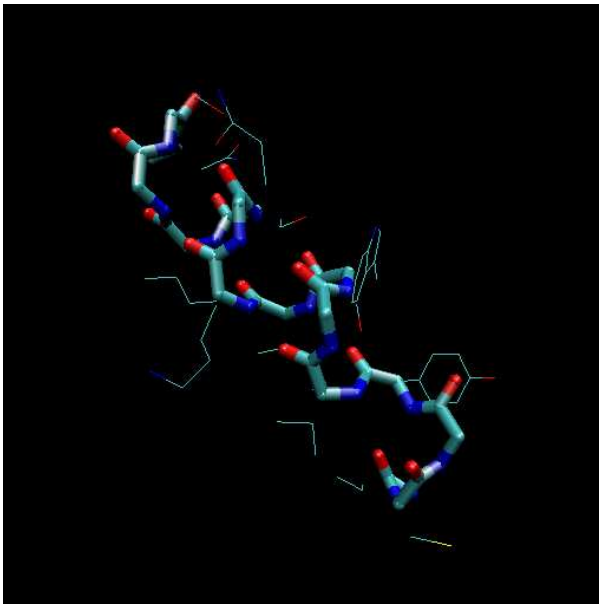


Figure 2.2.: Helix from the PDB structure *pdb2ptc*

In α-helices, the $C = O$ of the residue$_i$ can form a hydrogen bond with the NH atoms of the $i + 2th$ residue. Each residue is rotated by 100° compared to the next group. The distance between the two residues is 0.15Å. There are 3.6 amino acids in one turn of the helix. Whereas the right handed α-Helix ($3.6_{13}helix$) is the thermodynamically most stable Helix, there may be amino acid sequences which promote other helix types, e.g. the Pi and the $3_{10}helix$. In a $3_{10}helix$ which may occur at the end of $\alpha - helices$, the $residue_i$ can build hydrogen bonds with the residue 3 positions away. There are 3 residues within a turn and 10 atoms enclosed in a ring formed by each hydrogen bonds. A $Pi - helix$ is a rare helix type because the ϕ, ψ angles lie at the edges of the minimum allowed in Ramachandran maps, and some distorsion in the backbone occurs. In this helix, the water bridges are formed between residues which lie 4 positions away in the sequence [5]. An example of a α-helix can be seen in figure 2.2. The last residues of the helix shown at the bottom of the figure form a 3_{10}-helix.

In β-sheets, the water bridges are build between two strands which can be parallel or antiparallel (in the same or in the other direction). For changing the direction of the strand, hairpin loops (turns) are formed with a water bridge of residue i with the aa i+3 of the sequence.As example, two antiparallel strands of a sheet connected by a turn can be seen in figure 2.3.

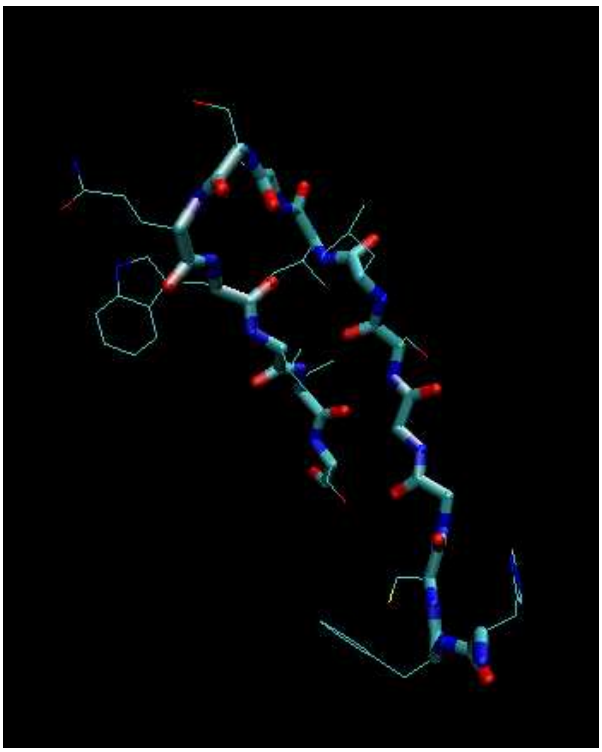


Figure 2.3.: β-sheet from the PDB structure *pdb2ptc*

The tertiary structure of a protein molecule, or of a subunit of a protein molecule, is the arrangement of all its atoms in space, without regard to its relationship with neighbouring molecules or subunits [20]. This 3D structure is formed and stabilised by electrostatic interactions, van der Waals interactions, hydrophobic interactions and hydrogen bonds. Hydrogen bonds can not only be formed between backbone atoms as mentioned before for the stabilisation of Secondary Structure elements, but can also be build from side chain atoms. The side chain atoms of TRP and ARG are hydrogen donors, the side chain of ASN, GLN, SER, THR can be donor or acceptor. For LYS, ASP, GLU the capatibility depends on the pH values. Hydrophobic interactions is another possibility of stabilisation of the 3D structure, because hydrophobic residues avoid the water at the proteinsurface and are mainly found in the core of the protein.

The quaternary structure of a protein molecule is the arrangement of its subunits in space and the ensemble of its intersubunit contacts and interactions, without regard to the internal geometry of the subunits [20]. For example hemoglobin is composed of four subunits, two α and two β chains which are bond to a hem group, so that it is placed in an environment so that oxygen can reversible be bound. The subunits of a protein are bond in a non-covalent fashion.

3. Proteindocking

Proteins are very important molecules for life. They play a crucial role in many processes like transport and cell excitation as carrier proteins or transmembrane channels, maintenance of structure and cell stability as structural proteins (like keratin), defence against microbes in the immune response, or cell growth and differentiation. Another important task where proteins are involved is the catalysis of metabolic reactions. These reactions are characteristic for living organisms. They catalyse the biochemical reactions so that they can take place under physiological conditions (temperature, pH values). Because not all reactions should take place in all cells under all circumstances, the regulation of the metabolism is an important task as well. This can be done by small organic molecules called ligands or by other proteins as inhibitors.

In many illnesses, proteins are affected. Therefore the knowledge about proteins, their interaction and regulation is important. The pharmaceutical industry wants to find and develop drugs which bind to special target proteins and regulate their activity. To prevent side effects, the target protein should not play a role in other than the affected pathway and the drug protein should not interact with other proteins from different pathways.

3.1. State of the art

One possibility to inhibit target proteins is to find another molecule which binds to this protein and prevents its activity. This problem is addressed in the field of protein docking which is introduced in the next section (3.1.1). To answer the question if and in which orientation two molecules interact with each other, flexibility of side chains has to be taken into account. An introduction to side chain flexibility is given in section 3.1.2. One possibility for integrating side chain flexibility in protein docking are side chain demangling tasks, where side chains with unfavourable conformations are placed in more likely rotamers. For side chain demangling, knowledge about probabilities for side chain conformations is needed, which is taken from rotamer libraries. The foundation for rotamer libraries is given in section 3.1.3.

3.1.1. Protein docking

Docking is the term for finding the best match between molecules. In molecular docking, the correct bound association of two molecules given their 3D coordinates is searched [16]. For protein docking, the molecules which are docked to the protein may be small organic molecules (ligands) or protein inhibitors. Two kinds of docking are distinguished: bound docking, where the two parts of a complex are taken in the bound form to reconstruct the complex and unbound

docking where the complex is reconstructed using unbound structures. Unbound structures can be crystallographic structures of unbound proteins, modeled structures or pseudo-native structures. For pseudo-native structures, the structure is taken from a complex where the partner in the complex differs from the partner of the reconstructed complex [16]. Because conformational changes take place during docking, one has to deal with flexibility during unbound docking.

In protein-protein docking, docking algorithm consists of two parts. In the first part, hypothesis how the two molecules may react are generated, in the second part, a scoring and ranking of the hypothesis has to take place. The first part is based on surface geometry. As representation of the geometrical features of the surface, the Conolly surface is taken [9]. This surface representation consists of the part of the van der Waals radii of the surface residues which get in contact with a probe sphere. Because of this, crevices are smoothed away. Geometrical features like concave regions, convex regions and saddle points are added to the surface . Based on this features, some critical points which lie in the center of these surface patches may be chosen [35]. During docking, a transformation and rotation of one of the docking partners is searched which matches triangles of critical points with the same internal length and opposing surface normals (that means opposed geometrical features) [16]. The complexity of this part of the algorithm depends on the number of critical points. Some more physicochemical features like charges and hydrophobicity may be added to the pure geometrical representation. During the scoring step, the different hypothesis have to be ranked according to their fitness. For this step, the integration of flexibility is important, especially when unbound molecules which show induced fit during docking are taken.

For the scoring of the generated hypothesis, features like geometrical complementarity, overlap of atoms, electrostatic interactions and possible hydrogen bonds between the two molecules are taken into account [16]. In rigid body approaches, the contact area is frequently employed for scoring of hypothesis [10]. If the geometrical fit of the surfaces is used for scoring in unbound docking, flexibility has to be integrated in the scoring as well [6]. Flexibility has also to be integrated for the penalisation of intramolecular overlaps, e.g. by connecting rigid parts of the ligands by anchor points [42]. By integrating electrostatics in the scoring function, electrostatic patches of the surface which interact with each other are taken into account [17],[21]. Another approach of ranking hypothesis is the ranking by one-dimensional H-NMR spectra [2]. A theoretical NMR spectra is calculated by the docking algorithm and compared to the H-NMR spectrum of the complex.

3.1.2. Flexibility

The docking algorithm based on surface geometry with added physicochemical features works for proteins which do not change their conformation upon complex formation. For many proteins, a conformational change upon binding occurs, e.g. for the docking of trypsin (a serine protease which is found in the gut of mammals) and BPTI (bovine pancreatic trypsin inhibitor) [14],[26]. To deal with such proteins, flexibility has to be integrated in the docking algorithm [1]. According to Nussinow et al. [42], there are three kinds of flexibility: small scale fast motions (side chain movements), large scale slow motions (domain movements) and the outcome of "disordered" proteins which are due to uncompensated buried charges and small hydrophobic cores.

For the docking of ligands into proteins, the integration of flexibility is more feasible in terms of computational time, because ligands are small molecules compared to proteins. Therefore flexibility is included in protein-ligand docking algorithms [28],[8]. The scoring in protein ligand docking is done with energy calculation. In the approach of Leach et al. [28], the minimal energy within a certain range is searched with a Dead End Elimination (DEE) procedure. The energy is calculated for a fixed backbone conformation and discrete low energy conformations of the side chain, and only conformations within an energy limit are taken into account. After applying the DEE procedure, the optimal combination is searched by A^* algorithm. This graph search algorithm searches the optimal path by adding the costs for the nodes visited so far and a heuristic cost function for reaching a goal node. The conformation with the lowest cost is taken as optimal conformation for the ligand and the active site of the protein. In the approach of Clausen et al. [8], flexibility is achieved by representing the side chains not only as one rigid conformation, but by a subset of rotamer conformations called ensemble. The optimal combination of ensembles is determined by a graph search algorithm and alternate rotamer subsets are combined to a so called unified protein model. A second method to integrate flexibility is done by Najmanovich et al. [33]. They investigate side chain rearrangements upon ligand binding. They build a database of proteins in the unbound form and protein with a ligand molecule. A side chain is characterised as flexible if the χ angles change more than 40°. Taking this measurement, only a small number of binding sites change their rotamer. The tendency that large, polar amino acids tend to be more flexible compared to smaller, apolar residues can be seen. The flexibility information can be used for a reduction of the search space in protein-ligand docking, because inflexible residues of the active site do not have to be counted for flexibility calculations.

Side chain flexibility between individual residues of independently solved unbound protein structures is done by Olson et al. [48]. The test set is composed of 123 proteins with a resolution higher than 2.0 Å and a fragment length of more than 50 amino acids. To eliminate domain movements as reason for side chain flexibility, the two structures are superimposed and only residues which show an C_α-RMSD of less than 0.5 Å are chosen for the comparison. The residues are classified in buried and exposed residues. As measurement for flexibility, the angular difference in which 90% of the side chains lie was taken. Because the most flexible amino acids are omitted from the calculation by this boarder, most of the side chains or for less flexible residues all side chains changing their rotamer are excluded.

For this comparison, again the tendency that large, polar or charged residues are most flexible whereas many smaller charged or polar residues are quite inflexible holds. Side chains which are buried in the protein show a smaller extend of flexibility because of the steric hindrance. A new confidence level to evaluate the significance of predicted side chain conformations is established by Olson [48]. The frequently used 40° standard deviation for the definition of correct or incorrect prediction is too low for most of the exposed residues and too tolerant for inflexible or buried residues.

Betts and Sternberg [6] investigate flexibility between unbound proteins and complexes. The test set is build from 31 complexes with a resolution of 2.8 Å or higher in 39 different complex-unbound pairs. Residues with a B-factor (temperature factor) larger than 50 Å^2 are not taken into account. Because the B-factor is a measure for the disorder or motion of atoms, very flexible side chains with a high B-factor are excluded from investigation. This may lead to an underestimation of flexibility. The structures of the unbound protein and the complex part

were superimposed by least square fitting of the C_α atoms of non-exposed residues. The side chains are considered if they change their energy minima. As measurement for conformational differences, the C_α-RMSD, the side-chain RMSD and the percentage of side chains changing their rotamers are taken. To remove outliers, a cutoff at 95% of the residues is chosen. As benchmark for flexibility, the angular differences of unbound, independently solved structures are investigated and only conformational changes higher than these values are taken into account. It can be seen that for half of the cases, the side chain flexibility is not higher than the experimental error. For the χ_2 of exposed interface residues, rotamer changes occur more frequently. For the comparison of exposed non-interface and interface regions Sternberg and Coworkers find larger conformational changes at the interface regions. This may be due to the specific causes for which interface regions move: to form specific interactions, to avoid steric clashes or to improve shape complementarity. A comparison of the differences between an unbound protein and distinct sequence identical complexes give hints whether the complexes among each have a higher degree of similarity than the complex and the unbound structure. No more similarity for the complexes could be observed by Sternberg and Coworkers for smaller changes in the interface.

Domain movements [15] can be separated into hinge bending and shear motions. Hinge bending is an angular movement around special hinges, shear motions are movements along interfaces. Domain movements during docking are of special interest for allosteric enzymes, which show large scale domain flexibility during docking. Nussinov et al. [42] integrate domain movements in their docking algorithm. During docking, they allow rotation around picked hinges while matching the triangles of interesting points during docking (see above).

To integrate flexibility in protein-protein docking, two methods can be chosen: statistical investigation of flexibility similar to Sternberg et al. [6] in the preprocessing phase, or side chain demangling similar to ligand docking during the scoring of hypothesis to avoid steric clashes [3]. For side chain demangling, information about the probability for a rotamer combination of a side chain is needed. This information can be taken from so called rotamer libraries.

3.1.3. Rotamer Library

One possibility to describe the conformation of a side chain is the discretisation of the side chain angles in so called rotamers. A rotamer library shows the probability for each possible rotamer combination of a side chain. Rotamer libraries are built for the prediction of unknown protein structures from sequences in folding tasks. For compiling a rotamer library, the probabilities for possible rotamer combinations are calculated from known protein structures. It is assumed that the side chain of unknown protein structures prefer the some conformation as seen in the test set of known structures.

First statistical investigations of side chain placement were done by Janin and coworkers in 1978 [22]. They calculate the probability from 19 PDB x-ray structures with a resolution better than 2.5 Å by statistics and energy calculation. As accessement for the quality of their data, they compare different measurements for the same structures. They investigate the rotamer distribution over the whole data depending on ϕ and ψ ranges and SAS area. It can be seen that the data points cluster near the theoretical predicted van-der-Waals interaction peaks and in the minima of the energy function. Therefore it can be seen that statistical investigation of preferred

conformations lead to the some results compared to more time consuming energy calculations. For residues in the inner part of the protein, some deviations forced by neighbouring amino acids can be seen. For folding tasks, they show that the probability for the rotamers is a real feature of the side chain and that the most probable side chains are selected during folding.

For Secondary Structure elements, other rotamer probabilities can be seen because of repetitive backbone angles and the related steric constraints. Sternberg and Coworkers investigate the placement of side chains according to Secondary Structure. They divide a test set of 61 proteins (8064 residues) with a resolution better than or equal to 2.5 Å in α-helices, β-sheets and non-α-non-β residues. Because the effect of placement within Secondary Structures is due to interaction to both sides, the residues which lay only three positions apart from the end of the helix are treated separately to get more pure data. Sternberg describes a few effects of the backbone angles on the χ_1 distribution. For $\alpha - helices$, there are two main groups: the first group consists of TRP, TYR, PHE, HIS, MET, GLU, GLN, LYS, ARG, LEU and CYS residues. For these residues, the first rotamer is nearly forbidden because of steric clashes with the $i_{th} - 3th$ residue. In the third rotamer, there are steric clashes as well. These steric clashes are not that severe, so that the probability for the third rotamer is reduced. For shorter residues with hydrogen binding capacity in the side chain (SER, THR), a shift towards the third rotamer can be seen due to hydrogen binding. For $\beta - sheets$, several changes which are due to complicated interactions can be seen.

Statistical investigation of the conformation of all angles within a side chain lead to rotamer libraries. Common rotamer libraries are built on the data of unbound structures. Ponder and Richards [39] chose 19 high quality PDB structures for compilation of their rotamer library, with R factors less than 0.18. By this high quality data, they were able to reduce the standard deviation of the χ angles in comparison to other rotamer libraries. Tuffery et al. [46] use cluster analysis to describe their rotamers, but they do an energy minimisation procedure before calculating the probability instead of using the raw data. Because of this, unfavourable rotamers which may be seen in nature are omitted and the rotamer library has a bias for energetical favoured rotamers.

In the approach of Schrauber et al. [44], the rotamericity of side chains is investigated to improve the library from Ponder et al. [39]. χ angles are considered as rotameric if they are not further than 20° from the mean of a rotamer. They use 70 chains from 68 pdb structures with a resolution better than 2 Å to show that large deviation from mean χ values can not be attributed merly to errors in the crystal structure determination, but occur systematically. To investigate backbone dependency, the divide the ϕ, ψ angle range in eight backbone classes according to Secondary Structure. They show that the rotamericity of side chain angles depend on backbone conformation and secondary structure, and the environment of the side chain.

The newer approach by Lovell et al. [30] uses 240 high quality PDB structures with a resolution better than 1.7 Å. For compiling the rotamer library, the modes instead of the average values are chosen, because the mode is not influenced by outliers.

In the approach of Dunbrack and coworkers [11, 7], Bayes statistic was chosen. The probability used in Bayes statistics consists of an a priori and a posteriori probability. The a priori probability can be calculated from part of the data or by an assumption how the data is distributed. The posterior distribution is calculated from the data. In regions with sparse or no data, the probability depends mainly or only on the prior distribution, in well populated regions, the

probability depends on both distributions. Dunbrack uses the data to calculate the probability of the prior distribution. Because for many unfavourable rotamer combinations only few or no examples can be found in the test set, the dependent probability for the whole rotamer combinations were calculated as multiplication of the dependent probability of the two angles.

$$p(r234|r1) \propto p(r2|r1) * p(r3|r2) * p(r4|r3) \tag{3.1}$$

The advantage of this approach is that more data for the combinations of two χ angles are available in comparison to data for the whole combination. The χ angles are modelled to be dependent only on the previous χ angle. For the counting of the different rotamer combinations, Dunbrack uses the adding one method, where 1 is added to the counts of each combination. By this method, zero probabilities are avoided.

The ranges of the rotamers used by Dunbrack and Coworkers for higher χ angles differ from the IUPAC rotamers [20]. They take into account crystallographic uncertainty, e.g. in branched amino acids, where the position of the functional group – which is important for the determination of the χ angle – can not be defined by crystallographers. To overcome this nomenclature problem, Dunbrack rotamers for branches the regions 180° apart are counted for the some rotamer. For higher χ angles of ring systems, Dunbrack and Coworkers use two rotamers, with the ring placed parallel or antiparallel to the backbone. In this case, the positions 180° away are again counted for the same rotamer, because these two positions are reflecting the two orientations of the ring system which is placed in the some way according to the backbone. For higher χ angles with tetrahedral C-atoms, the IUPAC rotamers are chosen.

3.2. Goal

The goal is to calculate a statistical based flexibility measurement which can be integrated in the scoring function of a 1:n protein-protein docking algorithm. Because in 1:n docking, many proteins have to be screened, the time to rank one hypothesis for one protein should be kept short. Therefore energy calculations for the validation of hypothesis are not feasible to rank the hypothesis. In contrast to Olson et al. [48], only side chains which change their rotamer upon complex formation are taken into account. The flexibility parameters are used for the scoring of hypothesis where steric clashes are penalised. It is assumed that flexible side chains which show a steric clash in a hypothesis of two unbound structures will move away so that the penalty for this steric clash can be reduced compared to inflexible residues. By weighting steric clashes, flexibility can be integrated without being too flexible. If too much flexibility is allowed, one inhibitor which would not bind to the target protein in vivo can be fitted into the target protein.

Another possibility to integrate side chain flexibility in protein-protein docking is side chain demangling, where clashing side chains are moved away according to their probability of conformation. Therefore, new rotamer libraries are compiled with the data for unbound and complexed proteins and for the data of different Secondary Structure elements. In contrast to Dunbrack et al. [11], a statistical approach from natural language processing is used where zero probabilities for rotamer combinations are avoided by redistributing the probability (see section 4.3.1). In addition information about the preferred direction of movement for flexible

side chains is calculated from the test set. Therefore the side chains can be additionally placed according to this direction, e.g. if the next probable conformation according to the rotamer library will lead to a steric clash as well.

4. Conformation of side chain angles

The χ angles describe the rotation around the bonds between the side chain atoms. There can be up to four torsions depending on structure and length of the side chain. The first χ angle is defined as the rotation between the C_α and C_β atoms, higher χ angles are defined by the rotation between the following atoms of the side chain ($C_\gamma, O_\gamma, C_\delta, O_\delta, N_\delta, C_\epsilon, O_\epsilon, N_\epsilon$). For branched side chains, the position of the functional group is chosen for the calculation of the angle [20].

In this section, the probabilities for different side chain conformations are described. The conformation of the side chain is determined by the side chain angles χ. In the first part of the chapter the calculation of χ angles and their discretisation[1] is described. After the introduction, the rotamer distributions for single χ angles are given followed by the dependent probabilities for combinations of two χ angles and χ_1 given the backbone conformation. The probabilities for the whole rotamer combination of the side chain can be seen in the rotamer libraries at the end of the section. The rotamer libraries can be used for placement of clashing side chains during docking in side chain demangling tasks.

The side chain angles of ALA, GLY, and PRO are not investigated because they are non-rotameric (see A.2 in the appendix). The side chains of ALA residues consist of a methyl group. Because of the symmetry of this group, no different rotamers can be assigned. The same holds for GLY residues where the side chain is composed of a small proton. PRO residues are iminoacids with their side chains bound to the NH group of the backbone forming a ring system. Some pseudo-rotameric puckering states can be assigned to PRO residues, but because of the additional bond to the backbone, the flexibility of PRO residues can not be investigated without taking into account backbone flexibility and is not considered throughout the thesis.

4.1. Calculation of Torsion Angles

Torsion angles can be calculated from the relative position of four atoms. The atoms used are depending on the side chain and defined by the IUPAC nomenclature [20]. The calculation is invariant to the position of the atoms in the coordinate system and torsion angles of different independently solved structures can be compared. The relative location of the different side chain angles can be seen in figure 4.1. The side chain atoms are shown in red, the angles can be seen in blue.

From the four atoms used for calculation, three vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are build, which span two intersecting planes. The intersection angle between the normals of the two planes a, b is the

[1] In pattern recognition the assignment of data to different classes is called quantisation. Because the term discretisation is often used for angle rotamers, it will be used in this thesis for consistency.

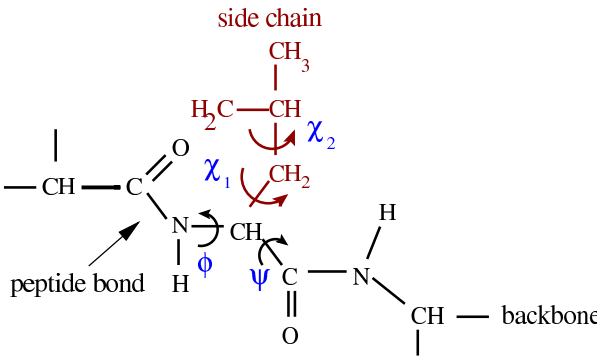


Figure 4.1.: relative position of different angles

torsion angle χ between the atoms (see figure 4.2). Its value can be calculated according to the following formula:

$$\chi = arcos \frac{a \times b}{||a||||b||} \tag{4.1}$$

To define the algebraic sign, the inner product between the vector ($\vec{v}_2$) and the plane which is defined by the normals ($\vec{a}, \vec{b}$) is calculated:

$$cos\chi = \frac{\langle \vec{v}_2(a \times b) \rangle}{||\vec{v}_2||||(a \times b)||} \tag{4.2}$$

If the value of *cos* χ is positive, the angle is smaller than 90° and the two vectors are parallel to each other. If this value is negative, the angle is larger than 90° and the two vectors are antiparallel. If the vectors are perpendicular, the values is zero.

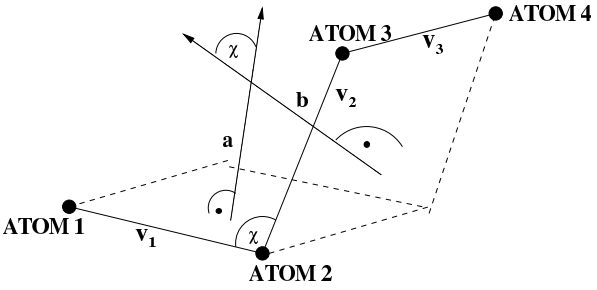


Figure 4.2.: Calculation of torsion angle

4.1.1. Discretisation of torsion angle in rotamers

One possibility for describing the conformation of a protein side chain is the description as set of discretised rotamers. Compared to float values, this description is easier to handle for different χ angles. For non-branched side chains the angle range of 360° is divided into three rotamers, but different ranges may be chosen for different structured side chains (see below). By discretisation of the conformation, only larger conformational changes upon complex formation

can be noticed. The borders of the rotamers are characterised by energy maxima. Changing the rotamer means crossing an energy barrier so that these larger conformational changes lead to a side chain in a new energy minimum.

The ranges of the rotamers are due to the geometry of bonds. For tetrahedral C-atoms which have sp^3 hybrid orbitals, the rotamers reach from $0° - 120°$, $120° - 240°$ and from $240° - 360°$ according to IUPAC nomenclature [20]. The four bonds of tetrahedral $C - atoms$ are pointing towards the four edges of a tetraeder. The three rotamers reflect these possible positions of the rest of the side chain between the three atoms bond to the C_α without steric clashes [20]. Not all C-atoms in amino acid side chains are tetrahedral, e.g. in higher χ angles double bonded C-atoms can be found. Because the geometry of these bonds differ, other rotamer ranges introduced by Dunbrack et al. [11] are chosen.

For branched amino acids the exact position of the functional group can not be seen in the electron density map. For the rotamer nomenclature, the position of the functional group is important, because the group with the highest priority (the one with the functional group) is taken for the calculation of the χ angle [20]. For higher χ angles where the exact position of the functional group is not clear, the rotamer may be calculated with the wrong atoms. To overcome this problem, the range 180° apart are taken for the same rotamer [11]. These positions are simular concerning steric hindrance, so that the difference does not matter for the calculation of rotamer probabilities. For VAL, the two ends of the side chain are indistinguishable so that different rotamers for the two ends are not reasonable. The mirrored positions of the side chains are simular concerning steric hindrance.

For ring systems, the two positions 180° apart are similar concerning the steric constraints, because it is the some placement of the ring in different orientations. Therefore these ranges are also counted as a single rotamer. The two possible rotamers for ring systems are the two positions parallel and antiparallel to the backbone. The rotamers for each amino acid can be seen in table 4.1. For divided rotamers, the region 180° is not shown.

	Rotamers		
	1	2	3
χ_1 all aa χ_2 ARG, GLN, GLU, ILE, LEU, LYS, MET χ_3 ARG, LYS, MET χ_4 ARG, LYS	$0 - 120°$	$120 - 240°$	$240 - 360°$
χ_2 ASN, ASP χ_3 GLN, GLU	$30 - 90°$	$330 - 360°$ & $0 - 30°$	$270 - 330°$
χ_2 PHE, TYR, HIS, TRP	$30 - 150°$	$330 - 360°$ & $0 - 30°$	

Table 4.1.: Rotamer ranges for different amino acids, the regions 180° apart for non-tetrahedral C-atoms belonging to the same rotamer are not shown

To point out in which rotamer a side chain is, the following nomenclature is chosen: r1 stands for χ_1, r2 for χ_2, the number afterwards gives the rotamer, e.g. r1=1 means χ_1 is in the first rotamer.

4.2. Distribution of χ angle rotamers

In the following part of the thesis, the rotamer distributions for single χ angles over the whole test set and in different Secondary Structures are described (sections 4.2.1 to 4.2.3) followed by the χ_1 distribution depending on the backbone in section 4.2.4. Afterwards the dependent probabilities for χ angle combinations are given in section 4.2.5. In the last section of the chapter the probabilities for whole angle conformations are given in backbone independent (4.3.1) and backbone dependent (4.3.2) rotamer libraries.

4.2.1. Rotamer distribution for χ_1

The C_α atom is tetrahedral (see section 4.1.1) with the bonds pointing towards the four edges of a tetraeder. Therefore the rotamer distribution for χ_1 is trimodal for all amino acids (cf. figure 4.3). The three rotamers reflect the possible position of the side chain between the three atoms bond to C_α. Side chains at the rotamer border clash with the backbone atoms so that the borders are characterised by high energy levels. Sparse or no data can be found at the rotamer borders.

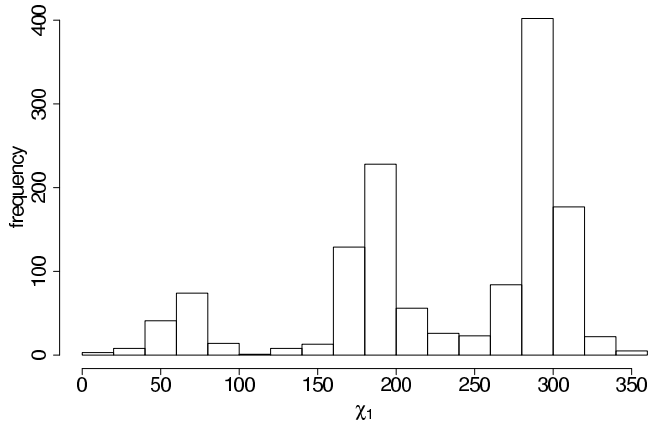


Figure 4.3.: χ_1 rotamer distribution for ARG residues

As an example for a χ_1 rotamer distribution, the histogram for ARG residues is shown in figure 4.3. The distribution is trimodal with a preference for the third rotamer.

The χ_1 rotamer distribution of all side chains can be seen in table 4.2. The histograms of the χ_1 distribution are shown in the appendix (B). Two groups can be distinguished: in the first

AA	#	counts and probabilities for χ_1 rotamers					
		P(r1=1)	P(r1=1)	P(r1=2)	P(r1=2)	P(r1=3)	P(r1=3)
ARG	2536	261	0.1029	971	0.3829	1304	0.5142
ASN	7897	622	0.0788	3115	0.3945	4160	0.5268
ASP	3616	440	0.1217	1871	0.5174	1305	0.3609
CYS	6321	580	0.0918	1172	0.1854	4569	0.7228
GLN	4450	282	0.0634	1277	0.2870	2891	0.6497
GLU	1840	162	0.0880	208	0.1130	1470	0.7989
HIS	1496	415	0.2774	868	0.5802	213	0.1424
ILE	6910	1519	0.2198	1081	0.1564	4310	0.6237
LEU	7249	2	0.0003	2509	0.3461	4738	0.6536
LYS	4551	358	0.0787	1700	0.3735	2493	0.5478
MET	1199	18	0.0150	404	0.3369	777	0.6480
PHE	1881	830	0.4413	51	0.0271	1000	0.5316
SER	13911	6454	0.4639	2604	0.1872	4853	0.3489
THR	5205	2399	0.4609	71	0.0136	2735	0.5255
TRP	2717	411	0.1513	601	0.2212	1705	0.6275
TYR	4754	1240	0.2608	1027	0.2160	2487	0.5231
VAL	8325	446	0.0536	7094	0.8521	785	0.0943

Table 4.2.: Probabilities for different χ_1 rotamers

group, the highest probability can be seen for the third rotamer [2]. In this position, the rest of the side chain is placed opposite to the largest atom bound to the C_α between the N and the $H-atom$. It is preferred especially for larger side chains. For THR residues, a more bimodal distribution can be seen (see B). This side chain is branched at the χ_1, the two ends of the side chain consisting of a *methylgroup* and an OH group. The torsion angle is defined by the position of the OH group, because the O atom has a higher priority compared to the methyl group [20]. If the OH group is in the third rotamer, the more bulky methyl group is in the second rotamer. If the side chain is in the second rotamer (the OH group), the larger CH_3 group would be in the unfavourable first rotamer. Therefore the probability for this rotamer is low. The (normally unfavourable) first rotamer has a higher probability compared to the second rotamer, because the larger methyl group is in the third rotamer if the OH group is in the first rotamer.

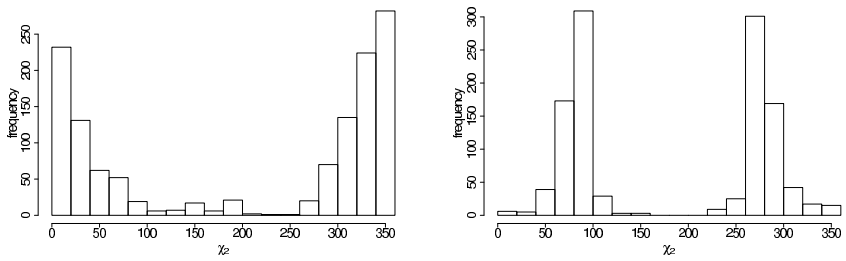
VAL, HIS and ASP are in the second group with the higher rotamer probability for the second rotamer. The side chains of these residues are branched. VAL has two identical *methylgroups* attached to the C_β, so that the second and third rotamer are identical. HIS and ASP are branched at the C_γ atom. Because of the branch, two ends of the side chain have to be positioned, the placement in the third and second rotamer for the two ends are favoured. SER residues prefer the r1=1 rotamer with a probability of nearly 50%. Normally this rotamer is the most unfavourable, because the side chain is positioned between the most bulky atoms around the C_α [22]. Because SER is a small residue and the OH group causes less steric hindrance, a more free rotation around the $C_\alpha - C_\beta$ bond is allowed. Furthermore, the OH group can form hydrogen bonds with backbone atomes, which also stabilises otherwise unfavourable positions. For CYS residues, the distribution is influenced by neighbouring CYS residues because of sulphur bridges which can be build between two SH groups of CYS residues.

4.2.2. Rotamer distribution of higher χ angles

The second rotamer describes the position of the rest of the side chain at the C_β. Because the C_β can be bond to other atoms than $C-atoms$ (e.g. O_γ, N_γ) or can be double bonded to

[2] Amino acids belonging to this group are: ARG, ASN, CYS, GLN, GLU, ILE, LEU, LYS, MET, PHE, THR, TRP, TYR

other atoms, not all C_β atoms are sp^3 hybridised with a trimodal distribution. For example the distribution for branched amino acids with a $C = O$ double bond is more bimodal.



(a) ASP (b) TYR

Figure 4.4.: χ_2 distribution for ASP and TYR residues

In figure 4.4, the χ_2 distributions of ASP and TYR residues are exemplarily shown. For side chains with a ring system like TYR (see figure 4.4(b)), there is sparse or no data in the second rotamer. A side chain in this rotamer would be antiparallel to the backbone leading to steric clashes. In the two remaining rotamers, the ring system is placed parallel to the backbone in two possible orientations, which is much more preferable.

The χ_2 rotamer distribution of ASP 4.4(a) is bimodal because of the branch after the C_β atom. It is double bonded to an O-atom and therefore has a sp^2-hybrid orbital. The branch will be positioned parallel to the backbone, with the negative charge positioned between the two C-atoms.

AA	probabilities for χ_2 rotamers			probabilities for χ_3 rotamers			probabilities for χ_4 rotamers		
	P(r2=1)	P(r2=2)	P(r2=3)	P(r3=1)	P(r3=2)	P(r3=3)	P(r4=1)	P(r4=2)	P(r4=3)
ARG	0.1092	0.7535	0.1372	0.3155	0.4456	0.2390	0.2086	0.5694	0.2220
ASN	0.1388	0.2773	0.5839	–	–	–	–	–	–
ASP	0.1358	0.5409	0.3233	–	–	–	–	–	–
GLN	0.1317	0.5939	0.2744	0.4764	0.1948	0.3288	–	–	–
GLU	0.2440	0.3560	0.4000	0.0707	0.6136	0.3158	–	–	–
HIS	0.9886	0.0114	0.0000	–	–	–	–	–	–
ILE	0.0845	0.8507	0.0648	–	–	–	–	–	–
LEU	0.3718	0.6213	0.0069	–	–	–	–	–	–
LYS	0.0512	0.7132	0.2356	0.0901	0.7882	0.1217	0.0769	0.8433	0.0798
MET	0.0017	0.9892	0.0092	0.4896	0.4854	0.0250	–	–	–
PHE	0.7592	0.2408	0.0000	–	–	–	–	–	–
TRP	0.3820	0.0088	0.6091	–	–	–	–	–	–
TYR	0.9975	0.0025	0.0000	–	–	–	–	–	–

Table 4.3.: Rotamer probabilities for higher χ angles

In table 4.3, the probabilities of the different rotamers in higher χ angles are shown. The histograms of the χ angle distributions can be seen in the appendix B. For tetrahedral C_β atoms in the side chains of ARG, GLN, GLU, ILE, LEU, LYS and MET, the distribution is trimodal. For ARG, LYS and MET residues, the second rotamer is preferred. These residues have long side chains without branches and the second rotamer of χ_2 leads to a relaxed, stretched conformation. Some branched side chain prefer the second rotamer as well (ILE, LEU, GLN). For ILE and LEU, two parts of the branch have to be positioned, so that one part is in the r2=2 rotamer, whereas the other part is in the r2=3 rotamer. For ILE residues, the probability of the second rotamer is higher compared to LEU. ILE has a $methyl - group$ attached to χ_1

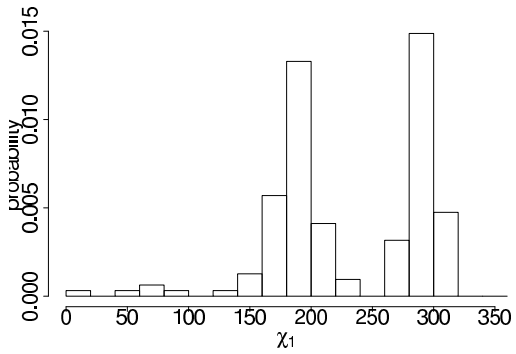
which is placed in the third rotamer if the $ethyl - group$ is in the second rotamer. For LEU residues, the probability for the second rotamer is reduced because the two ends of the side chain are equal (two $methyl - groups$ attached to the C_β). For GLU residues, the second and third χ_1 rotamer have a nearly equal probability. GLU and GLN are branched after the $C\gamma$ atom, so the branch influences the conformation of χ_2. The GLU side chains are negatively charged, the position of the charge is influenced mainly by χ_2, because the charge at the end of the side chain is positioned in the middle between the $C{=}O$ and $C\text{-}OH$ group. Because of this position, the conformation of χ_3 does not influence the position of the charge. The χ_2 conformation is more important for placing the charge. Charges have to be neutralised inside a protein or will lead to a destabilisation of the structure. The charge seems to be neutralised with r2=2 or r2=3. Side chains with ring systems prefer r2=1 and r2=3. These rotamers differ by the orientation of the ring, but are the same due to steric hindrance because the ring is placed parallel to the backbone for both rotamers.

For χ_3 tetrahedral $C - atoms$ prefer the third rotamer (e.g. ARG, LYS, MET). GLN and GLU are branched after the C_γ. For GLN residues, a more bimodal distribution can be seen whereas for GLU, the second rotamer is preferred, so that the two parts of the branch lie in the second and third rotamer. The position of the charge is not influenced by the χ_3 rotamer (see above), so that the sterical preferred conformation can be chosen.

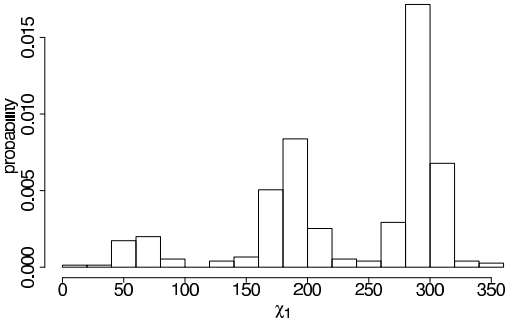
There are two residues with four χ angles: ARG and LYS. For both cases, the $C_\delta - atoms$ are tetrahedral, so that the distribution is trimodal. The main peak can be found in the second rotamer.

4.2.3. Rotamer distribution subject to Secondary Structure

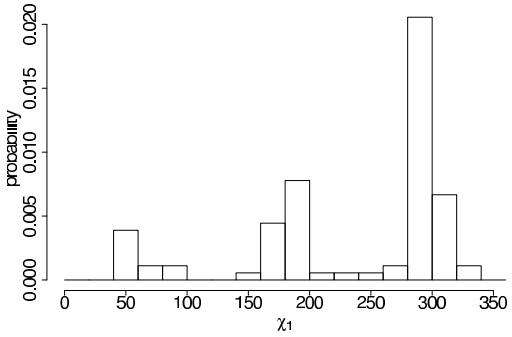
To get a first hint on the backbone dependency of side chain placement, differences of the rotamer distribution according to the Secondary Structure are investigated. The classification of the Secondary Structure elements for the different PDB entries from the test set is done with the DSSP algorithm [23] which calculates the Secondary Structure for given atomic coordinates. For special Secondary Structure elements, the backbone angles lie in distinct ϕ, ψ ranges. The fixed, repetitive backbone angles lead to sterical constraints and different probabilities for the placement of side chains in Secondary Structure elements due to steric clashes with the backbone in some rotamers. These constraints are moderated for ends of these elements because the following residues do not belong to the Secondary Structure and therefore are not in the fixed backbone ranges which lead to steric clashes. Because of these different constraints, the end residues are excluded in the investigation of McGregor et al. [31]. In helix types other than $\alpha - helices$, the backbone is distorted as well, so that they are also excluded by McGregor et al. [31]. In the thesis the investigation of the rotamer distribution in Secondary Structures is taken as a first hint for the backbone dependency of the conformation. Other helix type and end residues are included in the test set, because these residues are still in fixed backbone ranges. The following groups of Secondary Structure elements are investigated: helices ($\alpha - helix$, $3_{10} - helix$, $Pi - helix$), $\beta - strands$ (extended $\beta - ladder$), turns between elements of Secondary Structure and random coils (RND), that means residues which are in no Secondary Structure.



(a) Helices

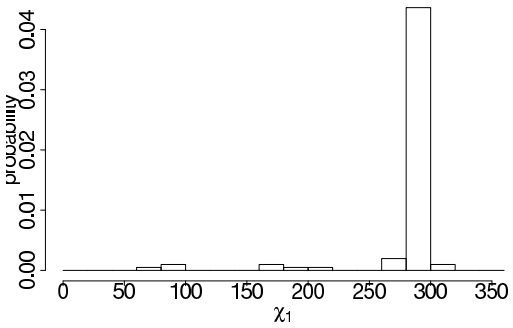


(b) Coils

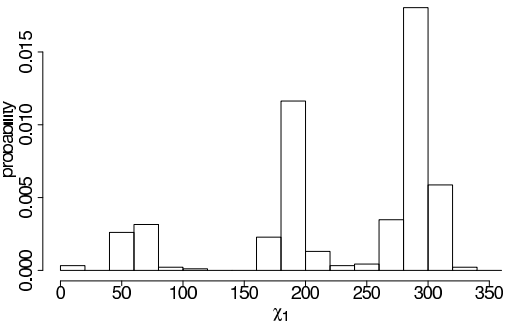


(c) β Sheets

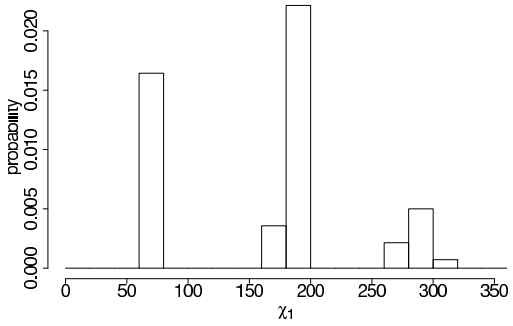
Figure 4.5.: χ_1 distributions for ARG residues



(a) Helices

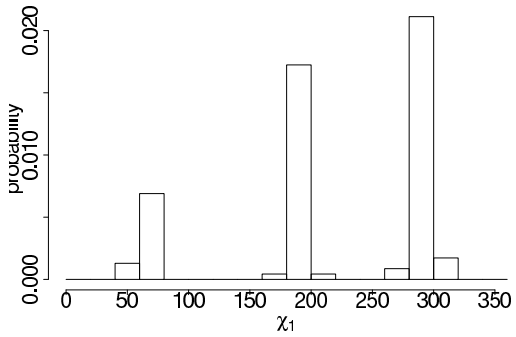


(b) Coils

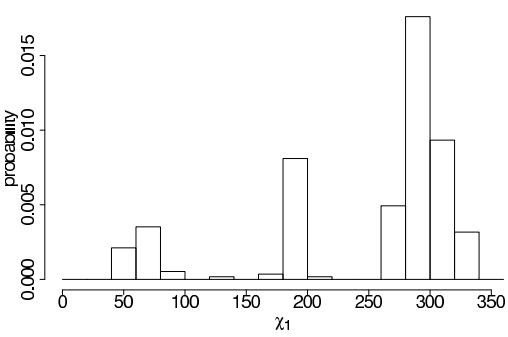


(c) β sheets

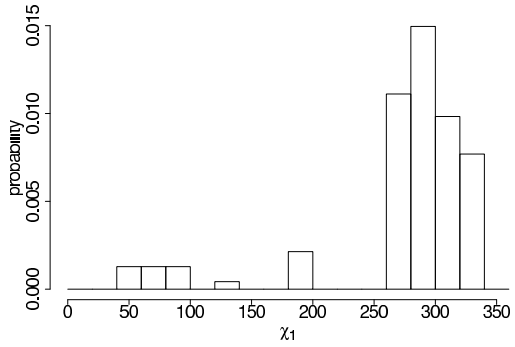
Figure 4.6.: χ_1 distributions for ASN residues



(a) Helices



(b) Coils



(c) β Sheets

Figure 4.7.: χ_1 distributions for CYS residues

The histograms 4.5, 4.6, and 4.7 show the rotamer distributions for unbound ARG, ASN and CYS residues depending on the Secondary Structure. Some differences between helices and sheets compared to the residues in random coils can be seen. Due to the restricted range of backbone angles, some unfavourable interactions on the one hand and some hydrogen binding possibilities on the other hand can be seen. These interactions have been described by McGregor et al. [31] for standard α-helices ($\phi = -57°$, $\psi = -47°$). In the first χ_1 rotamer, the C_γ atom of residue$_i$ and the Carbonyl Oxygen of residue$_{i-3}$ have a distance of 1.9 Å in standard $\alpha-helices$, so that these conformation is forbidden for long or bulky amino acids. This can be seen in the histograms for ARG in figure 4.5(a) and ASN residues in figure 4.6(a). This rotamer can be taken by a few residues if some distorsion of the $\alpha-helix$ occurs which increases the distance between these atoms or in other helix types [31]. Because other helix types and end residues are investigated as well (see above), this rotamer has an even higher probability for CYS residues in *helices* (see figure 4.7(a)). In the third rotamer an unfavourable interaction between the C_γ of residue$_i$ with the Carbonyl Oxygen of residue$_{i-4}$ which is not as severe as in the first rotamer because of the higher distance of 2.9 Å is described. Because of the inclusion of 3_{10} and *Pi helices*, there is only a slight reduction for this rotamer. In the second rotamer, the side chain is pointing outwards with no steric hindrance. For the core of $\alpha-helices$, this rotamer is energetically preferred.

For $\beta-sheets$, no clear tendencies can be seen in the histograms because the steric constraints in this Secondary Structure differ (see below). Shifts occuring in $\beta-sheets$ are described below. The complete data for all residues can be seen in table 4.4.

AA	probability for χ_1 in coils			probability for χ_1 in Helices			probability for χ_1 in Strands		
	P(r1=1)	P(r1=2)	P(r13)	P(r1=1)	P(r1=2)	P(r1=3)	P(r1=1)	P(r1=2)	P(r1=3)
ARG	0.2533	0.3067	0.4400	0.0316	0.5127	0.4557	0.1222	0.2778	0.6000
ASN	0.0870	0.5000	0.4130	0.0294	0.0392	0.9314	0.3286	0.5143	0.1571
ASP	0.1927	0.6055	0.2018	0.0417	0.1563	0.8021	0.0714	0.1905	0.7381
CYS	0.1628	0.2791	0.5581	0.1638	0.3621	0.4741	0.0769	0.0513	0.8718
GLN	0.0714	0.1667	0.7619	0.0083	0.4417	0.5500	0.0156	0.4609	0.5234
GLU	0.0222	0.2778	0.7000	0.1061	0.3182	0.5758	0.2222	0.4815	0.2963
HIS	0.0196	0.6667	0.3137	0.4651	0.1395	0.3953	0.0750	0.7500	0.1750
ILE	0.0952	0.4762	0.4286	0.1883	0.0844	0.7273	0.1505	0.0699	0.7796
LEU	0.0059	0.2899	0.7041	0.0081	0.3765	0.6154	0.0083	0.4292	0.5625
LYS	0.0280	0.1308	0.8411	0.0327	0.4444	0.5229	0.1324	0.3897	0.4779
MET	0.0000	0.3333	0.6667	0.0000	0.2281	0.7719	0.0185	0.4815	0.5000
PHE	0.1220	0.1951	0.6829	0.0122	0.4268	0.5610	0.4301	0.1290	0.4409
SER	0.5721	0.1674	0.2605	0.5155	0.0928	0.3918	0.2825	0.3898	0.3277
THR	0.5909	0.0000	0.4091	0.2479	0.0598	0.6923	0.2845	0.0776	0.6379
TRP	0.0000	0.3846	0.6154	0.0135	0.6216	0.3649	0.4600	0.0000	0.5400
TYR	0.0918	0.4388	0.4694	0.1818	0.3333	0.4848	0.2581	0.0430	0.6989
VAL	0.0610	0.6829	0.2561	0.1143	0.7600	0.1257	0.0749	0.8241	0.1010

Table 4.4.: Probabilities for different χ_1 rotamers in Secondary Structure elements

In table 4.4, the rotamer probabilities for *helices* and *sheets* compared to random coils are shown. For some rotamers, no representatives are found in the test set, so that zero probabilities can be seen. As comparison, the data for pure $\alpha-helices$ without end residues can be seen in table 4.5. Because of the exclusion of other helix type and end residues, the data in this table is comparable to the data of McGregor et al. [31].

For helices a reduction of the probabiltiy for r1=1 can be seen for ARG, ASN, ASP, GLN, PHE, SER, THR and TRP. This effect is due to the steric clash of a side chain in this rotamer with the backbone (see above). This reduction can not be seen for all residues. The r1=1 probability decreases for GLU, HIS, ILE, TYR and VAL residues in the core of $\alpha-helices$ (see table 4.5), therefore the high probability for all helix types is due to end effects and distorted backbone angles. For CYS residues the probability for random coils and *helices* levels off and

AA	n all	# r1	P(r1=1)	# r2	P(r1=2)	# r3	P(r1=3)
ARG	71	3	0.0423	39	0.5493	29	0.4085
ASN	35	0	0.0000	1	0.0286	34	0.9714
ASP	43	0	0.0000	15	0.3488	28	0.6512
CYS	47	0	0.0000	38	0.8085	9	0.1915
GLN	66	1	0.0152	28	0.4242	37	0.5606
GLU	76	0	0.0000	31	0.4079	45	0.5921
HIS	6	0	0.0000	5	0.8333	1	0.1667
ILE	83	2	0.0241	5	0.0602	76	0.9157
LEU	151	1	0.0066	69	0.4570	81	0.5364
LYS	94	0	0.0000	28	0.2979	66	0.7021
MET	36	0	0.0000	7	0.1944	29	0.8056
PHE	57	0	0.0000	24	0.4211	33	0.5789
SER	64	21	0.3281	13	0.2031	30	0.4688
THR	74	7	0.0946	5	0.0676	62	0.8378
TRP	51	0	0.0000	40	0.7843	11	0.2157
TYR	14	0	0.0000	5	0.3571	9	0.6429
VAL	56	0	0.0000	54	0.9643	2	0.0357

Table 4.5.: Counts and probabilities for a change of χ_1 rotamers in unbound α- helices without end residues

a reduction of the r1=1 probability occurs in the core of $\alpha - helices$ as well. The probability of the second rotamer increases for ARG, CYS, GLN, LEU, LYS, THR, TRP, TYR and VAL residues in *helices*. The second rotamer is most favourable because the side chain is pointing outwards the helix (see above), it is stronger in $\alpha - helical$ cores except for LYS residues. For PHE and THR residues, the probability for this rotamer is at the same level for all helix types and $\alpha - helices$. For ASN, ASP and MET residues, a shift from the second to the third rotamer is shown. For HIS residues, this shift can be seen for all helix types, but not in $\alpha - helical$ core residues. The effect is very pronounced for ASN and ASP residues. This more unfavourable rotamer may be stabilised by hydrogen bonds (ASN) or by electrostatic interactions.

For $\beta - sheets$, conformational constraints for several residues differ because parallel $\beta - strands$ may interact with each other or the strands may be twisted at the ends. Therefore in contrast to helices, no clear tendencies can be detected for $\beta - sheets$, but some shifts in the rotamer distribution compared to random coils can be seen. The side chains of ARG, THR show a shift from the first to the third χ_1 rotamer, for SER residues, a shift from the r1=1 to r1=2 can be observed. ILE residues show an increase in the third rotamer and a decrease in the second rotamer. For TYR residues, a decrease of the probability for r1=2 can be seen. A shift away from the third rotamer can be observed for ASN, LYS, PHE, GLU, GLN, MET and VAL. The first three residues show an increase for the first rotamer, whereas for the last three residues, a shift to the second rotamer can be seen. For ASP and CYS residues, the increase can be seen in the third rotamer. Because of the different steric constraints, many changes can be seen for residues in β sheets and in contrast to *helices*, no clear rules can be found for the change of the probability for different rotamers.

Because steric constraints within Secondary Structure elements differ, even within *helices* due to backbone distorsion and placement at the ends of helices, it is more accurate to use smaller ϕ and ψ ranges compared to regions of Secondary Structure elements for the placement of side chains.

4.2.4. Rotamer distribution depending on ϕ and ψ angles

In the previous section the dependency of the χ_1 distribution on Secondary Structure had been shown as first hint for backbone dependency. For Secondary Structure elements, the backbone ranges are fixed. In this chapter, the χ_1 distribution depending on the whole backbone angle

range is investigated. The ϕ and ψ angles are discretized into different rotamers comparable to the side chain rotamers. The backbone angle range of 360° for each backbone angle is divided into 18 rotamers with a range of 20°, starting at −180° with the first rotamer. The 18th rotamer runs from 160° to 180°.

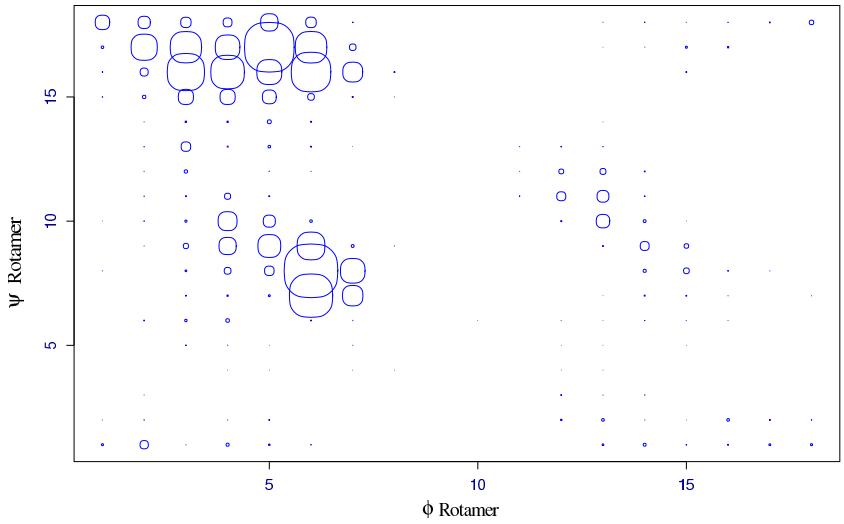


Figure 4.8.: Counts for different ϕ and ψ rotamers

The probabilities for different Phi and Psi rotamers can be see in figure 4.8. This plot is called Ramachandran plot [41] and shows the allowed backbone regions for proteins. The diameter of the circles correlate with the number of representatives being in this backbone rotamer combination. It can be seen that in some regions of the plot many representatives can be found, whereas other backbone conformations are forbidden.

Figure 4.9 shows the rotamer distributions of the first χ angle depending on the backbone rotamers. Along the x-axis, the rotamer for the ϕ angles can be seen, the y-axis shows the ψ rotamers. The χ_1 rotamer is shown on the z-axis. The diameter of the circles correlates with the probabilty for the rotamer combination. It can be seen that the probability for χ_1 rotamers vary according to the ϕ, ψ conformations.

The widest range of backbone conformations is allowed for the third χ_1 rotamer. Because this is the most favourable rotamer concerning the placement of the side chain alone (cf. section 4.2.1), it is allowed for more unfavourable backbone rotamers. For the r1=1 rotamer in which the placement of the side chain is not optimal, the probability for many backbone regions are reduced. The different probabilities for $P(\chi_1 \mid \phi)$ and $P(\chi_1 \mid \psi)$ for ARG residues are given in table 4.6.

The range of backbone rotamers used for table 4.6 is 40°, starting at −180° with rotamer 1. $P(\chi_1 \mid \phi, \psi)$ is not shown because the values are very small. The probabiltiy of the χ_1 rotamer varies according to the backbone conformation. The most restricted χ rotamer is r1=1, which has the lowest probability without taking into account the backbone conformation. Because of

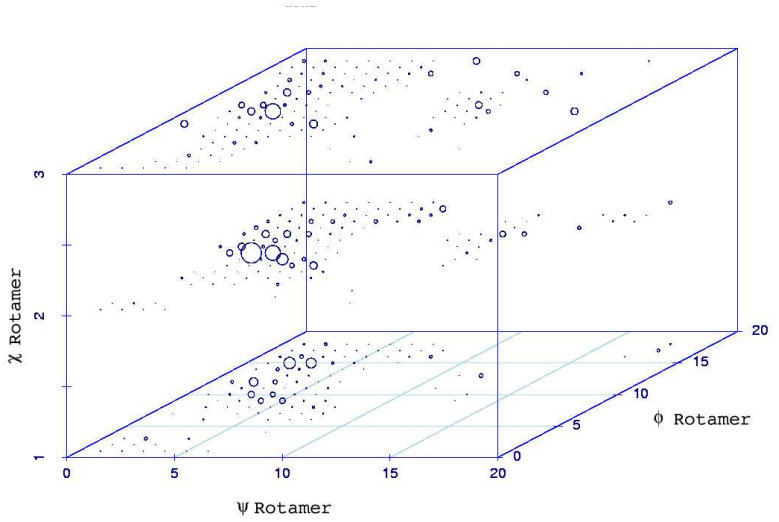


Figure 4.9.: χ_1 rotamer probabilities for different backbone rotamers

Rot1UDun	Phirot	$P(\chi_1 \mid \phi)$	Psirot	$P(\chi_1 \mid \psi)$
1	1	0.0318	1	0.0007
1	2	0.025	2	0.0003
1	3	0.028	3	0.0041
1	4	0.0033	4	0.0091
1	5	0.0000	5	0.0136
1	6	0.0006	6	0.0056
1	7	0.0003	7	0.0005
1	8	0.000003	8	0.0056
1	9	0.000008	9	0.0189
2	1	0.02498	1	0.0003
2	2	0.0333	2	0.0007
2	3	0.1942	3	0.0056
2	4	0.0455	4	0.0851
2	5	0.0001	5	0.0045
2	6	0.0014	6	0.0022
2	7	0.0003	7	0.0056
2	8	0.0001	8	0.057
2	9	0.0002	9	0.0136
3	1	0.028	1	0.0041
3	2	0.1943	2	0.0056
3	3	0.2236	3	0.0022
3	4	0.0265	4	0.0741
3	5	0.0001	5	0.0616
3	6	0.0174	6	0.0266
3	7	0.0117	7	0.0049
3	8	0.0001	8	0.0483
3	9	0.0003	9	0.0502

Table 4.6.: Backbone dependent rotamer probabilities for ASP residues

the unfavourable placement of the rest of the side chain between the largest atoms bound to the C_α, it is only allowed for some backbone conformations. The third χ_1 rotamer has the highest probability for all backbone ranges (compare figure 4.9). In this rotamer, the rest of the side chain is in the most favourable positition concerning the C_α atom. The most favourable region for the ϕ rotamer is the second and third rotamer ($-140°$ to $-60°$ in this 9-rotamer notation). This region has the highest probabiltiy for all χ_1 rotamers. For the ψ angle, the fourth and fifth rotamer is preferred ($-60°$ to $20°$). Because the backbone conformation influences the placement of side chains, a more accurate placement can be achieved by backbone dependent placement.

The possible conformation of a side chain is influenced by the backbone conformation, because steric hindrance of backbone atoms with the side chain may occur. Therefore it is more accurate to investigate conformation and flexibility depending on backbone conformation. The disadvantage of this approach is the sparse data, because the data has to be divided according to the backbone conformation.

4.2.5. Dependencies of χ angle conformations

In this section, the influences of the conformation of the χ angles within one side chain are investigated. After a short introduction to Bayes statistics, the dependent probabilities are investigated.

The Bayesian approach

To get a first hint wether two events are mutually independent of each other, the *Bayes theorem* [3] is used. For dependent events, the probability that event B occurs given event A has occured (written as $P(B \mid A)$ is given by

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)} \tag{4.3}$$

.

A solving of formula 4.3 for $P(A)$ results in

$$P(A) = \frac{P(A \mid B)P(B)}{P(B \mid A)} \tag{4.4}$$

<=>

$$P(B \mid A)P(A) = P(A \mid B)P(B) \tag{4.5}$$

.

Because

$$P(A \mid B)P(B) = P(A, B) \tag{4.6}$$

holds, the insertion of the right hand side from formula 4.6 in formula 4.5 results in

$$P(B \mid A)P(A) = P(A, B) \tag{4.7}$$

.

If two events are independent, $P(A, B) = P(A) * P(B)$, so that

[3] Thomas Bayes, 1702-1761

$$P(A \mid B) = P(A) \tag{4.8}$$

and

$$P(B \mid A) = P(B) \tag{4.9}$$

hold.

Using equation 4.8 and 4.9, a first hint on the independencies of two events can be gained. In table 4.7, the dependent probability for $P(r2 \mid r1)$ and the independent probability $P(r1, r2)$ for the nine χ_1, χ_2 combinations are shown. To test the independency of the two χ angles, $P(r2)$ is shown in column 6. If the two events are independent, $P(r2)$ and $P(r2 \mid r1)$ should be the same because for independent events the fact that event r1 occurred should have no effect for the occurance of r2.

Rot1	Rot2	# combinations	$P(r2 \mid r1)$	$P(r1, r2)$	P(r2)
1	1	1638	0.5991	0.1002	0.34
	2	917	0.3354	0.0561	0.4738
n=2734	3	179	0.0655	0.0109	0.1861
2	1	2174	0.4568	0.133	0.34
	2	2154	0.4526	0.1323	0.4738
n=4759	3	431	0.0868	0.0264	0.1861
3	1	1747	0.1973	0.107	0.34
n=8855	2	4675	0.528	0.286	0.4738
	3	2433	0.2748	0.1488	0.1861

Table 4.7.: Rotamer probabilities for χ_2 depending on χ_1

It can be seen that the dependent probability $p(r2 \mid r1)$ does not equal $p(r2)$. Therefore the considered χ angles are not independent of each other but the conformation of one χ angle is influenced by the previous χ angle. For the placement of side chains it is therefore more accurate to use the dependent probability if information about the previous angle is known.

Rot2	Rot3	# combinations	$P(r3 \mid r2)$	$P(r2, r3)$	$P(r3)$
1	1	277	0.4203	0.0521	0.2568
	2	295	0.4476	0.0546	0.497
n=659	3	87	0.132	0.0161	0.2462
2	1	1007	0.2759	0.1863	0.2568
	2	1796	0.4921	0.3322	0.497
n=3650	3	847	0.2321	0.1567	0.2462
3	1	104	0.0948	0.0192	0.2568
	2	596	0.519	0.1102	0.497
n=1097	3	397	0.3612	0.0734	0.2462

Table 4.8.: Rotamer probabilities for χ_3 depending on χ_2 angles

Rot 3	Rot2	# combinations	$p(r4 \mid r3)$	$p(r3, r4)$	$p(r4)$
1 n=501	1	159	0.3174	0.0565	0.1786
	2	288	0.5749	0.1022	0.628
	3	54	0.1078	0.0192	0.1935
2 n=1778	1	280	0.1574	0.0994	0.1786
	2	1159	0.6512	0.4114	0.628
	3	339	0.1907	0.1203	0.1935
3 n=538	1	64	0.119	0.0227	0.1786
	2	322	0.5985	0.1143	0.628
	3	152	0.2825	0.054	0.1935

Table 4.9.: Rotamer distribution for χ_4 depending on χ_3 angles

In tables 4.8 and 4.9, the dependent probabilities for higher χ angles are shown. The previous χ angle influences the rotamer probability for the next χ angle. For higher χ angle combinations, other rotamer combinations compared to the χ_1, χ_2 pair are preferred. The conformation of the higher χ angles is mainly influenced by neighbouring amino acids, whereas the conformation of the first χ angle is influenced by the backbone position.

Rot1	Rot3	# all	$P(r3 \mid r1)$	$P(r1, r3)$	$P(r3)$
1 n=465	1	138	0.2968	0.0255	0.2568
	2	191	0.4108	0.0353	0.497
	3	136	0.2925	0.0252	0.2462
2 n=1784	1	595	0.3335	0.1101	0.2568
	2	838	0.4697	0.155	0.497
	3	351	0.1967	0.0649	0.2462
3 n=3157	1	655	0.2075	0.1212	0.2568
	2	1658	0.5252	0.3067	0.497
	3	844	0.2673	0.1561	0.2462

Table 4.10.: Rotamer probabilities for χ_3 depending on χ_1 rotamers over all residues

To get a hint for cross dependencies for χ angles further apart, data for the χ_3, χ_1 pair is shown in table 4.10. It can be seen that the probability for χ_3 equals the dependent probability $P(r3 \mid r1)$. There are no influences between non adjacent χ angles.

Because the χ angle conformation is mainly influenced by the previous χ angle, the side chain conformation can be modelled as multiplication of the dependent probabilities for the single rotamer combinations (see next section). This model is used by Dunbrack and colleagues [11].

4.3. A rotamer library for the protein-protein docking problem

A rotamer library describes the possible rotamer combinations for all χ angles of the side chain and their associated probabilities from the test set. There are two approaches calculating a rotamer library: backbone independent and backbone dependent. For probability calculation

of the backbone independent case, the dependent probabilities for the rotamers of $\chi_{2,3,4}$ given r1 are taken. In the backbone dependent case, the $\chi_{2,3,4}$ conformations are described in dependency of χ_1 and the backbone angles ϕ and ψ. The backbone independent approach has the advantage that more data is available, because the rotamer data must not be divided according to backbone ranges, whereas the backbone dependent placement is more accurate, because the backbone atoms influence the possible rotamers of the side chain (see above).

For protein-protein docking, a rotamer library is used for side chain demangling tasks. If the side chain of a residue is clashing with other atoms, the side chain can be moved in its preferred direction (see chapter 5) or be placed according to the rotamer library. The side chain is placed in the rotamer combination with the highest probability without steric clashes. Because the probability for rotamer combinations depend on the environment of the residue (see table C.2), rotamer libraries for unbound data, complexes and different Secondary Structure elements are built.

4.3.1. Backbone independent

In the backbone independent approach, the independent probability for a χ_1, χ_2, χ_3, χ_4 combination is calculated approximated by the dependent probability. To calculate the probability for the whole rotamer combination, one has to deal with the problem of sparse or no data in some rotamer combinations. These unseen events will lead to a zero probability for the rotamer combination although in an infinite large test set, this (unfavourable) combination may be seen. So zero probabilities are not correct. To avoid unseen rotamers, Dunbrack et al. [11] use an approximation for the probability of the whole rotamer: because the conformation of one χ angle is mainly dependent on the previous χ angle (see above), they use

$$P(r234 \mid r1) \propto P(r2 \mid r1) \cdot P(r3 \mid r2) \cdot P(r4 \mid r3) \tag{4.10}$$

for the calculation of the probability. For the particular angle pairs, more data is available compared to the whole rotamer set. Less rotamer combinations with zero counts leading to a zero probability of the whole rotamer set are observed. To avoid zero counts Dunbrack et al. use the "adding one"method, where one count is added for each rotamer pair. Because all unseen rotamer pairs will get the same count, they will get the same probability independent of the probability for the previous χ angles. This may lead to an overestimation of the probability.

In this thesis, a statistical model from language processing is used for the calculation of the probabilities [13]. To calculate the probability of a sequence of symbols $P(w)$, it is approximated by the product of single probabilities

$$P(w) = P(w4 \mid w_3 w_2 w_1) \tag{4.11}$$

where $w_1 ... w_n$ is the history of w. For calculating the probability for side chain conformations, the values 1,2,3,4 for n where used (called mono-, di-, tri- and tetragram in language processing) according to the number of angles. Zero probabilities are avoided by discounting and backing off procedures. During linear discounting, the probability $c(yz)$ for each P(w) (each angle combination) is diminished by an amount $\beta(yz)$ proportional to its value. In the backing off procedure, this probability mass is redistributed for unseen events. The probability for the

unseen events is proportional to a background probability $\lambda(y)$ which is fixed and the probability $P(w_{n-1} \mid w_{n-2}...w_1)$, which is the dependent probability of the rotamer combination without the last angle. Zero probabilities are avoided without overestimating the probabilities for unseen events (rotamers not seen in the test set).

In contrast to Dunbrack [11], the independent probability for rotamer combinations are calculated which are approximated by dependent probabilities. In the approach of Dunbrack, the dependent probabilities are calculated by the multiplication of pairwise probabilities (see 4.10). In the new approach, influences of angles further apart are integrated as well, although they are quite small (see section 4.2.5). The rotamer probabilities are calculated for unbound proteins, complexes and unbound Secondary Structures (helix, sheet and random coils).

AS	r1	r2	r3	r4	$P(r1234)$ ALL	$P(r1234)$ HELIX	$P(r1234)$ SHEET	$P(r1234)$ RND
ARG	1	1	1	1	0.0008	0.0004	0.0017	0.0004
ARG	1	1	1	2	0.0042	0.0039	0.0083	0.0033
ARG	1	1	1	3	0.0011	0.0012	0.0019	0.0007
ARG	1	1	2	1	0.0057	0.0024	0.0056	0.0067
ARG	1	1	2	2	0.0097	0.0238	0.0167	0.0067
ARG	1	1	2	3	0.0048	0.0012	0.0043	0.0067
ARG	1	1	3	1	0.0004	0.0013	0.0001	0.0033
ARG	1	1	3	2	0.0089	0.0159	0.0028	0.0017
ARG	1	1	3	3	0.0034	0.0018	0.0028	0.0017
ARG	1	2	1	1	0.0040	0.0030	0.0056	0.0056
ARG	1	2	1	2	0.0101	0.0048	0.0083	0.0267
ARG	1	2	1	3	0.0030	0.0032	0.0051	0.0169
ARG	1	2	2	1	0.0091	0.0257	0.0278	0.0100
ARG	1	2	2	2	0.0256	0.0317	0.0139	0.0400
ARG	1	2	2	3	0.0137	0.0095	0.0140	0.0101
ARG	1	2	3	1	0.0015	0.0005	0.0015	0.0033
ARG	1	2	3	2	0.0175	0.0127	0.0226	0.0205
ARG	1	2	3	3	0.0067	0.0046	0.0056	0.0169
ARG	1	3	1	1	0.0004	0.0003	0.0007	0.0023
ARG	1	3	1	2	0.0023	0.0016	0.0007	0.0033
ARG	1	3	1	3	0.0006	0.0011	0.0003	0.0016
ARG	1	3	2	1	0.0055	0.0128	0.0056	0.0100
ARG	1	3	2	2	0.0109	0.0095	0.0194	0.0100
ARG	1	3	2	3	0.0063	0.0158	0.0060	0.0074
ARG	1	3	3	1	0.0013	0.0002	0.0003	0.0067
ARG	1	3	3	2	0.0076	0.0032	0.0085	0.0103
ARG	1	3	3	3	0.0025	0.0029	0.0056	0.0135
ARG	2	1	1	1	0.0048	0.0035	0.0056	0.0017
ARG	2	1	1	2	0.0133	0.0190	0.0167	0.0067
ARG	2	1	1	3	0.0110	0.0144	0.0032	0.0037
ARG	2	1	2	1	0.0095	0.0016	0.0024	0.0233
ARG	2	1	2	2	0.0301	0.0365	0.0306	0.0300
ARG	2	1	2	3	0.0186	0.0127	0.0168	0.0304
ARG	2	1	3	1	0.0015	0.0016	0.0005	0.0033
ARG	2	1	3	2	0.0108	0.0190	0.0197	0.0086
ARG	2	1	3	3	0.0070	0.0026	0.0056	0.0202
ARG	2	2	1	1	0.0196	0.0334	0.0167	0.0028
ARG	2	2	1	2	0.0232	0.0270	0.0250	0.0400
ARG	2	2	1	3	0.0124	0.0190	0.0204	0.0101
ARG	2	2	2	1	0.0276	0.0353	0.0222	0.0400
ARG	2	2	2	2	0.0629	0.0746	0.0194	0.0700
ARG	2	2	2	3	0.0462	0.0585	0.0617	0.0202
ARG	2	2	3	1	0.0059	0.0016	0.0067	0.0017
ARG	2	2	3	2	0.0470	0.0571	0.0649	0.0274
ARG	2	2	3	3	0.0272	0.0244	0.0084	0.0135
ARG	2	3	1	1	0.0010	0.0003	0.0028	0.0015
ARG	2	3	1	2	0.0059	0.0016	0.0056	0.0100
ARG	2	3	1	3	0.0029	0.0011	0.0045	0.0010
ARG	2	3	2	1	0.0194	0.0112	0.0194	0.0267
ARG	2	3	2	2	0.0449	0.0603	0.0694	0.0200
ARG	2	3	2	3	0.0377	0.0095	0.0179	0.0258
ARG	2	3	3	1	0.0034	0.0020	0.0006	0.0100
ARG	2	3	3	2	0.0272	0.0238	0.0141	0.0137
ARG	2	3	3	3	0.0188	0.0147	0.0168	0.0135
ARG	3	1	1	1	0.0006	0.0001	0.0011	0.0013
ARG	3	1	1	2	0.0027	0.0009	0.0028	0.0067
ARG	3	1	1	3	0.0008	0.0003	0.0006	0.0022
ARG	3	1	2	1	0.0019	0.0008	0.0004	0.0133
ARG	3	1	2	2	0.0086	0.0063	0.0083	0.0167
ARG	3	1	2	3	0.0023	0.0004	0.0013	0.0067

AS	r1	r2	r3	r4	$P(r1234)$ ALL	$P(r1234)$ HELIX	$P(r1234)$ SHEET	$P(r1234)$ RND
ARG	3	1	3	1	0.0006	0.0003	0.0001	0.0033
ARG	3	1	3	2	0.0029	0.0032	0.0028	0.0034
ARG	3	1	3	3	0.0010	0.0004	0.0056	0.0017
ARG	3	2	1	1	0.0055	0.0015	0.0083	0.0083
ARG	3	2	1	2	0.0249	0.0222	0.0167	0.0233
ARG	3	2	1	3	0.0042	0.0016	0.0025	0.0067
ARG	3	2	2	1	0.0185	0.0241	0.0194	0.0100
ARG	3	2	2	2	0.0483	0.0619	0.0722	0.0233
ARG	3	2	2	3	0.0202	0.0190	0.0112	0.0101
ARG	3	2	3	1	0.0027	0.0011	0.0030	0.0017
ARG	3	2	3	2	0.0377	0.0127	0.0226	0.0308
ARG	3	2	3	3	0.0156	0.0137	0.0168	0.0101
ARG	3	3	1	1	0.0027	0.0003	0.0020	0.0062
ARG	3	3	1	2	0.0046	0.0048	0.0020	0.0200
ARG	3	3	1	3	0.0006	0.0011	0.0008	0.0042
ARG	3	3	2	1	0.0097	0.0048	0.0083	0.0067
ARG	3	3	2	2	0.0310	0.0349	0.0139	0.0167
ARG	3	3	2	3	0.0120	0.0032	0.0209	0.0074
ARG	3	3	3	1	0.0034	0.0010	0.0012	0.0100
ARG	3	3	3	2	0.0179	0.0159	0.0226	0.0137
ARG	3	3	3	3	0.0120	0.0029	0.0393	0.0135

Table 4.11.: Probabilities for different ARG rotamers from new compiled unbound rotamer libraries

In table 4.11, the calculated rotamer probabilities for unbound ARG residues from diverse libraries are shown. Column 6 shows the over all probability, in column 7-9 the rotamer probabilities in Secondary Structures can be seen. The probabilities for rotamer combinations differ according to the rotamer library used due to the different steric constraints in Secondary Structure elements (see section 4.2.3). If information about Secondary Structure is available, the placement of the side chain can be done according to the library. The data for other unbound residues and the probability for complexes can be seen in the appendix (C.2, C.3).

4.3.2. Backbone dependent

Because the backbone conformation influences the probability for side chain rotamers, a backbone dependent rotamer library is compiled (see above). The dependent probabilities $P(\chi_1 \mid \psi)$ and $P(\chi_1 \mid \phi)$ are calculated with a language model described in section 4.3.1. For the calculation of the these dependent probabilities, more data is available compared to $p(\chi_1 \mid \psi, \phi)$, which is shown as well. The zero probabilities are due to rounding of the values to 6 digits. Because a language model is used, not all unseen combinations will get the same probability (see section 4.3.1).

AA	ϕ rotamer	ψ rotamer	r1	#	$P(\phi \mid \chi_1)$	$P(\psi \mid \chi_1)$	$P(\chi_1 \mid \phi\psi)$
ARG	1	17	1	2	0.0089	0.0097	0.0004
ARG	1	18	1	10	0.0089	0.0089	0.0024
ARG	2	1	1	1	0.0127	0.0006	0.0001
ARG	2	8	1	1	0.0127	0.0085	0.0001
ARG	2	10	1	4	0.0127	0.0066	0.0009
ARG	2	11	1	4	0.0127	0.0055	0.0009
ARG	2	15	1	1	0.0127	0.0009	0.0001
ARG	2	16	1	4	0.0127	0.0044	0.0009
ARG	2	17	1	8	0.0127	0.0097	0.0019
ARG	2	18	1	7	0.0127	0.0089	0.0016
ARG	3	8	1	1	0.0176	0.0085	0.0001
ARG	3	9	1	1	0.0176	0.0066	0.0001
ARG	3	10	1	11	0.0176	0.0066	0.0026
ARG	3	11	1	10	0.0176	0.0055	0.0023
ARG	3	16	1	1	0.0176	0.0044	0.0001
ARG	3	17	1	14	0.0176	0.0097	0.0034
ARG	3	18	1	7	0.0176	0.0089	0.0016
ARG	4	8	1	1	0.0013	0.0085	0.0001
ARG	4	9	1	1	0.0013	0.0066	0.0001
ARG	4	10	1	1	0.0013	0.0066	0.0001
ARG	4	17	1	1	0.0013	0.0097	0.0001
ARG	5	8	1	2	0.0028	0.0085	0.0004

AA	ϕ rotamer	ψ rotamer	r1	#	$P(\phi \mid \chi_1)$	$P(\psi \mid \chi_1)$	$P(\chi_1 \mid \phi\psi)$
ARG	5	9	1	1	0.0028	0.0066	0.0001
ARG	5	10	1	2	0.0028	0.0066	0.0004
ARG	5	11	1	1	0.0028	0.0055	0.0001
ARG	5	15	1	2	0.0028	0.0009	0.0003
ARG	6	8	1	14	0.0119	0.0085	0.0033
ARG	6	9	1	14	0.0119	0.0066	0.0033
ARG	6	16	1	3	0.0119	0.0044	0.0006
ARG	6	17	1	1	0.0119	0.0097	0.0001
ARG	7	7	1	1	0.0032	0.0002	0.0001
ARG	7	8	1	4	0.0032	0.0085	0.0009
ARG	7	16	1	4	0.0032	0.0044	0.0009
ARG	13	9	1	1	0.0002	0.0066	0.0001

Table 4.12.: r1=1 probabilities for ARG rotamers depending on the backbone conformation

Table 4.12 shows the probability for χ_1 depending on the backbone conformation. The backbone range is divided into 18 rotamers, rotamer 1 starting at $-180°$. In column 6 and 7, the dependent probability for χ_1 given one backbone angle can be seen, the dependent probability given both backbone angles can be seen in column 8. The probability of the first χ angle to be in the first rotamer differs depending on the ϕ, ψ conformation. The highest $P(\chi_1 \mid \phi\psi)$ can be seen for a ϕ range between -120 and -140 (Phi rotamer = 3) and a ψ range from 140 and 160 (rotamer 17) or a ϕ range between -60 and -80 (rotamer 6) in combination with a ψ range from -40 to 0. The disadvantage of the backbone dependent approach is the lack of data. Because of the division of the rotamer data into 18 rotamer ranges, there are many backbone regions with no data available. Because the language model uses a redistribution procedure to calculate the probability, a probability can be calculated without using the adding one method [11].

The knowledge about preferred conformations can be used for side chain demangling simular to ligand docking [28]. Side chain with steric clashes can be placed in more favourable conformations [3]. This is important if energy calculations are used for scoring the hypothesis, because this procedure does only make sense if all side chains are in a non-clashing, optimal position.

The rotamer libraries are evaluated in chapter 7. For evaluation the pruning of the search tree in using rotamer libraries compared to full search are calculated.

5. Flexibility

Rotamer libraries normally just give information on the side chain conformation, no flexibility of residues are taken into account. Mendes et al. try to integrate flexibility in their rotamer library by using not only three rigid rotamer, but a subset of rotamers clustering around this rigid conformation [32]. With the finer discretisation better results for the prediction of the conformation can be achieved, because more possibilities for the placement of the side chain are presented so that a more optimal solution can be found. But this model is still static. For protein-protein docking, side chains undergo conformational changes upon binding the second protein. The probabilities for these change are calculated in this chapter of the thesis.

Flexibility upon complex formation is investigated by comparing PDB structures from unbound proteins and complexes which show a nearly 100% sequence similarity in most residues of a chain. The test set of sequence identical proteins is composed of 39 PDB Entries of complexes with 46 chains, 55 Entries of unbound proteins with 55 chains containing 109309 corresponding amino acids. If there is more than one sequence identical unbound case for a complex or vice versa, all complex-unbound pairs are taken into account because different changing behaviour of one unbound protein in different complexes may be seen.

As measurement for flexibility, the percentage of amino acids changing their rotamer upon complex formation is used, therefore only the most flexible residues are taken into account. These residues have to cross an energy barrier to get in the next rotamer [22], so that these movements are forced by the environment. Angular differences of up to 20° in independently solved sequence identical structures are quite common and are due to slightly different methods for crystallisation. In the approach of Olson et al. [48], the 10% most flexible residues are omitted from the flexibility calculation, therefore most of the residues changing their rotamer are not taken into account. They consider a side chain as flexible if the χ angles change more than 40°. For changing to the next rotamer, an angular change of more than 40° is needed for most of the residues to get to the next rotamer. Because Olson et al. skip the most flexible residues and take only residues showing an angular movement larger than 40° into account, their flexibility is biased towards medium flexibility.

In section 5.1, the flexibility for individual χ angles is investigated, followed by the flexibility according to the rotamer combination in section 5.2. The direction of side chain movements can be seen in section 5.3. In section 5.4, the probability for concerted movements of more than one χ angle is investigated, followed by the calculation of individual χ angles without sequence identical chains 5.5.

5.1. Flexibility of individual χ angles

In this section, the flexibilities of individual χ angles are investigated as first hint for flexibility. Because the flexibilities differ according to the environment of the side chains, the flexibility was calculated over the whole test set 5.1.1, for exposed residues 5.1.2, for interface residues 5.1.3 and for residues in Secondary Structures 5.1.4.

5.1.1. Flexibility of the whole test set

In this section, the flexibility of all side chains in the test set is investigated without taking into account Secondary Structure and solvent accessible surface (SAS) area. An overview of the flexibility of different χ angles in different side chains can be seen in the flexibility scales, where the residues are ranked according to their probability for a rotamer change. Side chains in the same group show a probability difference less than 0.05, if empty bins (marked by >>) occur, the difference to the next group is larger than 0.05.

flexibility scale χ_1: **ARG >> SER, LYS, GLN >> VAL > ASP, THR, ASN, GLU > ILE, MET, HIS, PHE, TYR, TRP, CYS**

flexibility scale χ_2: **ASP, GLN, ASN, ARG >> LEU, GLU, LYS > ILE, PHE, HIS, MET, TRP, TYR**

flexiblity scale χ_3: **ARG» GLN » LYS> GLU » MET**

flexibility scale χ_4 **ARG >> LYS**

The ranking of the flexibility differs slightly in various χ angles, although some flexibility trends are constant in all angles. E.g. flexible amino acids like ARG can be found in the first flexibility group, inflexible side chains like the ones with ring systems in their side chain are inflexible in all angles because of the steric hindrance of this bulky system.

AA	P(change χ_1)	P(change χ_2)	P(change χ_3)	P(change χ_4)
ARG	0.3001	0.2449	0.5008	0.4984
ASN	0.0859	0.2493	–	–
ASP	0.1009	0.2707	–	–
CYS	0.0011	–	–	–
GLN	0.2115	0.2654	0.4200	–
GLU	0.0728	0.1027	0.2370	–
HIS	0.0060	0.0080	–	–
ILE	0.0505	0.0884	–	–
LEU	0.0128	0.1647	–	–
LYS	0.2197	0.1026	0.2694	0.3217
MET	0.0125	0.0050	0.0417	–
PHE	0.0058	0.0346	–	–
SER	0.2349	–	–	–
THR	0.0870	–	–	–
TRP	0.0011	0.0166	–	–
TYR	0.0017	0.0013	–	–
VAL	0.1204	–	–	–

Table 5.1.: Probabilities for a rotamer change of different χ angles

Table 5.1 shows the probabilities for rotamer changes in different χ angles upon complex formation. The probabilities of the most flexible residues for each angle are marked in yellow. The probability of a rotamer change differs according to the amino acid and the χ angle. High

percentages for rotamer changes can be seen for χ_3 of ARG residues (50% of the representatives changing their rotamer), χ_3 of GLN residues (42%), χ_4 of LYS side chains and χ_1 of ARG residues. For inflexible residues like HIS, MET, PHE, TYR and TRP, the flexibility is low for all χ angles (see above). PHE, HIS, TYR and TRP do not show high flexibility because of a ring system in their side chain, which leads to steric clashes if rotated. For most residues, the flexibility of χ_2 is higher compared to χ_1. This is due to the larger distance to the backbone atoms which prevent rotation by steric hindrance for the first χ angle. For ARG and LYS residues, the flexibility of the χ_2 is lower compared to χ_1. These are long side chains with four rotatable bonds, so that a rotation in χ_2 is not needed necessarily for a rotation of the side chain. The flexibility of LYS residues is reduced compared to ARG. ARG has a more complicated $H_2 - N - C = NH_2^+$ group at the end of the side chain which is more voluminous compared to the NH_3^+ group of LYS, so that steric clashes which lead to a conformational change occur more often in the side chains of ARG residues.

For side chains with branches the two ends of the branch influence the conformation and amount of flexibility. For apolar branched residues (VAL, LEU and ILE), different flexibilities can be seen in different angles: for VAL and LEU which are symmetrically branched, the highest probability for a rotamer change can be seen for the angle in front of the branch (χ_1 for VAL and χ_2 for LEU residues). A rotation of this angle leads to different positions of the two ends. ILE is asymmetrical branched with a methyl group bond to χ_1. The flexibility of ILE is reduced in χ_2 compared to LEU. A rotation in χ_2 for ILE moves a rotationally symmetric methyl group which does not lead to another conformation. Branched polar (GLN, ASN) or charged amino acids (GLU, ASP) show different amounts of flexibility as well. A reorientation of a charge is unfavourable, because a charge has to be neutralized in the inner part of a protein. Unneutralized charges inside the protein lead to destabilisation of the structure. For branched, charged amino acid (GLU, ASP), the charge at the end of the residue can not be localised exactly. If the charged atom is not involved in a bond, the free charge is localised somewhere in between the *C-OH group* and the *C=O group* of the side chain. Therefore a rotation of the χ angle in front of the branch (χ_2 for ASP and $\chi 3$ for GLU) will not lead to another position of the charge, whereas a rotation around the angle one *C*-atom earlier does change the position of the charge. This can be seen in the χ_2 of GLN and GLU, where the charged GLU side chain has a lower flexibility compared to the polar, but uncharged GLN residue. For ASN and ASP this effect can not be seen because the angle one rotatable bond earlier is the χ_1 angle, which rotation is restricted by the backbone atoms. Therefore the charged ASP residue has the some low amount of flexibility compared to the polar ASN side chain.

Another factor which influences the flexibility of side chains is its length. The longer a side chain is, the higher is the flexibility. For longer side chains, the probability for steric clashes is higher compared to smaller side chains, so that longer side chains tend to move away more often. One example can be seen in the χ_3 of GLN and χ_2 angle of ASN, which have the same structure, but a different length. The probability for a χ_3 movement of GLN is nearly two times higher compared to the probability for a rotation in χ_2 for ASN residues (0.42 compared to 0.24). The same effect can be seen for LYS side chains where χ_4 has the highest probability for a rotamer change. More rotatable freedom is allowed in the outer parts of the side chain. CYS and SER have small side chains. Although it can be assumed that smaller side chains are quite inflexible, SER shows a high probability for a rotamer change of 0.23. The *OH group* of SER residues can form hydrogen bonds. For these bonds, the atoms have to be placed exactly. A rotation around χ_1 can bring the OH group in a favourable position for building bonds. In

contrast to SER, CYS has a very low probability for rotamer changes compared to other χ_1 residues. CYS residues are mainly found in the core of the protein. The sulfur groups of two Cysteins can form covalent bonds (sulphur bridges) which stabilise the protein structure. Due to this bonds, CYS residues are quite inflexible.

5.1.2. Flexibilities of exposed residues

The data in table 5.1 is taken from the whole test set including exposed residues. Therefore some flexibility which would not or rarely take place in the core of a protein can be observed (e.g. rotation of charges). The data for exposed residues are shown in table 5.2 as comparison. For a classification of residues in buried and exposed, the solvent accessible surface (SAS) area is used. The SAS area is the part of the residue which gets in contact with a water molecule (diameter 1.5 Å) rolled along the protein surface without penetrating other atoms of the structure [29]. In the approach of Janin et al. [22], the surface area was calculated with the algorithm of Levitt et al.. A residue is defined as exposed if it has an accessible area more than 60 $Å^2$. Buried residues have an SAS area less than 20 $Å^2$ and intermediate residues have an SAS area in between. These thresholds are used for all residues, although large side chains tend to lie above the defined threshold for exposed residues, even if the larger part of the side chain does not get in contact with the water molecule. E.g. some ARG side chains have an accessible surface area of 60$Å^2$ although they are nearly buried. Therefore a threshold is difficult to determine and can not be taken for all residues. In our approach [25], we calculate the SAS area with the BALL software [26] which uses the algorithm from Eisenhaber et. al. [12], a modification of the algorithm used by Janin and Wodak. Instead of defining one thresholds for all residues we use the relative SAS area for each residue similar to [18],[37]. The SAS area of each amino acid is divided by the SAS area of the most exposed representative of this group. Because the threshold is relative, it can be used for all residues. A relative SAS area larger than 0.4 means that more than 40% of the residue gets in contact with the solvent.

AA	P(change χ_1)	P(change χ_2)	P(change χ_3)	P(change χ_4)
ARG	0.5491	0.2321	0.7589	0.5938
ASN	0.5000	0.6792	–	–
ASP	0.7333	0.6667	–	–
GLN	0.5755	0.8302	0.7170	–
GLU	0.5238	0.9048	0.5397	–
LYS	0.3053	0.1737	0.5421	0.5368
SER	0.5616	–	–	–
THR	0.2640	–	–	–
TRP	0.0000	0.0000	–	–
TYR	0.0000	0.0114	–	–
VAL	0.0000	–	–	–

Table 5.2.: Probabilities for a rotamer change of χ rotamers with a relative SAS area $>$ than 0.4

In table 5.2 the flexibility of amino acids with a relative SAS area > 0.4 is shown. Some residues are omitted because only a few representatives are found at the surface, e.g. MET, ILE and LEU. For some other residues like CYS, HIS and PHE, no exposed representatives can be found in the test set.

The flexibility for the χ_1 angle is higher compared to the over all flexibility in table 5.1 except for TRP, TYR and VAL residues. This is due to less hindrance by neighbouring amino acids at the surface. The sterical hindrance by backbone atoms is reduced as well, because the backbone

position at the surface of a protein is more stretched compared to the core. The ring systems of TRP and TYR do not show a higher flexibility if they are at the surface of the protein. The bulky ring system prevents rotation although more than 40% of the side chain area is exposed to the solvent. Exposed VAL residues show no flexibility at all. Because the exact position of the the two *methyl groups* can not be seen in the electron density map during crystallisation (personal communication with crystallographers) and the two ends of the branch are equal, they are placed in the most favourable rotamer. This is done for all structures, so that no differences between unbound and complexed structures can be found. It can be noticed that for residues with a high χ_1 flexibility over all data (ARG, SER, LYS) the increase of probability for exposed residues is smaller compared to the residues which have a smaller probability for a rotamer change in the whole test set (see table 5.1 compared to table 5.2). E.g. the flexibility for LYS residues increases at 8% for exposed residues compared to the whole test set, whereas this increase for ASP residues is 63%. The constraints which oppose rotation in buried branched amino acids cease to apply for exposed representatives, so that the flexibility is higher especially for exposed branched amino acids.

For χ_2 of exposed residues the flexibility increases for all amino acids except for ARG where the probability stays on the same level for the whole test set and for exposed residues. The highest χ_2 flexibility with 90% of the residues moving to a new rotamer can be seen for GLU residues. GLU is a branched and charged amino acid, which can rotate more freely at the surface of the protein (see above). The charge which opposes rotation in the core of the protein can be compensated by the solvent for exposed residues.

The higher flexibility of exposed residues can also be seen for all χ_3 and χ_4 angles. Because these angles show higher probabilities for a rotamer change for the whole test set, the increase of flexibility is lower compared to the angles whose flexibility is more restricted in buried amino acids. For these angles, more than 50% of the exposed representatives change their rotamer.

5.1.3. Flexibility of interface residues

The flexibility of interface residues is investigated separately to exposed residues because the flexibility differs from exposed non-interface residues. The information about interface residues are gained on complexes. Residues are characterised as belonging to the interface (active site) of a protein if the atoms of one complex part have a distance less than 4 Å to the second protein in the complex. With this definition, not only residues which form bonds with the partner in the complex, but also amino acids with are near the partner and may be influenced sterically are taken into account. The flexibility for interface residues can be seen in table 5.3.

In table 5.3, the probability of a rotamer change is calculated with the data of interface residues. Two groups can be distinguished: the first group shows an increase of flexibility in χ_1 (ASN, GLN, MET and THR), whereas the flexibility is reduced for the other residues. For χ_2 the same tendency can be seen with ASN, GLN, GLU, ILE and LEU residues having a high probability for a rotamer change, whereas the other residues are less flexible compared to the whole test set. For χ_3 and χ_4, the flexibility is higher for all residues.

The flexibility is not higher for all active site residues. For some residues, even no flexibility for χ_1, χ_2 can be observed. The restriction of rotation by the backbone may be very pronounced for these residues because an active site is a crevice in most of the cases. In contrast to χ_1,

AA	n activesite	P(change χ_1)	P(change χ_2)	P(change χ_3)	P(change χ_4)
ARG	8	0	1	1	1
ASN	612	0.3333	0.52288	–	–
ASP	646	0	0	–	–
CYS	1410	0	–	–	–
GLN	1866	0.6313	0.831	0.5413	–
GLU	16	0	1	1	–
HIS	3132	0	0	–	–
ILE	46	0.04348	0.9565	–	–
LEU	770	0	0.0338	–	–
MET	82	1	1	1	–
PHE	1508	0	0	–	–
SER	4082	0.0122	–	–	–
THR	100	0.42	–	–	–
TRP	1644	0.0024	0	–	–
TYR	1888	0	0	–	–
VAL	84	0	–	–	–

Table 5.3.: Probabilities for a rotamer change of χ angles at the activesite

a very flexible χ_2 can be found at interfaces for some residues (ARG, GLU, MET). For these residues, all interface representatives change their χ_2 rotamer. There are two tendencies at the interface: on the one hand, for some rather flexible residues, the conformational changes are restricted, whereas for other residues with medium or less flexibility in the whole test set (MET), the probability for conformational changes increases. If all residues would be flexible, the specificity of catalytic reactions could not be guaranteed, so just a few distinct side chains move away to let the second protein pass. Especially bulky amino acids at the active site may be involved in blocking tasks. In the active site of the three serine proteases chymotrypsin, trypsin and elastase which show a sequence similarity of 40% (60% in the interior of the protein), the specificity is achieved by a few distinct amino acids [45]. All three serine proteases are in the digestive system of mammals and cleave polypeptide chains. Chymotrypsin cleaves after aromatic or bulky unpolar residues (TYR, TRP, MET) whereas Trypsin cleaves polypeptides after positively charged ARG or LYS residues. Elastase needs small, uncharged side chains for cleavage. At the active site of chymotrypsin, the side chain of the polypeptide after which the cleavage takes place lies in a non-polar pocket. In the active site of trypsin, a SER of the pocket is replaced by a negatively charged ASP that forms an electrostatic bond with LYS and ARG residues of the cleaved peptide, so that the specificity changed towards positively charged residues. In the case of Elastase, two small GLY residues are exchanged by much bulkier VAL and THR residues, so that only small residues can get in the binding pocket and being cleaved.

Summarising the flexibility for single χ angles, it can be seen that the flexibility of a side chain is influenced by its structure and polarity. Longer, polar side chain tend to be more flexible than shorter or apolar ones. Branches hinder rotation, because branched side chains have a larger volume compared to long, stretched side chains, so that the probability for steric clashes with neighbour residues is higher. Because of the size of their side chain, residues with a ring systems are very inflexible. Polar groups like OH or NH_2 lead to an increase in the flexibility, because these groups can form hydrogen bonds for which the functional group has to be positioned carefully. Charges at the end of branched side chains (GLU, ASP) reduce the flexibility one angle before the branched $C - atom$. A rotation in this angle leads to a movement of the charge (see above). This effect can be seen for GLU residues. Residues at the surface are more flexible. The flexibility scale for exposed residues differs from the scale over all residues, because the steric constraints which oppose rotation in many cases cease to be active. For interface residues, not all residues show a high amount of flexibility because the specificity of the active site must be warranted.

5.1.4. Flexibility of Secondary Structure elements

In this section of the thesis, the flexibility is investigated for different Secondary Structure elements. Because Secondary Structure influences the conformation of the side chains, it is assumed that the amount of allowed flexibility is also influenced. The assignment for the Secondary Structure elements is done according to the DSSP algorithm [23]. The probability for a rotamer change is calculated for all helix types(α helices with end residues, $3_{10} - helices$ and $Pi - helices$), $\alpha - helices$ without two residues at the end (pure α helices), $\beta - sheets$ and random coils. The probability for pure $alpha - helices$ are shown for comparison because of the stricter steric constraints described by Sternberg et al. [31]. The flexibility of residues which are not assigned as *helix* or *sheet* residues ($non - helix - non - sheet$) are shown as comparison in table 5.5.

AA	χ_1		χ_2		χ_3		χ_4	
	Helix	pure α-Helix	Helix	pure α-Helix	Helix	pure α-Helix	Helix	pure α-Helix
ARG	0.4226	0.2902	0.1032	0.0121	0.5757	0.4000	0.6073	0.6909
ASN	0.0012	0.0000	0.2039	0.1856	–	–	–	–
ASP	0.7288	0.1250	0.5593	0.3125	–	–	–	–
CYS	0.0000	0.0000	–	–	–	–	–	–
GLU	0.0307	0.0062	0.0521	0.0186	0.2485	0.4161	–	–
HIS	0.0038	omitted	0.0076	omitted	–	–	–	–
ILE	0.2822	0.0423	0.2810	0.0448	–	–	–	–
LEU	0.0247	0.0567	0.1140	0.0567	–	–	–	–
LYS	0.1069	0.1137	0.0423	0.0364	0.1939	0.2085	0.3572	0.3061
MET	0.0000	0.0000	0.0035	0.0000	0.0140	0.0131	–	–
PHE	0.0265	0.0058	0.0212	0.0116	–	–	–	–
SER	0.2918	0.1040	–	–	–	–	–	–
THR	0.0189	0.0000	–	–	–	–	–	–
TRP	0.0044	0.0000	0.0044	0.0019	–	–	–	–
TYR	0.0000	0.0000	0.0000	0.0000	–	–	–	–
VAL	0.1041	0.0000	–	–	–	–	–	–

Table 5.4.: Probabilities for rotamer changes in helices

AA	flexibility in non-Helix-non-Sheet residues			
	χ_1	χ_2	χ_3	χ_4
ARG	0.3182	0.3474	0.6224	0.5674
ASN	0.0863	0.2622	–	–
ASP	0.0588	0.2440	–	–
CYS	0.0038	–	–	–
GLN	0.3141	0.3187	0.6382	–
GLU	0.0657	0.0985	0.2288	–
HIS	0.0116	0.0162	–	–
ILE	0.0492	0.0849	–	–
LEU	0.0187	0.2093	–	–
LYS	0.3311	0.1400	0.3964	0.5005
MET	0.0000	0.1905	0.0000	–
PHE	0.0130	0.1531	–	–
SER	0.3029	–	–	–
THR	0.0831	–	–	–
TRP	0.0000	0.0200	–	–
TYR	0.0035	0.0022	–	–
VAL	0.1384	–	–	–

Table 5.5.: Probabilities for rotamer changes in non-helix-non-sheet residues

The flexibility of residues in all helix types is higher compared to pure $\alpha - helices$ for many cases (cf. table 5.4). This is due to the different constraints in other helices: the backbone is distorted for some of these helix types, more conformations and therefore more conformational changes are allowed because steric clashes are moderated. This holds especially for end residues which have more freedom to move because residues next to them are not in the restricted backbone range of $\alpha - helices$.

Due to the steric constraints for residues in helices, the flexibility of the first χ angle is reduced for many cases compared to $non - helix - non - sheet$ residues. The probability for steric

clashes is higher in helices because of the tight packing of residues in helices. Polar or charged residues can form hydrogen bonds with the backbone, which stabilises helices and which opposes rotation. For pure $\alpha - helices$, the flexibility is reduced or stays on the same level except for ASP. The high flexibility for ASP side chains may be misleading because of the sparse data. ASP representatives are rarely found in the core of $\alpha - helices$, so that only 16 examples from 5 unbound structures and 4 complexes are in the test set. Two residues change their rotamer upon complex formation: ASP 53 from the chain A of the PDB structure *1bel* and ASP 337 from the structure *1cip* (chain A). For residues in all helix types, an increase can be seen for ARG and ILE. Most of the rotating amino acids of this types can be found at the ends of $\alpha - helices$ or in other helix types with a more distorted backbone, because the flexibility is lower for core residues in $\alpha - helices$. Residues like CYS, HIS, MET, THR, TYR which are inflexible in the whole test set do not show a high probability in helices and $non - helix - non - sheet$ as well. A high flexibility with 30% rotamer changers in $\alpha - helical$ cores can be seen for ARG residues. ARG residues have long side chains, so that a rotation in χ_1 may be needed to avoid some unfavourable interactions for the rest of the side chains.

GLN residues in pure $\alpha - helices$

The flexibility for GLN residues is not shown in table 5.4 because of the high flexibility which may be misleading. An increase of flexibility from 0.31 to 0.42 can be seen for the data of pure $\alpha - helices$. The test set of $\alpha - helices$ consists of 22 unbound proteins which can form 24 complexes, but 14 of this pairs are different trypsin unbound-complexed pairs. During docking, the GLN with the Residue ID 240 of trypsin changes its χ_1 rotamer from the third to the second rotamer in 104 out off the 114 cases. 10 residues move from the second to the third rotamer. The residue is exposed in the protein, so that a rotation is possible although it is in an $\alpha - helix$. It can be seen that flexibility of residues between unbound and complexed structures is quite conservative: most of the residues show the same behaviour during complex formation in different independently solved unbound-complex pairs. By including different unbound-complex pairs for the same protein, the conservative behaviour of the residues can be seen from the test set.

AA	χ_1		χ_2		χ_3		χ_4	
	Sheet	nonHelixnonSheet	Sheet	nonHelixnonSheet	Sheet	nonHelixnonSheet	Sheet	nonHelixnonSheet
ARG	0.1612	0.3182	0.2620	0.3474	0.2544	0.6224	0.2443	0.5674
ASN	0.2340	0.0863	0.3333	0.2622	–	–	–	–
ASP	0.0323	0.0588	0.5419	0.2440	–	–	–	–
CYS	NA	NA	–	–	–	–	–	–
GLN	0.0368	0.3141	0.0521	0.3187	0.0844	0.6382	–	–
HIS	0.0000	0.0116	0.0037	0.0162	–	–	–	–
ILE	0.0024	0.0492	0.0353	0.0849	–	–	–	–
LEU	0.0106	0.0187	0.1454	0.2093	–	–	–	–
LYS	0.2242	0.3311	0.0936	0.1400	0.2182	0.3964	0.1658	0.5005
MET	0.0000	0.000	0.0000	0.1905	0.0351	0.0000	–	–
PHE	0.0000	0.0130	0.0000	0.1531	–	–	–	–
SER	0.0421	0.3029	–	–	–	–	–	–
THR	0.1309	–	–	–	–	–	–	–
TRP	NA	0.0000	0.0000	0.0200	–	–	–	–
TYR	NA	0.0035	0.0000	0.0022	–	–	–	–
VAL	0.1045	0.1384	–	–	–	–	–	–

Table 5.6.: Probabilities for changes of χ rotamers in β-sheets

In $\beta - sheets$ (see table 5.6), the steric constraints for the residues differ. In *sheets*, more than one strand may be ordered in a parallel fashion or the ends of the strands may be twisted around each other. For parallel strands, the side chains above or underneath the plane may

rotate more freely if the sheet is positioned at the surface of the protein whereas the flexibility of the residues involved in hydrogen bonds stabilising the two strands is reduced.

For ARG, ASN, GLN and LYS residues, the χ_1 flexibilities are reduced in sheets compared to $non-helix-non-sheet$ residues, whereas this probability is raised for ASN residues. The changing probability for the second χ is lower for most of the residues as well compared to $non-helix-non-sheet$ structures except for ASN which has a higher probability of a rotamer change in sheets (0.33 compared to 0.26). The χ_2 flexibility is higher for ASN residues, although for other residues, the flexibilities in higher χ angles are reduced. ASN residues are branched, so this flexibility is due to the placement of the polar group for building bonds. For $\beta-sheets$, no clear tendencies can be seen for the flexibility because of the variety of steric constraints, whereas for helices, a reduction of flexibility can be noticed, which mainly occurs in the core residues of $\alpha-helices$.

GLU residues in $\beta-sheets$

GLU residues show a higher probability for a rotamer change in $\beta-sheets$ compared to $non-helix-non-sheet$ residues (0.07 compared to 0.32). There are 7 out off 23 residues which change their rotamer. These residues are in 4 unbound structures with two of them having two chains and 3 complexes with one complex part consisting of two chains. For the *1ckp-1fin* pair, the GLU with the Residue Id 12 in the chain A and C changes its χ_1 upon complex formation. For the pdb structure *1dm2*, this residue is in the first rotamer, for *1ckp* the GLU 12 is in the third rotamer. Upon complex formation both move to the second rotamer. A visual inspection using VMD this GLU does neither belong to the sheet nor to the following turn so that the sterical constraints are moderated. For this residue, the DSSP algorithm used for classification and the Secondary Structure assignment of VMD [19] have different measurements for sheets. Because this residue is exposed and not at the interface of the protein, rotation of the side chain is possible.

5.1.5. Flexibility depending on the backbone conformation

In this chapter, the influence of the backbone on the flexibility of the first χ angle is investigated. Because the probability for a special χ angle rotamer is influenced by the the backbone rotamers (see table 4.6 in section 4.2.4), it is assumed that the different backbone positions have an influence on the conformational flexibility as well.

Figure 5.1 shows the probability for a rotamer change in χ_1 depending on the backbone conformation. The backbone angles are discretised in 18 rotamers with a range of 20°, the first starting at $-180°$. The different diameters of the circles are correlated with the probability for a rotamer change. It can be seen that the probability for a rotamer change is not only depending on the χ rotamer, but also depending on the backbone conformation. There are some backbone conformations for which a change shows a higher probability in all three χ_1 rotamer compared to the over all value of the side chain flexibility without taking into account the backbone conformation, whereas in other regions, hardly any change can be observed. To predict the flexibility for a rotamer change, it is more accurate to take the backbone conformation into account.

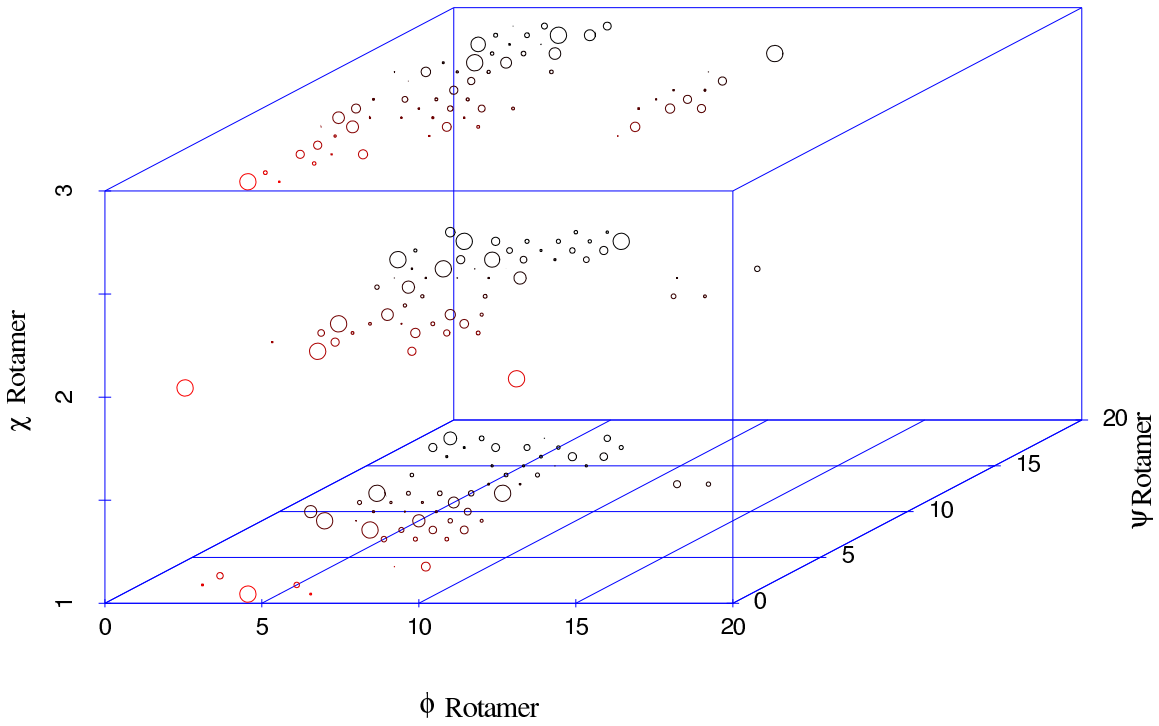


Figure 5.1.: Probability of a χ_1 rotamer change depending on the backbone rotamers

AA	ϕ rotamer	ψ rotamer	r1 Unbound	r1 Complex	# Unbound	# change	P(change)
ARG	1	18	1	2	6	2	0.3333
ARG	2	10	1	3	24	14	0.5833
ARG	2	11	1	3	24	14	0.5833
ARG	2	15	1	3	1	1	1.0000
ARG	2	16	3	2	4	2	0.5000
ARG	2	17	3	2	9	8	0.8889
ARG	2	18	3	1	6	1	0.1667
ARG	3	9	3	1	152	8	0.0526
ARG	3	10	1	3	48	28	0.5833
ARG	3	11	1	3	36	21	0.5833
ARG	3	15	2	1	1	1	1.0000
ARG	3	16	2	1	4	1	0.2500
ARG	3	17	1	3	4	1	0.2500
ARG	4	8	2	1	1	1	1.0000
ARG	4	8	3	1	18	1	0.0556
ARG	4	9	3	1	64	1	0.0156
ARG	4	14	2	3	2	2	1.0000
ARG	4	14	3	2	10	6	0.6000
ARG	4	15	2	3	10	4	0.4000
ARG	4	15	3	2	256	2	0.0078
ARG	4	16	2	3	2	2	1.0000
ARG	4	17	2	3	1	1	1.0000
ARG	4	18	3	1	60	2	0.0333
ARG	5	7	2	3	1	1	1.0000
ARG	5	8	2	3	2	2	1.0000
ARG	5	8	3	2	55	11	0.2000
ARG	5	9	3	1	67	16	0.2388
ARG	5	13	2	3	10	4	0.4000
ARG	5	14	2	3	42	18	0.4286
ARG	5	15	2	3	23	8	0.3478
ARG	5	15	3	2	12	6	0.5000
ARG	5	16	2	3	78	36	0.4615
ARG	5	17	2	3	25	12	0.4800
ARG	6	6	2	1	14	9	0.6429
ARG	6	7	2	1	115	11	0.0957
ARG	6	7	3	2	10	2	0.2000
ARG	6	8	1	3	29	26	0.8966
ARG	6	8	2	3	187	49	0.2620
ARG	6	8	3	2	52	21	0.4038
ARG	6	9	1	3	6	1	0.1667

AA	ϕ rotamer	ψ rotamer	r1 Unbound	r1 Complex	# Unbound	# change	P(change)
ARG	6	9	2	3	2	2	1.0000
ARG	6	9	3	2	3	3	1.0000
ARG	6	11	2	3	2	1	0.5000
ARG	6	15	3	2	30	18	0.6000
ARG	6	16	1	3	14	13	0.9286
ARG	6	16	2	3	109	74	0.6789
ARG	6	16	3	1	54	21	0.3889
ARG	6	17	1	3	14	13	0.9286
ARG	6	17	2	3	6	3	0.5000
ARG	6	17	3	2	5	2	0.4000
ARG	7	7	2	1	186	104	0.5591
ARG	7	7	3	2	5	2	0.4000
ARG	7	8	1	3	14	13	0.9286
ARG	7	8	2	3	118	85	0.7203
ARG	7	8	3	2	6	2	0.3333
ARG	7	16	1	3	28	26	0.9286
ARG	8	14	3	2	2	1	0.5000
ARG	12	2	2	3	2	1	0.5000
ARG	12	11	2	3	14	14	1.0000
ARG	12	11	3	1	117	4	0.0342
ARG	13	11	3	2	32	2	0.0625

Table 5.7.: Probabilities for rotamer changes of the different χ_1 rotamers depending on the backbone, here shown for ARG residues

In table 5.7 probabilities for conformational changes in χ_1 are given for ARG residues. The probabilities vary according to the backbone conformation. For example the probability for a change from r1=2 to r1=3 calculated over all data is 0.33, whereas the probabilities depending on the backbone rotamers vary from 0.17 to 1.0 (the probabilities for this change are highlighted in yellow). The backbone rotamers do not only influence the side chain conformation, but also the flexibility shown upon complex formation. A problem of the backbone dependent flexibility prediction is the lack of data: there are no or only few representatives for some backbone rotamers. For rotamers with no data points, no flexibility can be calculated. To overcome this problem, backbone rotamers which are large enough to fit data in the sparse regions of the ramachandran plot as well can be taken.

5.1.6. Flexibility depending on Rotamericity

The rotamericity of an angle is its distance to the median of the rotamer [44]. For the investigation, the rotamericities of the unbound angles are taken and divided into three classes: rotamericities larger than 20°, larger than 25° and larger than 30°.

	over all		rotamericity20		rotamericity25		rotamericity30	
AS	#	P(change)	#	P(change)	#	P(change)	#	P(change)
ARG	2536	0.30	305	0.46	202	0.61	158	0.55
ASN	7897	0.09	520	0.21	271	0.29	123	0.55
ASP	3616	0.10	157	0.35	67	0.55	56	0.54
CYS	6321	0.00	697	0.00	463	0.00	210	0.00
HIS	1496	0.01	378	0.01	6	0.17	2	0.50
GLN	4450	0.21	516	0.57	355	0.47	241	0.42
GLU	1840	0.07	120	0.33	64	0.30	38	0.32
ILE	6910	0.05	127	0.07	72	0.08	29	0.07
LEU	7249	0.01	600	0.04	273	0.08	219	0.10
LYS	4551	0.22	561	0.46	270	0.41	169	0.46
MET	1199	0.01	325	0.05	213	0.00	71	0.01
PHE	1881	0.01	55	0.15	21	0.29	4	1.00
SER	13910	0.23	1942	0.45	1188	0.56	807	0.68
THR	5205	0.09	155	0.35	107	0.50	91	0.57
TRP	2717	0.00	120	0.00	14	0.00	0	0.00
TYR	4754	0.00	79	0.03	7	0.00	1	0.00
VAL	8325	0.12	229	0.54	149	0.79	134	0.84

Table 5.8.: Probabilities of rotamer changes according to rotamericity

In table 5.8, the probabilities for rotamer changes of the residues depending on the rotamericity can be seen. The counts and probabilities for rotamer changes are shown for the whole test set (second and third column), for a rotamericity more than 20° (fourth and fifth column), rotamericity above or equal 25° (sixth and seventh column) and a rotamericity exceeding 30° (eight and ninth column).

Two trends can be seen: for flexible residues the probability for a rotamer change increases especially for rotamericities smaller than 30°. Residues with higher rotamericities are in a higher energy niveau because of their larger distance to the median of the rotamer which marks the energy minimum of the rotamer. They need less energy to cross the energy barrier to the next rotamer, so that the probability to do so is higher. For inflexible residues like the ones with a ring system in their side chain, the probability for changing in the next rotamer is low no matter what the rotamericity is. For a rotamericity larger than 30°, an increase can not be seen for all cases. The energy niveau of proteins with a rotamericity of 20° is high enough to cross the border to the next rotamer, so that a further increase of the energy niveau (a higher rotamericity) does not necessarily lead to higher probability of a rotamer change. Only residues with a medium flexibility (like ASN, THR, VAL) show a further increase of their flexibility for a rotamericity of 30°.

PHE residues have a probability of a rotamer change of 1.0 for a rotamericity of 30°. This high probability is misleading because there are only 4 representatives in this rotamericity class. They belong to the PDB structure *1ckp* (Residue Ids 146 and 248, chains A and C). A movement from the second to the third rotamer can be observed for both residues upon complex formation (*1fin*). The high probabilities for a rotamer change of residues with a rotamericity larger than 25° is misleading as well because not only the four cases mentioned above do also count for this rotamericity, but PHE 248 from the PDB structure *1dm2* which has a rotamericity of 26.78 is again counted twice because it is sequence identical to both chains of the complex *1fin*. This residue moves from the second rotamer in the unbound protein to the third rotamer in the two chains of the complex. Because of this, 6 out of the 21 PHE residues are counted as rotamer changers. The 8 representatives which are flexible and have a rotamericity larger than 20° are due to 4 residues from two sequence identical unbound structures which are paired with the two chains of the 1fin complex. Therefore the high amount of rotamer changing residues for PHE in table 5.8 is due to complexed-unbound pairs of more than one sequence identical cyclin dependent kinase in the test set.

For HIS residues, the representative with a rotamericity above 30° changing the rotamer has a rotamericity of 50.79 (Entry *1br5*, chain A, residue Id 65). The range for a rotamer is 120°, therefore a distance of 51° from the median means that the side chain is nearly in the next rotamer. Because of this high rotamericity, a rotamer change is even possible for ring systems.

5.1.7. Flexibility of different rotamers

To investigate the influence of the rotamers the side chains are in of the flexibility, the flexibilities depending on the χ_1 rotamer of the side chains are shown in table 5.9.

It can be seen that the flexibility is related to the initial rotamer. In most cases, the rotamer with the lowest probability for the conformation shows the highest probability of moving away. If a side chain is in an unfavourable position (that means a high energy niveau), the energy

AName	# r1=1	# change	P(change)	# r1=2	# change	P(change)	# r1=3	# change	P(change)
ARG	261	173	0.6628	971	446	0.4593	1304	142	0.1089
ASN	622	184	0.29582	3115	82	0.02632	4160	412	0.0990
ASP	440	40	0.09091	1871	87	0.0465	1305	238	0.1824
CYS				1172	6	0.0051	4569	1	0.0002
GLN	282	273	0.9681	1277	117	0.0916	2891	551	0.1906
GLU	162	87	0.5370	208	6	0.0288	1470	41	0.0279
HIS				868	4	0.0046	213	5	0.0235
ILE	1519	325	0.214	1081	13	0.0120	4310	11	0.0026
LEU	2	1	0.5	2509	44	0.0175	4738	48	0.0101
LYS	358	333	0.9302	1700	258	0.1518	2493	409	0.1641
MET	18	14	0.7778	404	1	0.0025	51	6	0.1176
PHE							1000	5	0.005
SER	6454	1080	0.1673	2604	459	0.1763	4853	1728	0.3561
THR	2399	275	0.11463	71	51	0.7183	2735	127	0.0464
TRP				601	2	0.0033	1705	1	0.0006
TYR	1240	4	0.0032				2487	4	0.0016
VAL	446	162	0.3632	7094	798	0.1125	785	42	0.0535

Table 5.9.: Probabilities for rotamer changes of different rotamers

level which has to be added to reach the next rotamer is lower so that the side chain moves more easily.

5.2. Flexibility according to the rotamer set of the side chain

In the last section the flexibility was calculated for individual χ angles. In protein-protein docking, the flexibility of the whole side chain is important for the prediction of flexibility. In this section, the dependent flexibility is calculated according to the rotamer set of the side chain.

In table 5.10, the χ_1 flexibility is given for unbound rotamer sets of ARG residues with more than 5 representatives in the unbound conformation. In column 2 to 4, the unbound rotamers are shown, the counts for this combination can be seen in column 6. In column 7, the number of representatives from the unbound conformation which change their χ_1 rotamer upon complex formation are shown. In column 8, the probabilities for a rotamer change are given, probabilities larger than 0.5 are highlighted in yellow. It can be seen that the probability for a χ_1 rotamer change varies according to the rotamer set of the side chain, e.g. the probability for a χ_1 change if r1 is in the first rotamer varies from 0.2 to 0.58. The χ_1 flexibility is dependent on the conformation of the whole side chain.

AA	r1	r2	r3	r4	#	# change	P(change χ_1)
ARG	1	1	1	2	10	2	0.2000
ARG	1	2	2	1	16	14	0.8750
ARG	1	2	2	2	130	76	0.5846
ARG	1	2	2	3	70	65	0.9286
ARG	1	2	3	1	24	14	0.5833
ARG	2	1	1	1	24	12	0.5000
ARG	2	1	1	2	72	10	0.1389
ARG	2	1	2	1	25	15	0.6000
ARG	2	1	2	3	74	30	0.4054
ARG	2	1	3	2	35	16	0.4571
ARG	2	2	1	1	264	165	0.6250
ARG	2	2	1	2	120	74	0.6167
ARG	2	2	2	1	37	14	0.3784
ARG	2	2	2	2	45	21	0.4667
ARG	2	2	2	3	33	25	0.7576
ARG	2	2	3	2	163	49	0.3006
ARG	2	2	3	3	65	9	0.1385
ARG	3	1	2	2	9	1	0.1111
ARG	3	1	3	2	18	1	0.0556
ARG	3	2	1	1	64	14	0.2188
ARG	3	2	1	2	160	8	0.0500
ARG	3	2	1	3	21	3	0.1429
ARG	3	2	2	1	22	5	0.2273

AA	r1	r2	r3	r4	#	# change	P(change χ_1)
ARG	3	2	2	2	422	7	0.0166
ARG	3	2	2	3	79	19	0.2405
ARG	3	2	3	1	16	9	0.5625
ARG	3	2	3	2	68	12	0.1765
ARG	3	2	3	3	75	19	0.2533
ARG	3	3	1	1	11	6	0.5455
ARG	3	3	1	2	44	15	0.3409
ARG	3	3	2	1	21	2	0.0952
ARG	3	3	2	2	91	4	0.0440
ARG	3	3	2	3	49	7	0.1429
ARG	3	3	3	2	44	0	0.0000
ARG	3	3	3	3	83	7	0.0843

Table 5.10.: Probabilities for χ_1 rotamer changes depending on the rotamer set

The complete table for all residues and all rotamers can be seen in the appendix (table D.3).

5.3. Direction of side chain movement

Side chains do not only show different percentages of amino acids changing their rotamer, but also preferred directions of movements.

AA	probability for change in direction						P(all)
	1→2	1→3	2→1	2→3	3→1	3→2	
ARG	0.3142	0.3487	0.1339	0.3254	0.0406	0.0683	0.3001
ASN	0.2444	0.0514	0.0055	0.0209	0.0106	0.0885	0.0859
ASP	0.0659	0.0250	0.0390	0.0075	0.0092	0.1732	0.1009
GLN	0.1241	0.8440	0.0023	0.0893	0.0076	0.1830	0.2115
GLU	0.0679	0.4691	0.0096	0.0192	0.0061	0.0218	0.0728
ILE	0.0652	0.1488	0.0093	0.0028	0.0014	0.0012	0.0505
LEU	0.0000	0.5000	0.0000	0.0175	0.0000	0.0101	0.0128
LYS	0.1955	0.7346	0.0065	0.1453	0.0040	0.1600	0.2197
MET	0.7778	0.0000	0.0000	0.0025	0.0000	0.0000	0.0125
SER	0.0705	0.0967	0.1118	0.0645	0.3291	0.0270	0.2349
THR	0.0117	0.1030	0.4225	0.2958	0.0340	0.0124	0.0870
VAL	0.3386	0.0247	0.0214	0.0911	0.0102	0.0433	0.1204

Table 5.11.: Directions for a change of χ_1 rotamers

In table 5.11, the directions for changing the first χ angle can be seen. The probability given for each direction is the probability based on the residues being in the given rotamer. The notation 1→2 means a rotamer shift from the first to the second rotamer upon complex formation. The flexibilities of the angle not taking into account directions are given in column 8. Inflexible side chains which show a low probability for a rotamer change (CYS, HIS, PHE, TRP, TYR) are not considered.

For ARG residues r1=1 is the most unfavourable rotamer, 66% of the residues being in this rotamer move away. For the moving residues, a slight preference for the most favourable third rotamer compared to a movement to the second rotamer can be noticed. If residues from the second rotamer move, a preference for a movement to the favourable third rotamer can be seen. Side chains which are in the third rotamer show less flexibility. The main movement from the rotamer with the lowest probability can also be seen for GLN, GLU, LEU (with 50% of the residues in the first rotamer changing to the r1=3) and VAL. A preference of changing to the rotamer with the second highest probability can be seen for MET, ASP, ILE and SER residues. For MET residues this movement has a probability of 78%. A probability over 15% for a change from the most favourable to the rotamer with the second highest probability can be seen for GLN (0.183) and LYS (0.16). Side chains tend to move to a more favourable rotamer. The flexibility in unfavourable rotamers is higher. In some cases a movement away from the

most preferred rotamer can be observed. This may be due to steric clashes or reorientation of functional groups in bonds (e.g. for GLN) which force the side chain to move away.

AA	probability for change in direction						P(all)
	r1→2	r1→3	r2→1	r2→3	r3→1	r3→2	
ARG	0.5993	0.2924	0.0262	0.1162	0.0546	0.2385	0.2449
ASN	0.1597	0.1807	0.2064	0.1452	0.0913	0.0878	0.2493
ASP	0.1670	0.0591	0.1324	0.0884	0.1488	0.2241	0.2707
GLN	0.0870	0.0410	0.2414	0.0533	0.1048	0.1630	0.2654
GLU	0.0601	0.0089	0.0122	0.0061	0.0788	0.1196	0.1027
ILE	0.1130	0.0068	0.0757	0.0134	0.0067	0.0313	0.0884
LEU	0.0787	0.0234	0.1314	0.0622	0.4800	0.4600	0.1647
LYS	0.1545	0.0987	0.0262	0.0139	0.0578	0.2015	0.1026
MET	0.0000	0.0000	0.0008	0.0008	0.0000	0.3636	0.0050

Table 5.12.: Directions for a change of χ_2 rotamers

The directions for χ_2 are shown in table 5.12. The flexibility of this angle is higher, so that more movements in different directions (even within the favourable rotamers) can be seen and the tendencies are not that clear compared to χ_1. A change towards a more favourable rotamer can be seen for ARG, ASP, GLN, ILE, LYS and MET. For ASN residues, 34% of the first rotamer and 36% of the second rotamer are changing in both directions. This side chain is branched after χ_2, so that the steric constraints of placing the two ends of the branch influence this angle. For GLU side chains, the highest probability of moving can be seen in the (favoured) third rotamer. GLU is a branched side chain as well with the branch one $C-atom$ later. The position of the branch with a charge is influenced by this angle. Because the charge has to be positioned carefully and a bond mediated by the charge can compensate for energetically unfavoured movements, e.g. a change out of the rotamer with the highest probability. The sterical constraints by neighbouring amino acids are more pronounced in this angle, so that more movements which seem to be unfavourable on the first glance take place.

AA	probability for change in direction						P(all)
	r1→2	r1→3	r2→1	r2→3	r3→1	r3→2	
ARG	0.3300	0.3250	0.1230	0.2575	0.1337	0.3878	0.5008
GLN	0.1962	0.1127	0.1684	0.1165	0.4094	0.2515	0.4200
GLU	0.2385	0.2385	0.0744	0.0390	0.2513	0.1721	0.2370
LYS	0.8512	0.0707	0.0541	0.0530	0.0505	0.7870	0.2694
MET	0.0017	0.0562	0.0000	0.0000	0.4667	0.0667	0.0417

Table 5.13.: Directions for a change of χ_3 rotamers

For χ_3 (cf. table 5.13), the probabilities for changing the rotamers are high for all rotamers. The highest probabilities for a movement to the most favourable rotamer can be seen for ARG (39%) and LYS (85%) and MET(47%). For ARG and GLN residues, a high percentage of side chains changing away from the most favourable rotamer can be seen, e.g. 38% of the ARG side chains in the r3=2 and 31% of the LYS side chains being in the r3=1 change their rotamer. For χ_3, the influence of neighbouring amino acids does influence the direction of rotamer change, so that not all residues change to the energetically preferred rotamer.

AA	probability for change in direction						P(all)
	r1→2	r1→3	r2→1	r2→3	r3→1	r3→2	
ARG	0.4310	0.4423	0.1468	0.2382	0.2504	0.3535	0.4984
LYS	0.8657	0.1143	0.0683	0.1527	0.1129	0.7080	0.3217

Table 5.14.: Directions for a change of χ_4 rotamers

In the fourth χ angle (cf. table 5.14), sterical influences of neighbouring amino acids which lead

to more unfavourable movements can be seen as well. For both residues, the probability for a movement to the most favourable rotamer is high (1→2), but movements in other directions have high probabilities as well, e.g. the 1→3 movement for ARG residues or the 3→2 direction for LYS side chains.

Summarising the results of this section it can be seen that movements to more favourable rotamers are preferred,especially for the first χ angle, although some other movements can be observed which may be forced by the neighbourhood of the residues. For higher χ angles, these forced movements have a stronger influence.

5.4. Concerted rotamer changes within one side chain

The conformation of a side chain is influenced by the rotamers of all χ angles. Therefore conformational changes within a side chain are not in all cases caused by the movement of one individual χ angle, but by a concerted movement of an ensemble of χ angles. The probabilities for concerted movements are shown in this chapter.

5.4.1. Dependent probability for concerted movement

To investigate concerted movements within one side chain, the probabilities for a rotamer change given another rotamer changes its rotamer is calculated.

AA	P(1&2)	P(1&3)	P(1&4)	P(12&3)	P(12&4)	P(13&4)	P(123&4)
ARG	0.26	0.78	0.74	0.18	0.16	0.58	0.12
ASN	0.83	-	-	-	-	-	-
ASP	0.66	-	-	-	-	-	-
GLN	0.67	0.70	-	0.42	-	-	-
GLU	0.85	0.61	-	0.51	-	-	-
HIS	0.44	-	-	-	-	-	-
ILE	0.72	-	-	-	-	-	-
LEU	0.83	-	-	-	-	-	-
LYS	0.07	0.67	0.45	0.04	0.05	0.28	0.03
MET	0.07	0.93	-	0.07	-	-	-
PHE	0.18	-	-	-	-	-	-
TRP	0.67	-	-	-	-	-	-
TYR	0.25	-	-	-	-	-	-

Table 5.15.: Probabilities for a change in higher χ angles given χ_1 changed

Table 5.15 shows the dependent probability for a rotamer change if χ_1 moves. The notation p1&2 means a combined movement of the first and second χ angle. A probability larger than 0.5 for a change in χ_2 given χ_1 moved can be seen for ASN, ASP, GLN, GLU, ILE, LEU and TRP side chains. For this cases, the probabilities for a rotation in χ_2 are increased by a rotation in χ_1. A positive influence on χ_3 flexibility can be noticed for a rotation in χ_1 for all residues with a dependent probability above 0.6. These concerted movement is especially important for long amino acid side chains (ARG, LYS and MET) where the probability for a combined movement of χ_1 and χ_2 is low. For these side chains, a rotation of χ_1 moves a large volume, so that a countermovement in χ_3 is needed to avoid steric clashes.

A collective movement of more than two angles with a probability larger than 0.5 can be seen for χ_1, χ_2, χ_3 in GLU residues and χ_1, χ_3, χ_4 in ARG residues. The concerted movement of χ_1, χ_2,χ_3 for GLN residues has a probability of 0.42. For branched side chains (GLU, GLN) χ_3

moves as counterpart for χ_1 because of the large volume of the side chains, whereas a rotation in χ_2 positions the charge. For ARG residues an additional rotation in χ_4 has a probability which is 20% lower compared to the χ_1, χ_3 rotation. This movement may be important for avoiding steric clashes of the outer part of the side chain.

AA	change χ_2						
	P(1&2)	P(2&3)	P(2&4)	P(12&3)	P(12&4)	P(23&4)	P(123&4)
ARG	0.31	0.76	0.64	0.22	0.19	0.49	0.14
ASN	0.28	-	-	-	-	-	-
ASP	0.25	-	-	-	-	-	-
GLN	0.60	0.66	-	0.37	-	-	-
GLU	0.60	0.66	-	0.37	-	-	-
HIS	0.33	-	-	-	-	-	-
ILE	0.41	-	-	-	-	-	-
LEU	0.06	-	-	-	-	-	-
LYS	0.15	0.30	0.48	0.08	0.06	0.14	0.06
MET	0.17	0.33	-	0.17	-	-	-
PHE	0.03	-	-	-	-	-	-
TRP	0.04	-	-	-	-	-	-
TYR	0.33	-	-	-	-	-	-

Table 5.16.: Probability for a change in other χ angles if χ_2 or χ_3 changes

AA	change χ_3						
	P(3&1)	P(3&2)	P(3&4)	P(31&2)	P(31&4)	P(32&4)	P(312&4)
ARG	0.47	0.37	0.67	0.11	0.35	0.07	0.07
GLN	0.35	0.45	-	0.21	-	-	-
GLU	0.19	0.28	-	0.16	-	-	-
LYS	0.55	0.12	0.40	0.03	0.22	0.02	0.02
MET	0.28	0.04	-	0.02	-	-	-

Table 5.17.: Probability for a change in other χ angles if χ_2 or χ_3 changes

Tables 5.16 and 5.17 shows the probabilities for a concerted movements given one of the higher χ angles has moved. For a rotamer change in χ_2, the probability for a rotamer change in χ_3 for GLN and GLU is higher than 0.5. The movement in χ_2 increases the χ_3 flexibility. These side chains are branched, so that the rotation in χ_2 and χ_3 is needed for positioning the two ends of the branch. For ARG residues, the dependent probability for a movement of χ_2 and χ_3 is 0.76. Because ARG residues have a long side chain, more than one angle has to be moved for an optimal placement(see above). In some cases, even three angles have to be moved. The movement of χ_2, χ_3, χ_4 shows a dependent probability of 0.49.

Combined movements with probabilities larger than 0.5 given χ_3 moves can be seen for χ_3, χ_1 movements of ARG and LYS residues. Movements of these angles are also correlated in table 5.15 with a higher probability. For longer side chains, the countermovement in χ_3 is important for the avoidance of steric clashes (see above). The probability for this combination given χ_3 moved (and not χ_1 moved) is lower because the movement of χ_1 is more restricted.

5.4.2. Concerted movements from the test set

In this section, the probabilities for changes from an unbound rotamer set to a complex rotamer set is calculated. In table 5.18, the unbound rotamer combinations (columns 2-5), the complex rotamer combinations (columns 6-9), the counts for the unbound combination (column 10), the counts for the representatives which change into the complex conformation (column 11) and the probabilities (column 12) for a change can be seen for ARG residues which r1 is in the first rotamer. The probability for the perpetuation of the rotamer upon complex formation

is included and marked in yellow. Different probabilities for a change into different complex rotamers sets can be seen. As example there are 14 possible complex rotamers in which the unbound residues with the rotamer combination 1222 change with different probabilities. For the 1212 combination, no change in other rotamers of the 4 representatives can be observed. For some changes, the perpetuation of the rotamer combination upon complex formation does not show the highest probability, e.g. for the 1221 combination, a greater amount of test cases show a change to the 2231 combination (56,25%) compared to the proteins in the test set which stay in the same rotamer (6.25%). For some cases, only residues which change at least one rotamer can be seen, e.g. for the 1213 combination. This again shows that flexibility information is important when dealing with protein docking. A change especially in the third rotamer occurs quite often. The data for all amino acids without the cases where no rotamer changes in the side chain are noticed is shown in table D.4 in the appendix.

The flexibility of a rotamer combination depends on the rotamers of the χ angles itself so that a prediction of the flexibility for a rotamer combination can only be done accurately if the possibilities seen in the test set for this combination are investigated.

AA	unbound				complex				#	# change	P(change)
	r1	r2	r3	r4	r1	r2	r3	r4			
ARG	1	1	1	2	1	1	1	2	10	8	0.8000
ARG	1	1	1	2	2	2	1	2	10	1	0.1000
ARG	1	1	1	2	3	3	3	1	10	1	0.1000
ARG	1	2	1	2	1	2	1	2	4	4	1.0000
ARG	1	2	1	3	3	2	2	3	1	1	1.0000
ARG	1	2	2	1	1	2	2	1	16	1	0.0625
ARG	1	2	2	1	1	2	2	3	16	1	0.0625
ARG	1	2	2	1	2	2	1	2	16	1	0.0625
ARG	1	2	2	1	2	2	3	1	16	9	0.5625
ARG	1	2	2	1	2	2	3	3	16	2	0.1250
ARG	1	2	2	1	3	1	1	2	16	1	0.0625
ARG	1	2	2	1	3	2	3	3	16	1	0.0625
ARG	1	2	2	2	1	2	1	3	130	1	0.0077
ARG	1	2	2	2	1	2	2	1	130	9	0.0692
ARG	1	2	2	2	1	2	2	2	130	23	0.1769
ARG	1	2	2	2	1	2	3	2	130	3	0.0231
ARG	1	2	2	2	1	2	3	3	130	18	0.1385
ARG	1	2	2	2	2	2	1	2	130	1	0.0077
ARG	1	2	2	2	2	2	2	3	130	18	0.1385
ARG	1	2	2	2	2	2	3	2	130	2	0.0154
ARG	1	2	2	2	2	3	2	2	130	9	0.0692
ARG	1	2	2	2	3	2	1	1	130	9	0.0692
ARG	1	2	2	2	3	2	1	3	130	9	0.0692
ARG	1	2	2	2	3	2	3	1	130	9	0.0692
ARG	1	2	2	2	3	2	3	3	130	10	0.0769
ARG	1	2	2	2	3	3	2	2	130	9	0.0692
ARG	1	2	2	3	1	2	1	3	70	3	0.0429
ARG	1	2	2	3	1	2	2	3	70	2	0.0286
ARG	1	2	2	3	2	2	1	2	70	5	0.0714
ARG	1	2	2	3	2	2	3	1	70	18	0.2571
ARG	1	2	2	3	2	2	3	2	70	6	0.0857
ARG	1	2	2	3	2	2	3	3	70	4	0.0571
ARG	1	2	2	3	3	2	3	3	70	32	0.4571
ARG	1	2	3	1	1	2	2	1	24	2	0.0833
ARG	1	2	3	1	1	2	2	2	24	4	0.1667
ARG	1	2	3	1	1	2	3	3	24	4	0.1667
ARG	1	2	3	1	2	2	2	3	24	4	0.1667
ARG	1	2	3	1	2	3	2	2	24	2	0.0833
ARG	1	2	3	1	3	2	1	1	24	2	0.0833
ARG	1	2	3	1	3	2	1	3	24	2	0.0833
ARG	1	2	3	1	3	2	3	1	24	2	0.0833
ARG	1	2	3	1	3	3	2	2	24	2	0.0833
ARG	1	2	3	2	1	3	2	3	2	1	0.5000
ARG	1	2	3	2	3	2	1	1	2	1	0.5000
ARG	1	2	3	3	1	2	3	3	4	4	1.0000

Table 5.18.: Probabilities for changes from unbound to complex rotamer sets for ARG residues with r1=1 including the perpetuation of the rotamers

5.5. Flexibility for the test set with standardisation of sequence identical chains

For the tables shown in this chapter so far, unbound proteins which are sequence identical to the same complex partner are all taken into account to reflect different changing behaviours of residues from independently solved unbound structures. If more than one sequence identical case is known for one complex, flexibilities for residues may be over- or underrepresented depending on the occurance of a rotamer change in the identical structures. For exclusion of these effects, sequence identical unbound proteins pairing with the same complex are summed. The counts for each rotamer is divided by the number of representatives of unbound proteins, so that all rotamer changers are taken into account, but over- or underrepresentation is avoided by normalising the counts to one.

AA	change $\chi 1$	change $\chi 2$	change χ_3	change χ_4
ARG	0.3107	0.2475	0.5076	0.4434
ASN	0.1332	0.3241	–	–
ASP	0.1364	0.3306	–	–
CYS	0.0108	–	–	–
GLN	0.2106	0.255	0.4422	–
GLU	0.1801	0.2224	0.3973	–
HIS	0.0457	0.0672	–	–
ILE	0.0719	0.1418	–	–
LEU	0.0543	0.1993	–	–
LYS	0.2385	0.1512	0.2794	0.3688
MET	0.0163	0.044	0.1035	–
PHE	0.0402	0.0402	–	–
SER	0.2503	–	–	–
THR	0.1169	–	–	–
TRP	0.0159	0.0344	–	–
TYR	0.0129	0.0107	–	–
VAL	0.1329	–	–	–

Table 5.19.: Probabilities of different χ angles with normalised sequence identical unbound chains

In table 5.19, the probability for a rotamer change is calculated for non-sequence identical unbound chains. The probability for residues with an increase more than 0.05 are highlighted in yellow. The flexibility is comparable with the flexibility over all data shown in table 5.1 with a difference less than 0.05. Higher flexibilities compared to table 5.1 are marked in yellow. Higher flexibilities can be noticed for all angles of GLU which is now more comparable with the flexibility of GLN. These two side chains have a similar structure, so that identical flexibility is expected. Higher probabilities for changing the χ_2 rotamer can be seen for ASN, ASP, HIS and ILE residues, MET is more flexible in the third χ angle. A reduced flexibility can be seen for the χ_4 of ARG residues. For a few cases, an over- or underestimation for some angles due to more sequence identical unbound proteins can be seen.

The flexibility of non-sequence identical unbound chains in *sheets* can be seen in table 5.20. For the first χ angle the flexibility is comparable to the flexibility in table 5.6 (without exclusion of sequence identical chains) except for ASP residues where the flexibility is twice as high. The χ_2 flexibility stays on the same level, while for χ_3, an increase of the flexibility can be seen for ARG and GLN side chains. The fourth χ angle is more flexible compared to the data with sequence identical unbound chains.

In table 5.21, the helix data with exclusion of sequence identical unbound cases is shown. For helices, two trends can be noticed: a higher flexibility or a decrease of the flexibility. A higher

AA	change $\chi 1$	change $\chi 2$	change χ_3	change χ_4
ARG	0.1735	0.2991	0.4466	0.3637
ASN	0.2462	0.3429	–	–
ASP	0.1200	0.48	–	–
CYS	0.0000	–	–	–
GLN	0.0667	0.0685	0.1859	–
GLU	0.3125	0.3438	0.5106	–
HIS	0.0000	0.0333	–	–
ILE	0.0219	0.0949	–	–
LEU	0.0244	0.1644	–	–
LYS	0.2273	0.1263	0.2661	0.2667
MET	0.0130	0.0000	0.0384	–
PHE	0.0000	0.0000	–	–
SER	0.0747	–	–	–
THR	0.1789	–	–	–
TRP	0.0000	0.000	–	–
TYR	0.0000	0.000	–	–
VAL	0.1449	–	–	–

Table 5.20.: Probabilities of different χ angles with normalised sequence identical unbound chains in sheets

AA	change $\chi 1$	change $\chi 2$	change χ_3	change χ_4
ARG	0.4045	0.1396	0.5805	0.4729
ASN	0.0112	0.2584	–	–
ASP	0.3158	0.3905	–	–
CYS	0.0000	–	–	–
GLN	0.3106	0.5084	0.5766	–
GLU	0.1277	0.1902	0.3555	–
HIS	0.0696	0.0971	–	–
ILE	0.1502	0.1767	–	–
LEU	0.0840	0.1862	–	–
LYS	0.1953	0.1296	0.249	0.3632
MET	0.0000	0.0238	0.0714	–
PHE	0.0667	0.0444	–	–
SER	0.3097	–	–	–
THR	0.0962	–	–	–
TRP	0.0448	0.0448	–	–
TYR	0.0000	0.0000	–	–
VAL	0.1357	–	–	–

Table 5.21.: Flexibility of different χ angles without sequence identical chains in helices

χ_1 flexibility for the dataset without sequence identical chains compared to table 5.4 can be seen for GLU, LEU and LYS residues. The side chains of ASN, GLU, HIS, LEU and LYS show a higher flexibility for in the second χ angle. For χ_3, a higher flexibility can be seen for GLU, LYS and MET residues. A reduction of the probability for a rotamer change occurs for the χ_1 of ASP and ILE residues, and in the χ_2 of LEU residues.

AA	over all		helix		sheet		RND	
	test set	without identical	test set	without identical	test set	without identical	test set	without identical
ARG	2536	353	601	89	397	49	392	60
ASN	7897	565	819	89	795	65	831	87
ASP	3616	294	177	31	155	25	729	89
CYS	6321	369	1342	120	1276	86	989	79
GLN	4450	326	439	63	1114	90	167	33
GLU	1840	198	326	64	22	16	773	57
HIS	1496	124	264	24	270	30	277	31
ILE	6910	454	815	104	1699	137	1070	70
LEU	7249	537	649	119	2084	164	1242	120
LYS	4551	320	851	99	1164	88	514	51
MET	1199	91	286	42	513	37	1	1
PHE	1881	199	189	45	793	73	156	26
SER	13911	803	1323	150	2018	131	1800	141
THR	5201	452	424	52	825	95	1069	103
TRP	2717	189	685	67	510	34	23	7
TYR	4754	310	78	20	679	73	1032	76
VAL	8325	595	1249	137	2363	207	308	46

Table 5.22.: Counts for test cases with and without normalising unbound chains

For some residues different flexibilities with and without normalised sequence identical unbound chains can be observed, whereas the flexibilities stay on the some level for most residues. By excluding these residues, the representatives for each amino acid especially for different Secondary Structures fall off (see table 5.22).

5.6. Visual inspection of residue flexibility

For visualising the flexibility of different complexes, the 3D visualisation tool viwish [24] is used. It is TCL/TK based with a client-server architecture. The server is listening for events, e.g. strings send from clients. For the visualisation of flexibility, the amino acids are coloured according to float values selected from the MYSQL database system. An example of the SQL query realising the colouring can be seen below:

mysql -B -N -e " select "send viwish pdb1gp2 1 residuecolor", Res_IdC, getRGB(if(countsWechsel=0, 0.2, if(countsWechsel=1, 0.4, if(countsWechsel=2, 0.6,if(countsWechsel=4, 0.8, 1.0)))),"rainbow") from Vergleich_Winkeldaten where Entry_Complex="1GP2" and Res_NameC!="PRO" and Res_NameC!="ALA" and Res_NameC!="GLY" and Entry_Unbound="1CIP" and Chain_IdU="A" " kerstin | /vol/elmar/bin/wish

The SQL statement selects the Residue Identifier of the amino acid which should be coloured without taking into account non-rotameric amino acids. The getRGB() function gets a float value and a colour string (e.g. "red", "rainbow") and returns RGB values according to colour string and input values. The if statement within the getRGB() function selects the number of χ angles which change their rotamer and and sets the float value accordingly, so that residues with different numbers of rotamer changers in the side chain are shown in different colours.

In figure 5.2, the residues of the complex *pdb3bth* (chain E) are coloured depending on the number of angles in the side chain of the enzyme which change their rotamer upon complex formation compared to the unbound structure *pdb1c2j*, chain A. The grey residues are non-rotameric (PRO, GLY, ALA), they are not coloured with the if statement used for RGB value generation. Blue coloured residues show no flexibilities upon complex formation, residues shown in green change in one side chain rotamer. A higher flexibility in two side chain angles can be observed for amino acids shown in orange. Side chains with a purple colour change in three rotamers. Very flexible side chains which change in four angles are shown in red (not seen in figure 5.2). It can be seen that large parts of the *3bth* E chain do not undergo conformational changes upon complex formation, especially residues in the core. The PDB structure *3bth* is a β-trypsin in complex with the bovine pancreatic trypsin inhibitor (BPTI, chain I). The inhibitor is shown in red and grey cylinders at the right hand side. The structure was deposited in 1999 with a resolution of 1.75 Å. The unbound trypsin structure used for comparison (*1c2j*) has a resolution of 1.4 Å. The LYS residues with the Residue Identifier 109 (chain A) changes in 3 rotamers upon complex formation. This residue is at the surface of a protein so that it can rotate freely. At the active site, there are a few residues which are flexible. For GLN 192 a change in two of the three side chain angles can be observed. Other residues in the active site

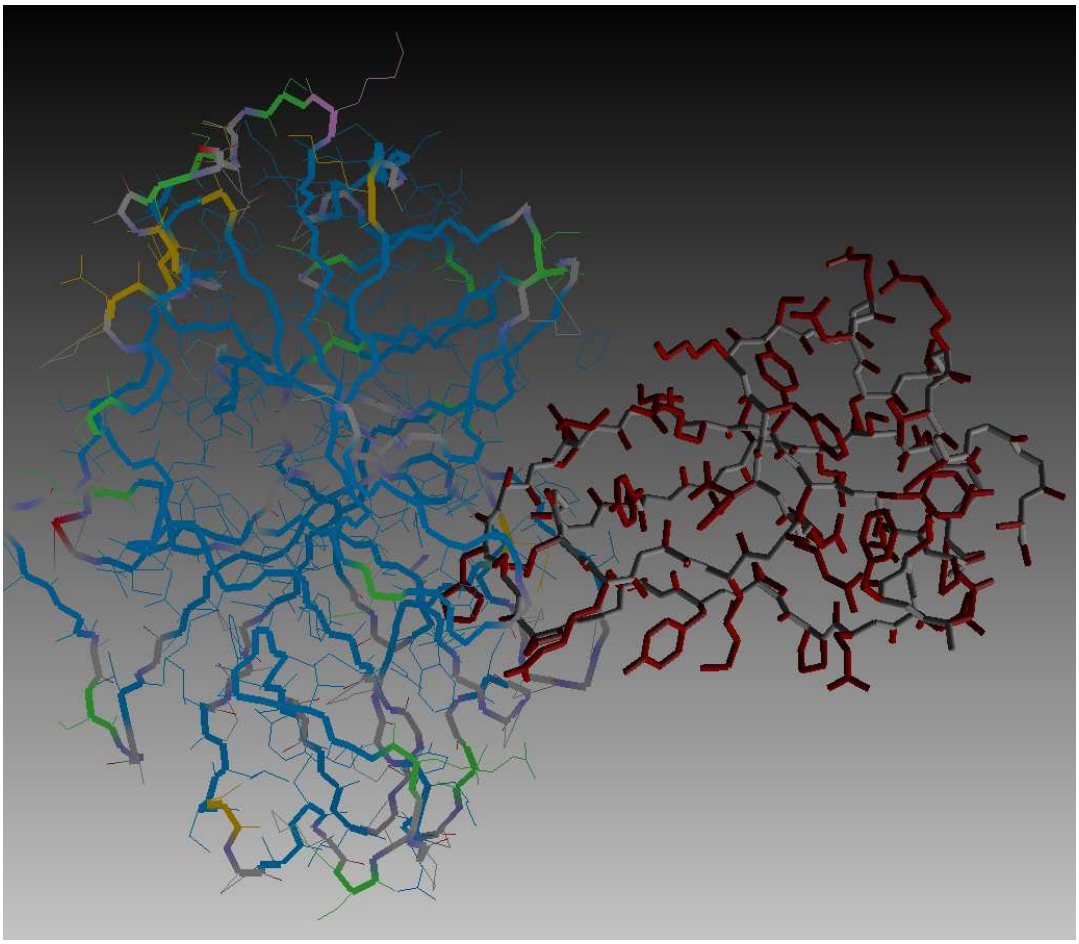


Figure 5.2.: Flexibility of a trypsin-BPTI complex (*pdb3bth*), the residues of chain A are coloured according to the number of angles which change their rotamer upon complex formation, the inhibitor is shown in red-gray cylinders

are inflexible. The other representatives which change in two rotamers are at the surface of the protein: ASN 48, GLN 50, LYS 87, ASP 165, GLN 240 and ILE 242.

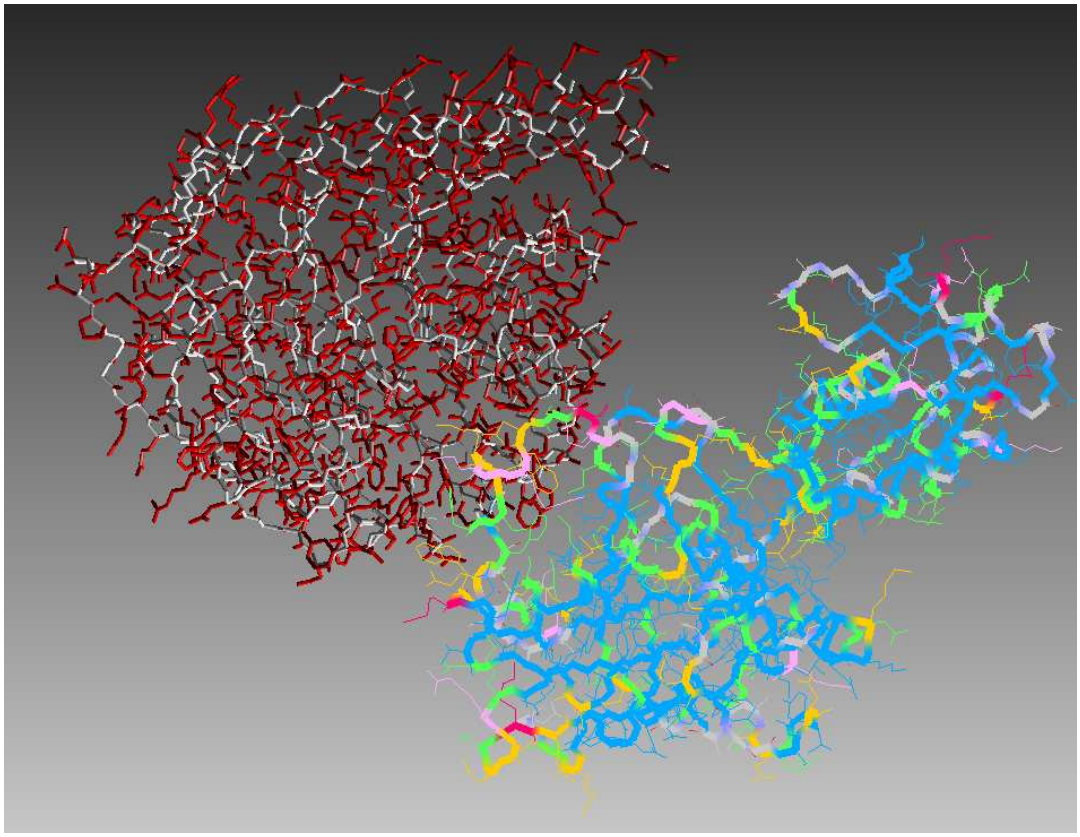


Figure 5.3.: Flexibility for a G-protein (*1gp2*), the residues of the α chain are coloured according to the number of angles which change their rotamer upon complex formation

In figure 5.3 the flexibility of a Guanine Nucleotide binding protein (G-protein) α-fragment upon binding its transducer (GTP) can be seen. The structure *1gp2* is from *Rattus norvegicus* and has a resolution of 2.3 Å (deposited in 1996). A comparison with PDB *1cip* (resolution 1.5 Å) was done to gain flexibility information. This molecules shows a larger amount of flexibility compared to the enzyme shown in 5.2. There are five residues changing in 4 rotamers upon binding: the ARG residues with the Residue identifiers 105, 129 and 205, and the LYS side chains with the identifier 257 and 317. These residues are shown in red. Eleven residues change their rotamer in three side chain angles. The transducer binds at the crevice between the upper right and lower left part of the chain. The flexiblity of this protein is higher compared to the enzyme seen in figure 5.2. The very flexible amino acids can be found at the surface of the protein.

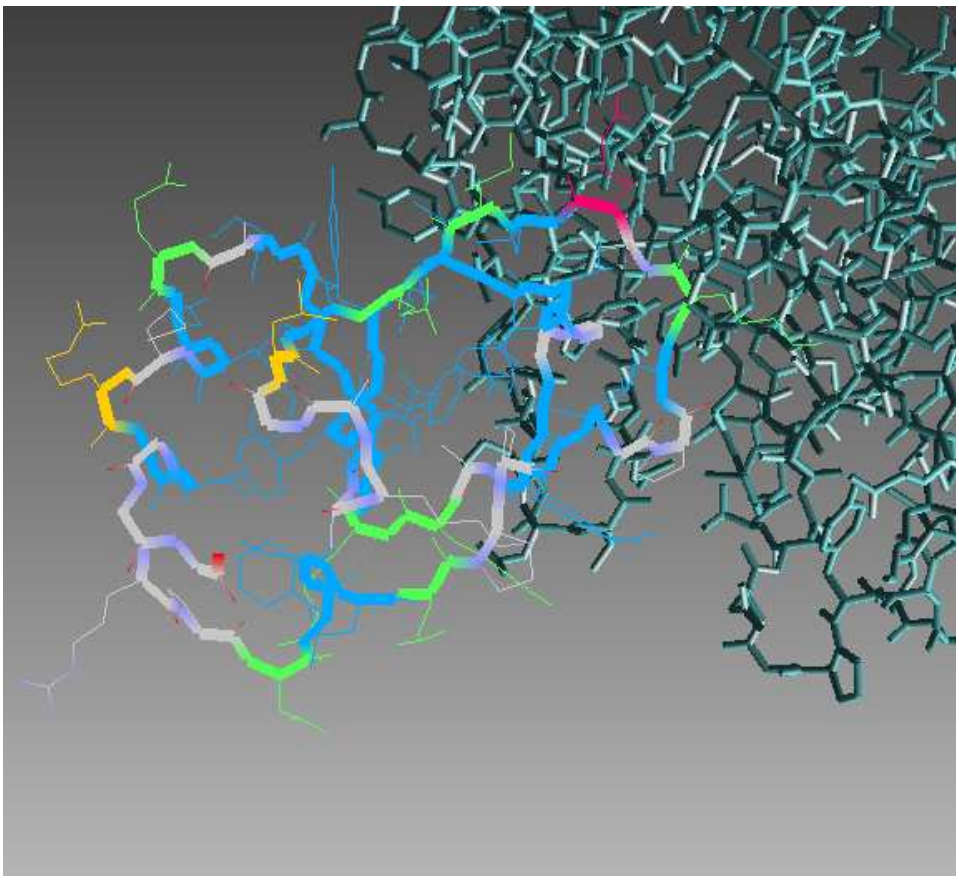


Figure 5.4.: Flexibility for BPTI in complex with β-trypsine (*pdb2ptc*), the residues of the inhibitor are coloured according to the number of side chain angles changing their rotamer

Figure 5.4 shows the flexibility of the bovine pancreatic trypsine inhibitor in complex with β-trypsin compared to the unbound inhibitor *1BPI*. At the contact site of the inhibitor, the ARG residue with the Residue Identifier 17 changes in all 4 rotamers (coloured in red). A change in 2 rotamer can be seen for LEU 29 and ARG 53. These residues are at the surface of a protein so that a free rotation is possible. There are two residues at the active site which change in one side chain rotamer: LYS 14 and ILE 18. These residues are coloured in green. For the BPTI molecule, more flexibility can be seen compared to the enzyme (see figure 5.2) .

6. Database

The data of the test set and the statistical results are stored in a MYSQL database system, so that the data used for calculation is easily accessible and the statistical results can easily be stored. Information about the different amino acids (e.g. flexibility) can be accessed by the docking algorithm and added to the proteins in a preprocessing phase. Because this is done before the docking run starts, the time needed for the calculation of docking hypothesis is not affected.

The residues in the different tables can be identified by the PDB Entry, Chain Identifier and the Residue Identifier. These attributes are called Entry, Chain Id and Res Id in the MYSQL tables. For comparing residues in unbound and complexed structures an alignment is done to introduce an unique residue identifier called Res Pos. Comparable amino acids can be found by Entry, Chain Id and the Res Pos. The residue identifier given in the PDB file is not unique because the length of two comparable proteins may differ or sometimes a number letter combination is used as identifier if the residue enumeration is based on (shorter) homologous proteins. Some other information about the residues is stored in the database system as well, e.g. Secondary Structure, rotamericity, SAS area and information from the header of the protein like the classification.

6.1. The test set

The test set used in the thesis is taken from the PDB [4]. All proteins with a length of more than 35 residues and a resolution between 0.1 and 2.5 Å are taken as test cases. By this length restriction shorter peptides which can often be found as peptide inhibitors of enzymes are excluded. The resolution is a measurement for the quality of the proteins. In structures with a resolution larger than 2.5 Å, the position of the atoms needed for calculating the torsion angles is not exact enough. Structures with zero resolution are non x-ray structures of proteins, but nuclear magnetic resonance (NMR) structures or theoretical models. Protein structures gained by the NMR technique consists of an ensemble of structures in different conformations, so that the calculation of torsion angles is not possible, although flexibility information can be extracted from some NMR structures (personal communication with crystallographers).

For the classification of the test set in unbound and complexed proteins, some heuristics were used [34]. Proteins consisting of only one chain are characterised as unbound cases, whereas structures with two non-identical chains are classified as complex. Another possibility for complex classification are the chain identifiers from the PDB file. The chain identifier for inhibitor chains in complexes is I in most cases, so that a protein with two chains where one of these chains has the identifier I can be taken as complex. This methods works (in contrast to

the first heuristic) likewise for complexes with three chains where one is classified as inhibitor. A third possibility for classification of PDB structures are the pdb_at_a_glance [40] and CATH [38] databases. In the pdb_at_a_glance database, the PDB structures are classified according to their group (e.g. cofactor binding proteins, calcium binding proteins) and complexed or unbound representatives. The CATH classification is based on domains. The complexes in the test set are characterised according to the first method. The unbound PDB Entries of the test set can be seen in table 6.1, the complex Entries in table 6.2.

Entry Unbound	Entry Unbound	Entry Unbound	Entry Unbound
132L	135L	193L	194L
1AKI	1AKL	1AKZ	1ANE
1ANG	1AQ1	1AQ7	1AQP
1AUJ	1AXB	1AZ8	1AZF
1B0D	1B0E	1BE6	1BE8
1BEL	1BFK	1BFU	1BGI
1BJI	1BJU	1BJV	1BOF
1BP4	1BPI	1BQI	1BR5
1BRA	1BT5	1BTL	1BTP
1BTU	1BTY	1BVX	1BWH
1BWI	1BWJ	1C10	1C1M
1C1N	1C1O	1C1P	1C1Q
1C1R	1C1S	1C1T	1C2D
1C2E	1C2F	1C2G	1C2H
1C2I	1C2J	1C2K	1C2L
1C2M	1CCP	1CE5	1CHG
1CIP	1CJC	1CKP	1CVZ
1DM2	1DPW	1DPX	1E1K
1E1L	1E1M	1E1N	1EAS
1EAT	1EAU	1ECY	1ELF
1ELG	1F0W	1F10	1FKB
1FKD	1FKF	1FKG	1FKH
1FKJ	1GBT	1GCD	1GDD
1GFI	1GIA	1GLN	1GMH
1HCK	1HCL	1HEL	1HEW
1HOE	1HPT	1HSW	1HSX
1JEF	1JIM	1JPO	1JSE
1LPI	1LSA	1LSB	1LSC
1LSD	1LSE	1LSF	1LVY
1LZN	1LZT	1LZY	1MAX
1MAY	1MWE	1P09	1POH
1QA0	1QB1	1QB6	1QB9
1QBN	1QBO	1QCP	1QGF
1QL7	1QNJ	1QTK	1TIO
1TNJ	1TNK	1TNL	2ANG
2BZA	2CI2	2TIO	8EST

Table 6.1.: Unbound structures from the test set

Entry Complex	Entry Complex	Entry Complex	Entry Complex
1A14	1A4Y	1ACB	1AK4
1AZZ	1BRB	1BRC	1BVN
1BZX	1C08	1CGI	1CGJ
1CHO	1CSE	1D0D	1D6R
1DFJ	1DQJ	1DZB	1EAI
1F5Q	1FDL	1FIN	1FLE
1FY8	1G7H	1G7I	1G7J
1G7L	1G7M	1GP2	1KIP
1KIQ	1KIR	1MEL	1NSG
1SBN	1SIB	1STF	1TAB
1TGS	1TPA	1UGH	1VFB
1WQ1	2AAI	2FAP	2JEL
2KAI	2PCC	2PTC	2SEC
2SIC	2SNI	2TGP	2TPI
3BTD	3BTE	3BTF	3BTG
3BTH	3BTK	3BTM	3BTQ
3BTT	3BTW	3FAP	3PRO
3SIC	3TGI	3TPI	4PRO
4TPI	5SIC		

Table 6.2.: Complex structures from the test set

6.1.1. Composition of the test set considering enzyme families

The enzyme family composition of the used test set is investigated using the EC numbers of the enzymes given in the BRENDA database [43]. The EC numbers are a hierarchical classification for enzymes related to the chemical reaction they catalyse. These numbers are given to enzymes from the International Union of Biochemistry and Molecular Biology (IUBMB). The complete list of EC numbers and appropriate enzymes can be seen in [36]. A four letter code is assigned for each enzyme with a progressively finer classification. In the **EC 1** group *Oxidoreductases* which catalyses oxidation/reduction reactions can be found, the **EC 2** group consists of *transferases* which transform functional groups between molecules. **EC 3** stands for *hydrolases* which use a water molecule to break chemical bonds. In the **EC 4** group *lyases* which cleave chemical bonds by other means than hydrolysis or oxidation are combined. The fifth group contains enzymes which catalyse isomerisation reactions within a molecule, the *isomerases*. In the last EC group (**EC 6**) *ligases* which join two molecules by covalent bonds can be found. The EC classification of the test proteins can be seen in 6.3.

EC_number	classification	unbound	complex
1.11	acting on peroxide	281	-
1.18	acting on iron sulfur proteins	906	-
2.7	transfer of P-groups	1408	598
2.7.1	alcohol as acceptor	1136	598
3.1	acting on ester bonds	244	364
3.2	glucosylases	5447	1953
3.2.1	glucosidases	4961	1467
3.4	acting on peptide bonds	15000	7225
3.4.21	Serin Endopeptidases	14322	7015
3.4.22	Cysteine Endopeptidases	210	210
3.4.24	Metallo Endopeptidases	468	-
3.5	acting on carbon-nitrogen bonds other than peptides	783	-
3.5.2	cyclic amides	783	-
5.2	cis-trans isomerases	105	638

Table 6.3.: EC classification

For 5714 unbound and 4922 complexed test cases, no EC numbers are available. Because EC numbers are a classification method for enzymes, not all pdb structures have EC numbers. There are some proteins in the test set which are not enzymes, e.g. receptors, proteins involved in transport of electrons or oxygen, viral proteins,cell adhesion proteins and the enzyme inhibitors inhibitors. For some cases, no EC numbers are assigned in the header of the PDB Entry. The EC number classification for unbound and complexed enzymes of the test set can be seen in table 6.4, with the counts for unbound and complexed representatives in column 5 and 6.

EC_number	unbound	complex	classification
1.11.1.5	281	-	Cytochrome C Peroxidase
1.18.1.2	906	-	Ferrodoxin-NADP Reductase
2.7.1.37	1136	-	Protein Kinase
3.1.27.5	244	122	Pancreatic Ribonuclease
3.2.1.17	4189	1012	Lysozyme
3.2.1.18	772	386	Exo-α-Sialidase
3.2.2.3	-	321	
3.2.2.22	265	265	rRNA N glucosidase
3.4.21.1	464	469	Chymotrypsin
3.4.21.4	9491	3803	Trypsin
3.4.21.8	-	55	Kallikrein
3.4.21.11	238	-	Elastase
3.4.21.12	195	-	α lytic Endopeptidase
3.4.21.14	-	335	transferred, new EC number
3.4.21.32	-	3	Brachyrin
3.4.21.36	2847	714	Elastase (pancreatic)
3.4.21.62	1087	1636	Subtilisin
3.4.22.2	210	210	Papain

EC_number	unbound	complex	classification
3.4.24.40	468	-	Serralysin
3.5.2.6	783	-	β-lactamase
5.2.1.8	105	638	Peptidyl-Isomerase

Table 6.4.: EC statistics for unbound and complexed residues in the test set

6.2. MYSQL tables

In the MYSQL database the general information for each protein like Entry, Chain Identifier, residue name, sequence and the residue position identifier given by an alignment of sequence identical structures is stored on the one hand, the outcome of the statistical investigation on the other hand. The main tables with the data used for statistical investigations are the Winkeldaten_Unbound and Winkeldaten_Complex table. In this tables, the PDB Entry, Chain identifier, residue name, residue position, residue identifier and the ϕ, ψ and χ angles with their corresponding rotamers for unbound proteins or complexes are stored. Furthermore information about Secondary Structure and rotamericity can be accessed. In the Vergleich_Winkeldaten table, these informations from the unbound and complexed proteins are joined with the information about sequence identical proteins and the unique residue identifier. In this table, the data of comparable unbound and complexed residues is stored. The number of residues used for statistical calculations can be seen in 6.5.

AA	Vergleich Winkeldaten	Winkeldaten Unbound	Winkeldaten Complex
ARG	2536	1314	585
ASN	7897	1873	1040
ASP	3616	1288	626
CYS	6321	1145	620
GLN	4450	1210	596
GLU	1840	993	486
HIS	1496	512	268
ILE	6910	1597	830
LEU	7249	2202	1067
LYS	4551	1503	756
MET	1199	386	209
PHE	1881	740	432
SER	13911	2942	1561
THR	5205	1707	917
TRP	2717	664	324
TYR	4754	1145	602
VAL	8325	2257	1214

Table 6.5.: Counts for residues in database tables

It can be seen that not all residues are represented in the same amount in the test set. The residues where most representatives are available in the test set are SER (13% of the residues in the Vergleich Winkeldaten table, 10% in the complex and unbound table), VAL (8%) and ASN (7%). Less representatives can be found for TRP (2% of the residues in the test set), ARG (2%), PHE 2%), GLU (2%), HIS (2%) and MET (1%) residues [1].

In table 6.6, the number of complex residues having atoms with a distance less than 4 Å from atoms of the second complex part are shown. These representatives are taken as active site residues. Because the distance of 4 Å is more than the distance for bonds, not only residues building hydrogen bonds but also residues at the border of the active site are taken as active site residues.

[1]The percentages are calculated taking into account ALA, GLY and PRO residues which are included in the database tables but not shown here and not considered for statistical calculations

Res Name	number of representatives
ARG	749
ASN	952
CYS	391
GLN	281
GLU	380
HIS	948
ILE	529
LEU	1055
LYS	583
MET	226
PHE	988
SER	1142
THR	911
TRP	305
TYR	721
VAL	902

Table 6.6.: Number of representatives which are at the active site of a complex

Res Name	Helix	α-Helix	α-Helix without ends	Strand	Turn	RND
ARG	158	153	72	90	73	75
ASN	102	58	35	70	175	92
ASP	96	84	43	42	68	109
CYS	116	65	47	117	23	86
GLN	120	100	66	128	90	42
GLU	132	114	76	27	68	90
HIS	43	18	6	40	15	51
ILE	154	142	83	186	44	105
LEU	247	214	151	240	93	169
LYS	153	141	94	136	53	107
MET	57	47	36	54	2	3
PHE	82	71	57	93	11	41
SER	194	170	64	177	184	215
THR	117	109	74	116	64	132
TRP	74	62	51	50	25	13
TYR	33	28	14	93	32	98
VAL	175	132	59	107	14	82

Table 6.7.: Test cases for unbound Secondary Structure

The number of residues in Secondary Structure elements can be seen in table 6.7 for unbound residues and table 6.8 for amino acids in complexes. The percentage of residues of one type being in helices varies from 3% (TYR) to 15% (MET, LEU). For LEU, MET, GLU, ARG, TRP, LYS, CYS and PHE residues at least 10% of the representatives can be found in helices. The preference for $\alpha - helices$ (column 2) over other helix types (column 3) varies from 56% (CYS) to 97% (ARG). Residues with a percentage of helices above 90% additional to the mentioned ARG are THR, ASP, ILE and LYS. For pure $\alpha - helices$ without end residues less data can be found for some amino acids which do not prefer helical environments, e.g. for HIS residues. For residues in $\beta - sheets$ (see column 6), the percentage of the representatives found in this Secondary Structure varies from 14% (MET) to 3% (GLU). For MET, PHE, ILE, LEU, GLN and CYS, more than 10% of the residues can be found in strands. Turns are only short parts of Secondary Structures, so that only few representatives can be found. The fraction of the residues in turns varies from 7% (GLN) to 0.5% (MET). The highest percentage of residues in random coils can be seen for THR residues (14%), the lowest amount of representatives in random coils can be seen for VAL (4%), GLN (3%) and TRP (2%) residues.

In table 6.8, the number of residues in Secondary Structure elements can be seen for proteins in complexes. The fraction of residues being in helices varies from 5% (TYR) to 20% (MET) for complex residues. The percentage differs slightly from the percentage calculated for unbound structures. The residues with at least 10% of their representatives in helices not including MET can be found for TRP (17%), CYS (15%), ARG (14%) GLU (13%), PHE (12%) and ILE (10%). Less representatives can be found for THR (7%), SER (6%) and TYR (5%). For

Res Name	Helix	α-Helix	Strand	Turn	RND
ARG	84	83	59	58	58
ASN	77	62	48	127	98
ASP	43	37	31	47	108
CYS	94	65	83	26	58
GLN	50	38	91	42	34
GLU	65	62	38	32	52
HIS	24	12	30	18	33
ILE	86	77	125	31	66
LEU	100	76	157	56	117
LYS	96	91	87	54	76
MET	42	34	31	8	4
PHE	52	38	75	17	47
SER	101	79	101	111	149
THR	67	54	125	71	98
TRP	55	40	34	29	20
TYR	29	22	80	26	43
VAL	105	79	221	9	86

Table 6.8.: Test cases for complex Secondary Structure

strand residues, the number of residues found in this Secondary Structure are higher for many residues with a fraction above 10% for VAL (18%), PHE (17%), GLN, ILE , MET, LEU (all 15%),THR (14%) and TYR, CYS (both 13%), LYS (12%), HIS (11%), TRP and ARG (10%). For SER, ASN and ASP, less residues are in strands (6% for SER, 5% for ASN, ASP). The highest amount of amino acids assigned as belonging to random coils and thus to no Secondary Structures can be seen for MET (36%), ASP (17%), GLU (12%) LEU, PHE, GLU and THR (11%) and LYS, ARG, SER (10%), the lowest amount can be seen for GLN (6%) and TYR (3%).

The percentages shown for the different Secondary Structure elements are calculated with the representatives in the test set having a Secondary Structure assigned by the DSSP database. Because this information is not available for all structures in the test set, the fraction of representatives in Secondary Structures is underestimated.

7. Evaluation

In this section of the thesis, the rotamer libraries are evaluated by their relative saving in pruning the search tree. After a short introduction showing the calculation of the relative saving in section 7.1, the rotamer libraries compiled on unbound data are evaluated in section 7.2 followed by the evaluation for rotamer libraries on complex data (see section 7.3).

7.1. Evaluation criteria for rotamer libraries

As measurement for the performance of the libraries, the pruning of the search tree when searching for a given side chain conformation compared to full search is used. Therefore the rotamer combinations of the side chains are ranked according to their probabilities. To calculate the relative saving, the actual rank of the rotamer combination found in a PDB structure is divided by the maximum rank of the investigated amino acid. The maximum rank is the number of possible rotamers for the investigated amino acid. In the worst case where the searched conformation is at the last leaf of the search tree, the number of visited nodes is the maximum rank.

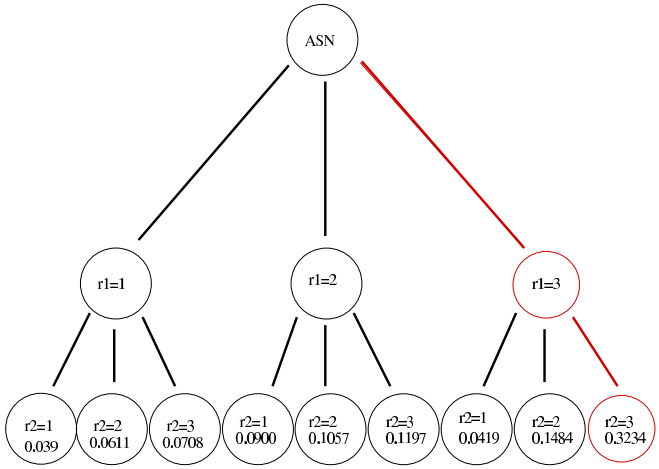
$$relative \quad saving = 1 - (\frac{rank}{maximum \quad rank}) \tag{7.1}$$

A sample search can be seen in figure 7.1. The example ASN residue in the protein has the rotamer combination r1=3 and r2=3 (which has the highest probability). If the search is done in the unsorted tree without ranking the rotamers according to their probability (see figure 7.1(a)), all nodes of the tree have to be visited to find the right conformation. The relative saving for this example can be seen in formula 7.2.

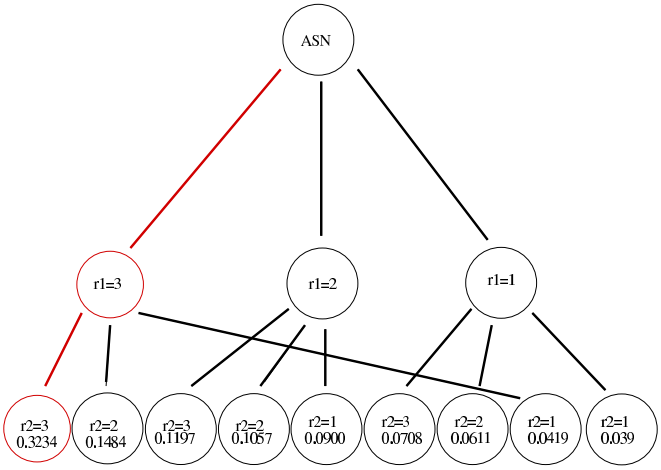
$$relative \quad saving = 1 - (\frac{9}{9}) = 0 \tag{7.2}$$

If the rotamer combinations are ranked according to their probability, the combination is found at the first leaf and the rest of the tree can be pruned (see figure 7.1(b)). The relative saving for this case is 0.99. The relative saving is calculated as average of the relative saving over all amino acids of a protein and written into MYSQL tables.

The independent probabilities of the rotamer combinations are given in the rotamer libraries, because no priori information about the χ_1 rotamer is known if the whole side chain is placed. The language model (see chapter 4.3.1) used for calculation of the probability yields these probabilities although they are internally approximated by dependent probabilities. In the



(a) not sorted



(b) sorted

Figure 7.1.: Unsorted and sorted search tree for one ASN residue. The probabilities are written below the rotamer combinations

Dunbrack rotamer library [11], the dependent and independent probabilities are given for each rotamer combination. The dependent probabilities are higher because they sum up to one for each χ_1 rotamer. The rank assigned for the new rotamer libraries are compared to the rank of the Dunbrack library (as of 1999) given by the independent probability.

7.2. Evaluation of the rotamer library on unbound data

In this section of the thesis, the average relative saving for the unbound rotamer libraries are calculated on the unbound proteins of the test set.

AA	ALL	HELIX	STRAND	RND	Dunbrack
ARG	0.8051	0.776	0.6773	0.7055	0.8585
ASN	0.6474	0.5938	0.2887	0.6603	0.6594
ASP	0.6413	0.4967	0.4889	0.5981	0.6341
CYS	0.5222	0.5223	0.4972	0.5223	0.5223
GLN	0.7051	0.7032	0.6986	0.6459	0.7817
GLU	0.7148	0.6787	0.5803	0.6789	0.7119
HIS	0.6824	0.6451	0.5365	0.6062	0.6838
ILE	0.6605	0.6651	0.6771	0.6904	0.7479
LEU	0.7436	0.7443	0.7755	0.7436	0.8154
LYS	0.8527	0.8327	0.817	0.7949	0.8826
MET	0.7596	0.6944	0.6295	0.3965	0.8239
PHE	0.6763	0.6283	0.6471	0.6767	0.6506
SER	0.4402	0.4402	0.2265	0.4402	0.4402
THR	0.4847	0.4847	0.4847	0.4747	0.4747
TRP	0.6264	0.553	0.6172	0.3894	0.6284
TYR	0.7086	0.7116	0.7116	0.7118	0.7272
VAL	0.5462	0.5462	0.5462	0.5462	0.5462

Table 7.1.: Relative saving for different unbound rotamer libraries according to the amino acid

In table 7.1, the average relative saving in spanning the search tree for different rotamer libraries are shown. The highest average relative saving that can be achieved is marked in yellow if not more than two libraries show the same saving. The rotamer libraries are calculated on data of unbound proteins and tested on all unbound proteins of the test set. The residues used for compilation of these libraries are the whole unbound test set (column 2), residues in all helix types (column 3), amino acids in β sheets (column 4) and representatives in random coils (column 5). As comparison, the relative saving using the Dunbrack library (as of 1999) is shown.

The relative savings which can be reached differ according to the amino acid. A large part of the search tree can be pruned for ARG, GLN, GLU, LEU, LYS and MET, whereas for SER, THR, CYS and VAL residues, a larger part of the tree has to be visited for all rotamer libraries. CYS, SER and VAL have short side chains which position is strongly influenced by the backbone conformation. Modelling of these side chains in a backbone independent approach is more difficult compared to longer side chain. For these side chains the backbone influence is reduced concerning the whole side chain because the angles in the outer part are not influenced by the backbone conformation. THR has side chain with an OH group. This group can form hydrogen bonds, which may stabilise unfavourable conformations. Therefore the prediction of the conformation is more difficult. An average relative saving for the Dunbrack library which has a difference more than 0.05 compared to the new compiled over all library can be seen for ARG, GLN, ILE, LEU, LYS and MET. For the other residues, the Dunbrack library shows a

slight increase of the average saving compared to the ALL library. Examples where less saving in spanning the search tree can be achieved are ASN and SER residues with the SHEET library (average saving of 0.29 or 0.23) and TRP side chains with the library calculated on random coil data. Because the libraries are tested on all unbound proteins (see above), the average saving using the libraries compiled on Secondary Structure data is smaller. The boxplots of the relative saving using the different rotamer libraries can be seen in figures 7.3 and 7.2.

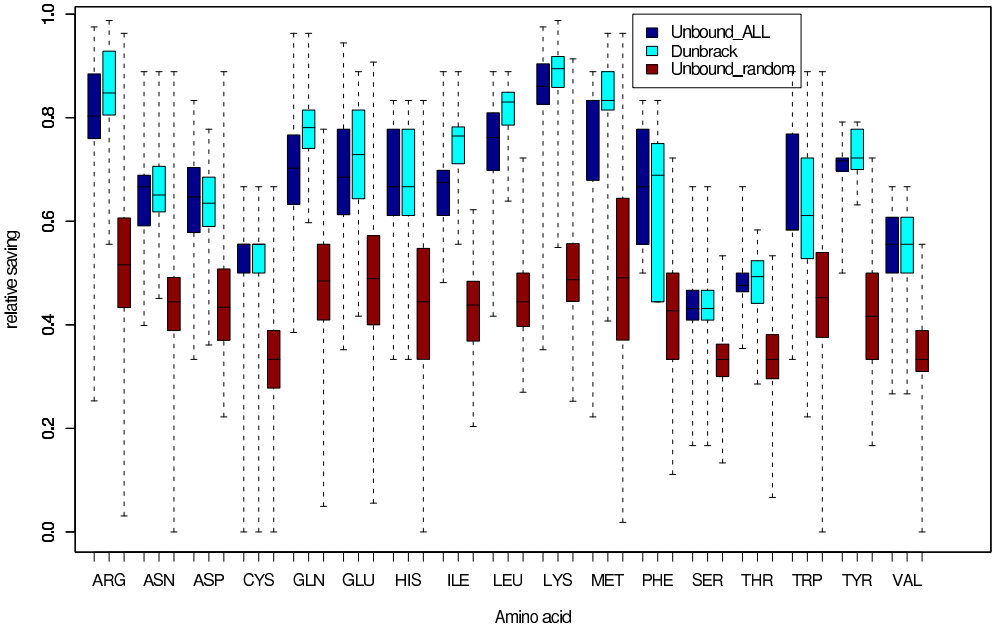


Figure 7.2.: Relative savings for rotamer libraries over all data, the Dunbrack library and a random library

In figure 7.2, the boxplot of the relative saving for the rotamer library compiled with the unbound data of the whole test set compared to the Dunbrack library and a random library can be seen. In contrast to table 7.1, the medians and not the mean values are shown in the plot. The whiskers of the boxes are extended to the extreme values, so that the minimum and maximum saving can be seen for each amino acid. The ranks assigned for the rotamer combinations of the random rotamer library shown in dark red have nothing to do with the probabilities of the rotamer combinations. They are assigned by the MYSQL random function which returns a random value between 0 and 1[1]. To calculate the random rank, the returned random value is multiplied with the maximum rank of the amino acid (number of rotamers) to avoid ranks larger than the maximum possible rank. The range of the random value multiplied with the maximum rank is from 0 to maximum rank-1. Because a zero rank is not assigned

[1] MYSQL Query for generating the random rank: *Select round(((RAND()*(max(rank)-1))+1),0) from Rotamerbibliothek_Unbound_ALL where Res_NameU="ARG"*

in rotamer libraries, one is added to this value. For compensation, one is subtracted from the maximum rank. The value is rounded as integer to gain the random rank ranging from rank 1 to the maximum possible rank of the amino acid.

The assignment of the rank according to the probability leads to higher relative saving compared to the random assignment of the rank mentioned above (cf. figure 7.2). The median for the new over all library compiled with the language model is higher for ASN, ASP and TRP. A higher maximum value of the relative saving using this library can be seen for ASP and GLU residues. A noticeable higher median for the Dunbrack library is found for GLN, ILE, LEU and MET. For the side chains of MET residues, the maximum saving is higher compared to the other rotamer libraries. For most residues, there are only slight differences of the median for both probability ranked rotamer libraries.

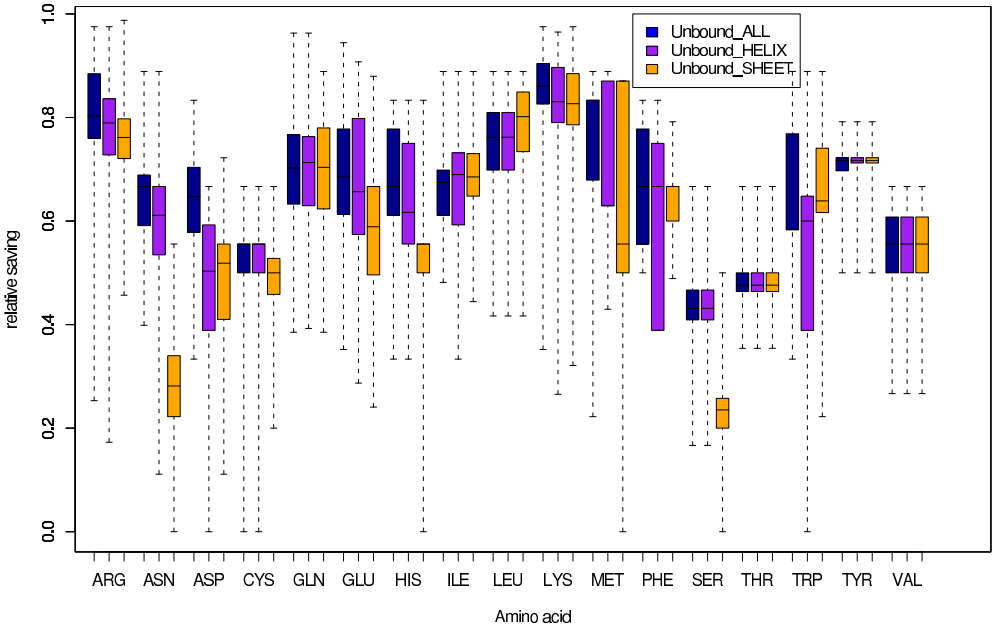


Figure 7.3.: Relative savings for the libraries calculated on residues in Secondary Structures compared to the over all library

Figure 7.3 shows the relative saving for the rotamer libraries compiled on Secondary Structure data and the over all library for comparison. Because the libraries are not only tested on residues in Secondary Structures, the highest median of the relative saving can be seen for the ALL library for many residues (ARG, ASN, ASP, GLU, HIS and LYS). For other amino acids the median of the relative saving is comparable to one of the other libraries. The highest maximum of the saving can be seen for ASP, GLU and LYS residues if the over all library is used. Using the sheet library a high maximum value can be seen for ARG residues, although the median is smaller compared to the other libraries. A low amount of pruning the search

tree can be seen for ASP and SER residues with the SHEET library. For these two residues, the probabilities calculated for sheet data differs from the probabilities for representatives in helices which leads to complete different ranks of the different conformations.

The average saving using random coil library are shown in figure 7.4. A higher saving of the random coil library can be seen for ASN, ILE and THR residues compared to the library compiled with the whole unbound data. For ILE side chain, the minimum value of the saving is smaller. For this residue, the RND rotamer library does not only show a better average result in pruning the search tree (a higher median), but the minimum is higher as well. The relative saving is low for this library when searching for TRP and MET residues. Residues which are assigned to random coils are in no Secondary Structures, so that residues belonging to this structure are a good model for many residues of unbound proteins which are in no Secondary Structure.

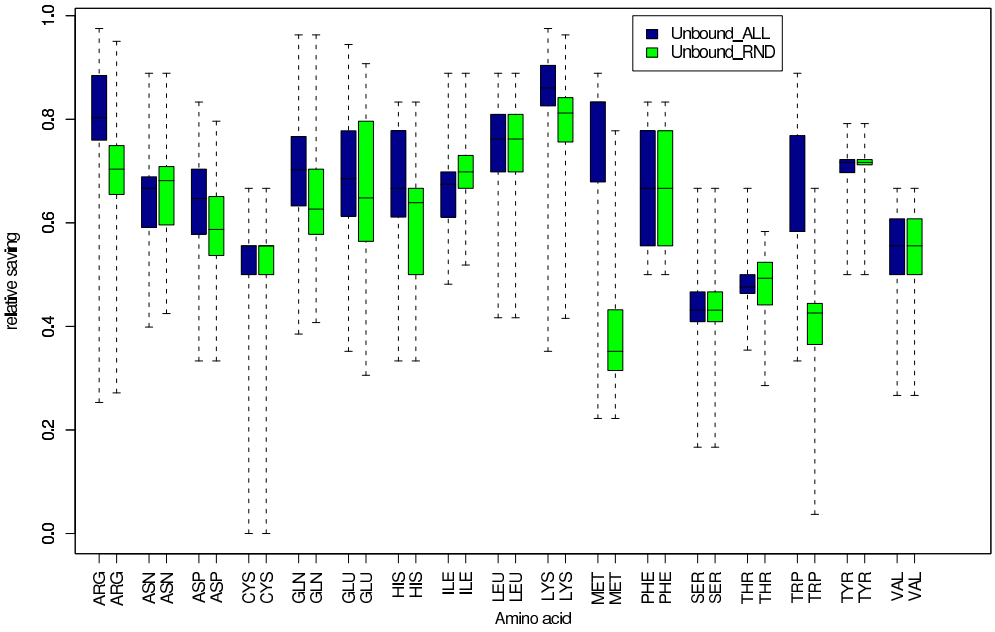


Figure 7.4.: Relative savings

The libraries are tested on residues in Secondary Structure elements to test the specificity of the different libraries for these elements.

AA	HELIX	ALL	SHEET	RND	Dunbrack
ARG	0.8263	0.8057	0.7599	0.5928	0.8487
ASN	0.7945	0.7951	0.2031	0.7139	0.7927
ASP	0.7437	0.6869	0.6630	0.3898	0.7455
CYS	0.4423	0.4423	0.3869	0.4423	0.4423
GLN	0.6234	0.6160	0.6281	0.5824	0.7029
GLU	0.7864	0.7940	0.6733	0.7012	0.7737
HIS	0.6380	0.5448	0.1864	0.4408	0.5448
ILE	0.6951	0.6542	0.7096	0.7124	0.7781
LEU	0.7506	0.7396	0.7772	0.7396	0.8014

AA	HELIX	ALL	SHEET	RND	Dunbrack
LYS	0.9216	0.9255	0.9036	0.8570	0.9421
MET	0.7585	0.7485	0.4921	0.3751	0.8469
PHE	0.7191	0.6903	0.5918	0.6920	0.7339
SER	0.4878	0.4878	0.1789	0.4878	0.4878
THR	0.6164	0.6164	0.6164	0.3603	0.3603
TRP	0.6811	0.5000	0.4327	0.4241	0.6401
TYR	0.6440	0.6228	0.6440	0.6440	0.6676
VAL	0.5725	0.5725	0.5725	0.5725	0.5725

Table 7.2.: Relative savings for different rotamer libraries for data in helices

In table 7.2, the conformation predictions of the different libraries are tested on unbound residues in helices. The average saving using the HELIX library is higher compared to the saving which can be reached on all unbound data with this rotamer library (shown in table 7.1) except for CYS, GLN, TRP and TYR. For GLU, THR and TRP residues, an increase of the average of 0.11, 0.24 or accordingly 0.13 can be noticed. These residues have longer side chains which conformations are influenced by neighbouring amino acids especially in helices. Therefore the probabilities calculated on helix data lead to the pruning of the largest part of the search tree. Although the Dunbrack library is not compiled on Secondary Structure data, it shows an average saving at least 0.05 higher compared to the HELIX library for GLN (+0.08), ILE (+0.08), LEU (+0.05) and MET(+0.09). For THR residues, the average saving can be increased by 0.26 using the HELIX library instead of the Dunbrack library. The highest saving (with an increase of the average of at least 0.05) are marked in yellow.

AA	SHEET	ALL	HELIX	RND	Dunbrack
ARG	0.8533	0.8293	0.7960	0.7315	0.8553
ASN	0.5326	0.4483	0.5088	0.4352	0.4128
ASP	0.6919	0.6541	0.7071	0.5253	0.7298
CYS	0.5803	0.5679	0.5679	0.5679	0.5679
GLN	0.7329	0.7128	0.7198	0.6018	0.7644
GLU	0.7963	0.7137	0.7623	0.6515	0.7528
HIS	0.7148	0.6139	0.4889	0.7296	0.6250
ILE	0.7941	0.7817	0.7802	0.8054	0.8259
LEU	0.7940	0.7817	0.7444	0.7462	0.8064
LYS	0.8737	0.8528	0.8291	0.7921	0.8654
MET	0.7924	0.7599	0.6289	0.4068	0.7864
PHE	0.6441	0.6180	0.5414	0.6180	0.5745
SER	0.3546	0.3121	0.3121	0.3121	0.3121
THR	0.5615	0.5615	0.5615	0.4021	0.4021
TRP	0.8377	0.6333	0.4200	0.2822	0.6845
TYR	0.7805	0.7795	0.7805	0.7805	0.7438
VAL	0.5782	0.5782	0.5782	0.5782	0.5782

Table 7.3.: Relative savings for different rotamer libraries for data in sheets

The relative saving of the different libraries in searching a given conformation for residues in sheets can be seen in table 7.3. Using the SHEET library improves the average saving in pruning the search tree for all cases sheet residues compared to the usage of all unbound data (cf. table 7.1). The highest increase is shown by ASN residues (+0.24), GLU, TYR(+0.22 both) and ASP (+0.2). For this test residues, an increase of the average saving of at least 0.05 compared to Dunbrack can be achieved using the SHEET library on HIS, PHE (both +0.08), THR (+0.16) and TRP (+0.15) residues (marked in yellow). The residues with a small saving seen in figure 7.2 show an increase on sheet data, although the prediction of SER is poor using all libraries.

AA	RND	ALLL	HELIX	SHEET	Dunbrack
ARG	0.7418	0.7087	0.6862	0.6792	0.7741
ASN	0.7163	0.6374	0.4863	0.1851	0.6422
ASP	0.7347	0.7325	0.4340	0.4914	0.6389
CYS	0.4042	0.4042	0.4042	0.3916	0.4042

AA	RND	ALLL	HELIX	SHEET	Dunbrack
GLN	0.8008	0.7976	0.7651	0.7342	0.8548
GLU	0.7268	0.6734	0.6481	0.5856	0.7174
HIS	0.7576	0.6591	0.5318	0.7379	0.6618
ILE	0.5995	0.6111	0.5746	0.5752	0.6684
LEU	0.7928	0.7928	0.7926	0.8206	0.8415
LYS	0.8817	0.8166	0.7820	0.7047	0.8654
MET	0.3333	0.6019	0.4722	0.6019	0.8056
PHE	0.7712	0.7712	0.7785	0.6506	0.7814
SER	0.4822	0.4822	0.4822	0.1844	0.4822
THR	0.5095	0.4905	0.4905	0.4905	0.5095
TRP	0.5505	0.4394	0.3586	0.4041	0.4091
TYR	0.6763	0.6763	0.6763	0.6763	0.7012
VAL	0.5861	0.5861	0.5861	0.5861	0.5861

Table 7.4.: Relative savings for different rotamer libraries for data in random coils

For residues in random coils, a higher saving compared to the usage of the RND library over the whole test set can be seen for all residues except CYS and MET. The highest increase can be seen for ASP (+0.21), TRP (+0.16), GLN, HIS (both +0.15) and ILE (+0.10) residues. Residues which show an average saving which is at least 0.05 higher than the average achieved by the Dunbrack library are ASN, ASP, HIS and TRP. An average at least 0.05 higher using the Dunbrack library can be seen for GLN, ILE and MET residues. These residues are marked in yellow.

Summarising the evaluation for unbound residues, it can be seen that the saving which can be achieved using the different rotamer libraries are dependent on the amino acids. For most residues without taking into account Secondary Structure, the Dunbrack library shows the highest saving. For residues in Secondary Structures, the highest average saving can be noticed for the libraries compiled on the Secondary Structures where the test residues belong to.

7.3. Evaluation of the rotamer libraries on complex data

The complex rotamer libraries are evaluated on the complex data of the test set. In table 7.5, the different average relative saving are shown using different complex libraries. The rotamer library with the highest relative saving are again marked in yellow, if not more than two libraries show the same amount of saving. The amino acids which show the highest saving using rotamer libraries are (as for the unbound libraries) ARG, LEU, LYS and MET. Because all complex residues from the test set are used for evaluation, the libraries compiled on Secondary Structure elements show less relative saving. The new compiled rotamer library over all data shows the best pruning of the search tree for ASN and TRP residues. The boxplots showing the saving which can be achieved by the different libraries can be seen in figures 7.5,7.6 and 7.7.

In figure 7.6, the average saving for the rotamer libraries compiled over the whole test set compared to a random library can be seen. In the random library, the ranks are not related to the probabilities of the rotamer combinations and are assigned randomly (see above). A higher amount of saving can be reached using the probability ranked libraries compared to the random library. The new compiled library yields a higher saving for ARG, ASN, ASP, GLU, HIS, MET, PHE, SER, THR , TYR and VAL residues compared to the Dunbrack library.

AA	ALL	HELIX	STRAND	RND	Dunbrack
ARG	0.8106	0.7493	0.7625	0.7485	0.8395
ASN	0.6285	0.5399	0.4816	0.6335	0.6273
ASP	0.619	0.4322	0.5234	0.5929	0.6088
CYS	0.5179	0.5179	0.512	0.5179	0.5179
GLN	0.685	0.6132	0.6841	0.5904	0.7721
GLU	0.6767	0.6445	0.6042	0.6522	0.7225
HIS	0.6816	0.6471	0.6469	0.5427	0.6825
ILE	0.6389	0.6306	0.652	0.6579	0.7372
LEU	0.7002	0.7325	0.7315	0.6794	0.7890
LYS	0.8654	0.8469	0.8438	0.8565	0.8991
MET	0.7207	0.6901	0.6841	0.3072	0.8050
PHE	0.6646	0.6417	0.6512	0.6646	0.6333
SER	0.4323	0.4323	0.3114	0.4323	0.4323
THR	0.5015	0.5015	0.5015	0.5015	0.4653
TRP	0.6595	0.6125	0.6353	0.5455	0.5971
TYR	0.6994	0.7181	0.6950	0.6941	0.7227
VAL	0.5504	0.5504	0.5504	0.5504	0.5504

Table 7.5.: Relative saving for different complexed rotamer libraries according to the amino acid

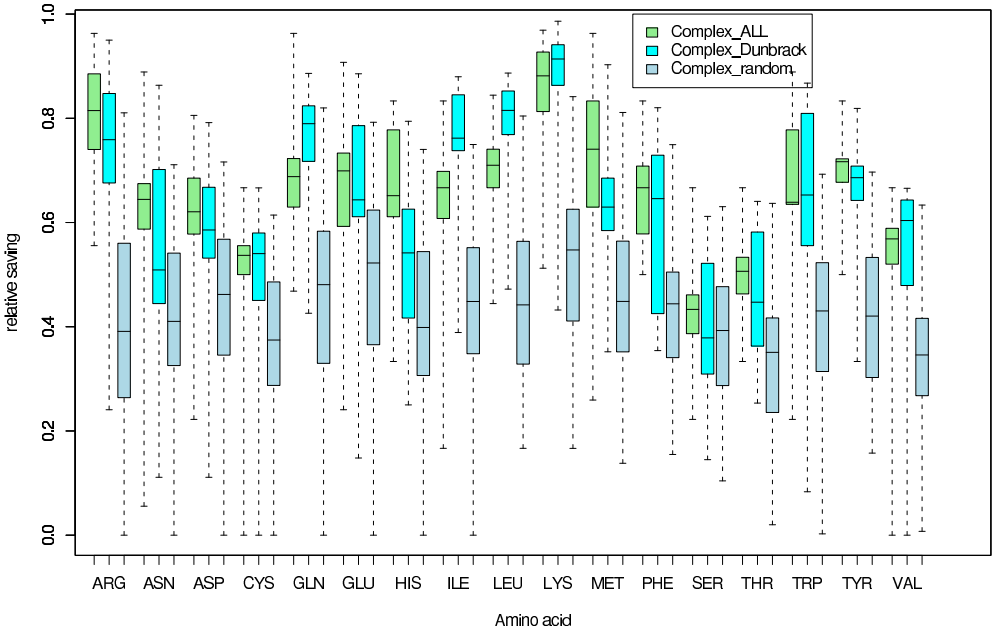


Figure 7.5.: Relative savings

The probabilities in the Dunbrack library are calculated on unbound data, not on complexes. Therefore constraints on the placement occurring in complexes are not taken into account.

The highest median and maximum values of the relative saving in figure 7.5 using the over all library can be seen for ARG, ASN, ASP, MET, PHE, SER and TYR residues. Although the Dunbrack library is not calculated on complexes, a higher median can be seen for GLN, ILE

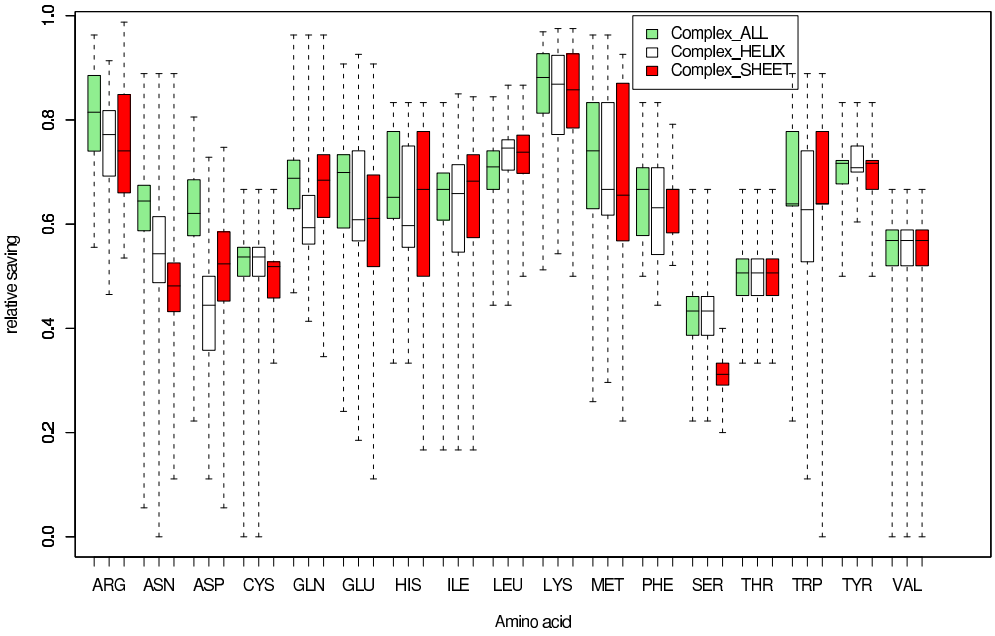


Figure 7.6.: Relative saving in searching a conformation using complex rotamer libraries compiled on different Secondary Structures

,LEU, LYS and TRP. Both libraries show a higher average saving compared to the random library whose rank is assigned randomly (see section 7.2).

The comparison of the the ALL library and the libraries compiled on Secondary Structures (cf. figure 7.6 and figure 7.7) show a higher saving using the ALL library for ARG, ASN, ASP,GLU, LYS, MET and PHE, for ARG and ASP a higher maximum value can be found as well. Using the RND library a slightly higher saving can be achieved for LYS residues.

The specificity of the different libraries towards the data they are compiled on are tested on residues in helices, sheets and random coils.

AA	HELIX	ALL	SHEET	RND	Dunbrack
ARG	0.7562	0.7386	0.6618	0.682	0.8116
ASN	0.74	0.6967	0.4889	0.6167	0.7033
ASP	0.6157	0.5069	0.5269	0.4914	0.5754
CYS	0.4868	0.4868	0.3625	0.4868	0.4868
GLN	0.7558	0.6616	0.6944	0.5321	0.7469
GLU	0.7451	0.6906	0.555	0.7067	0.7162
HIS	0.6463	0.5648	0.6315	0.2426	0.5648
ILE	0.7536	0.7387	0.7612	0.7346	0.8080
LEU	0.7239	0.685	0.7254	0.6635	0.8019
LYS	0.8519	0.8539	0.8281	0.8300	0.9063
MET	0.8349	0.8312	0.6434	0.2746	0.8851
PHE	0.7197	0.7213	0.6562	0.7213	0.7259
SER	0.4885	0.4885	0.2303	0.4885	0.4885
THR	0.6084	0.6084	0.6084	0.6084	0.3607
TRP	0.777	0.6443	0.5964	0.4097	0.6543
TYR	0.6891	0.6581	0.6367	0.6870	0.7105
VAL	0.5817	0.5817	0.5817	0.581	0.5817

Table 7.6.: Relative savings for different rotamer libraries for complex data in helices

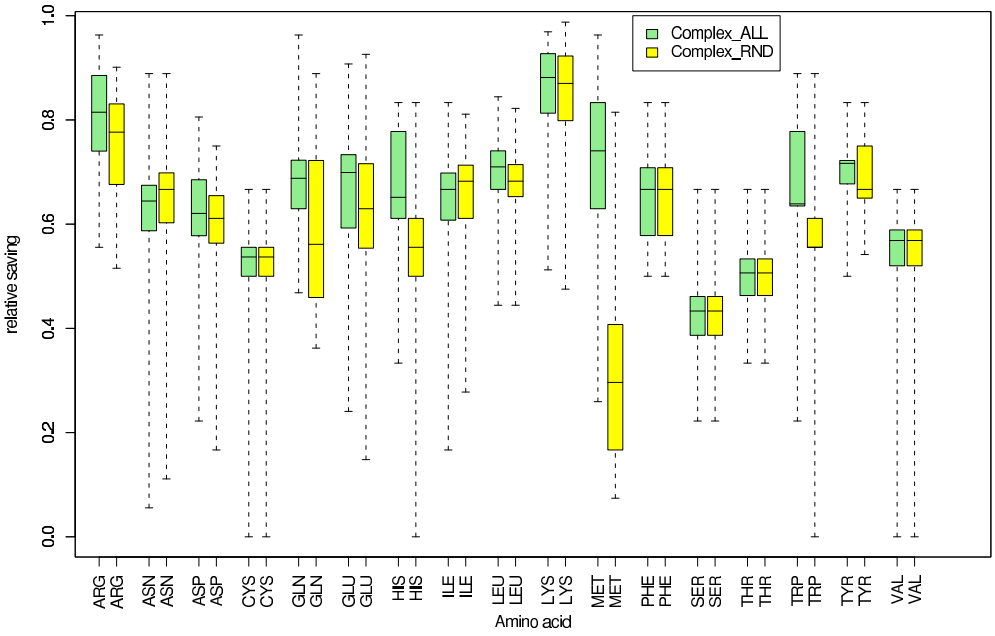


Figure 7.7.: Relative saving in searching a conformation using complex rotamer libraries compiled over all data and random coil

In table 7.6 the average saving in searching a given conformation of complex helix residues using different rotamer libraries can be seen. An increase compared to the calculated saving of the HELIX library over all complex residues (cf. table 7.5) can be noticed for all residues except for CYS, LEU and TYR residues. The highest increase of the relative saving is shown for ASN residues (+0.20), ASP residues (+0.19) and TRP residues (+0.16). The relative saving for the Dunbrack or the HELIX library where the difference of the average values is larger than 0.05 compared to the Dunbrack or HELIX library is marked in yellow. For THR residues, only 36% of the search tree could be pruned using the Dunbrack library.

AA	STRAND	ALL	HELIX	RND	Dunbrack
ARG	0.5006	0.4732	0.4455	0.4009	0.7975
ASN	0.4433	0.3686	0.2937	0.3406	0.5224
ASP	0.4406	0.4275	0.3642	0.3835	0.6862
CYS	0.5522	0.5559	0.5559	0.5559	0.6451
GLN	0.6436	0.6358	0.6102	0.5024	0.7797
GLU	0.5801	0.5227	0.5091	0.5119	0.7660
HIS	0.4432	0.3883	0.3301	0.4015	0.6406
ILE	0.5585	0.5316	0.5484	0.5346	0.7554
LEU	0.6394	0.6027	0.6358	0.5837	0.7234
LYS	0.7518	0.7277	0.7070	0.7081	0.8924
MET	0.5849	0.5436	0.4909	0.1250	0.8187
PHE	0.5681	0.5695	0.5452	0.5695	0.5990
SER	0.3234	0.2545	0.2545	0.2545	0.2690
THR	0.3996	0.3996	0.3996	0.3996	0.3753
TRP	0.6705	0.6276	0.6275	0.5301	0.7299
TYR	0.5449	0.5449	0.5507	0.5359	0.7775
VAL	0.5137	0.5137	0.5137	0.5137	0.5330

Table 7.7.: Relative savings for different rotamer libraries for complex data in sheets

The average saving for the rotamer libraries using residues in sheets can be seen in table 7.7. Compared to the test of the SHEET library on all complex data in table 7.5, the average saving

is smaller for all residues except for SER and TRP which show a slightly higher average saving using the SHEET library. The relative saving of the Dunbrack or STRAND library which show at least 0.05 more average saving compared to the other library are marked in yellow. The largest part of the search tree from most residues can be pruned using the Dunbrack library. For SER residues, the SHEET library shows the higher average saving. SER residues are difficult to predict because their side chain is small and can form hydrogen bonds which may stabilise unfavourable conformations. The saving in spanning the search tree which can be reached for SER residues is small in all libraries. Because the steric constraints in β-sheets are diverse depending on the arrangement of the strand within the sheet, the residues used for calculating this library enfolds a range of conformations and is therefore not very specific for sheet residues.

AA	RND	ALL	HELIX	SHEET	Dunbrack
ARG	0.5632	0.5363	0.5092	0.4713	0.7508
ASN	0.4906	0.4906	0.4010	0.4239	0.5480
ASP	0.5397	0.5367	0.3467	0.4847	0.6230
CYS	0.3566	0.3566	0.3566	0.3616	0.4829
GLN	0.4060	0.4279	0.4137	0.4283	0.7551
GLU	0.4954	0.4836	0.4637	0.5165	0.7004
HIS	0.5208	0.5008	0.4167	0.5268	0.6966
ILE	0.5153	0.4641	0.4465	0.4838	0.7000
LEU	0.6065	0.6131	0.6253	0.6266	0.8316
LYS	0.6680	0.6578	0.6295	0.6289	0.8722
MET	0.4290	0.2593	0.2778	0.2500	0.7531
PHE	0.5150	0.5150	0.5088	0.4365	0.7623
SER	0.3442	0.3442	0.3442	0.3316	0.4666
THR	0.3815	0.3815	0.3815	0.3815	0.4336
TRP	0.4263	0.4071	0.1635	0.3921	0.3632
TYR	0.4676	0.3994	0.4676	0.3928	0.7031
VAL	0.4287	0.4287	0.4287	0.4287	0.5434

Table 7.8.: Relative savings for different rotamer libraries for complex data in random coils

In table 7.8, the rotamer libraries are tested on residues in random coils. An increase of the average saving compared to column 4 of table 7.5 can only be seen for MET residues where the average value increases by 0.12. The highest saving (again marked in yellow) can be achieved using the Dunbrack library for most residues. For TRP residues, the average saving using the SHEET library is increased by 0.07. Because random coil residues are in no Secondary Structure, the conformations seen are diverse and not restricted by repetitive backbone ranges. Therefore the Dunbrack library which is compiled over unbound data without taking into account Secondary Structures can be taken as model for this Secondary Structure.

The specificity of the complex data towards the Secondary Structure they are compiled on is low. For most residues a higher saving can be reached using the Dunbrack library. Different proteins are included in the test set, so that same effects of the distribution which may be seen for same families are smoothed away.

8. Summary and outlook

In the thesis, amino acid side chain placement and flexibility was investigated to improve scoring functions of protein-protein docking algorithms. To gain knowledge about side chain placement which can be used in side chain demangling tasks, new rotamer libraries were compiled. The flexibility information gained upon unbound proteins and sequence identical complexes is used for a dynamisation of steric clash penalty.

In the first part of the thesis the conformations of side chains were investigated. First the rotamer probability for single χ angles was calculated. For χ_1, a trimodal distribution of χ angles has been found which reflects the geometry of the bonds connected to the C_α atom. For most residues the third rotamer is preferred due to the position of the side chain between the smallest atoms connected to C_α and opposite to the largest atom. For higher χ angles, a more bimodal distribution can be seen for ring systems or branched C-atoms due to the other geometry of the bond. For higher angels with a tetrahedral C-atoms, the distribution is trimodal with a preference for the second or third rotamer.

Afterwards the distribution of χ angle rotamers were investigated according to Secondary Structure. Because of the restricted backbone ranges in these structures, the χ_1 distribution differs. Side chains in the r1=2 rotamers are pointing outwards the helix, which led to an increase of the probability for this rotamer. The probability for r1=1 is reduced for most residues due to an unfavourable interaction of side chain atoms with the backbone. A similar interaction reduces the probability for the third rotamer of bulky amino acids. These reductions can not be seen in all cases because of the inclusion helix types other than $\alpha - helices$ and end residues which may have distorted backbones preventing unfavourable interactions. For core residues of $\alpha - helices$ ($\alpha - helices$ without the last two residues) the effects described above can be seen. For $\beta - strands$, no clear tendencies could be seen because of the different steric constraints. Same shifts in sheets were observed, e.g. a decrease in the third χ_1 rotamer for ASN, LYS, PHE, GLU, GLN, MET and VAL.

The χ distribution of residues in Secondary Structure elements was taken as first hint on the backbone dependency of the side chain conformation. For investigation of the χ_1 probabilities over the whole backbone range, the χ_1 rotamer distribution depending on the ϕ and ψ angles was calculated. It has been shown that the probability for the χ_1 rotamer varied depending on the backbone range. The most favourable third rotamer is allowed for a wider range of ϕ and ψ angles compared to the more unfavourable first rotamer. So unfavourable χ_1 rotamers are allowed only for some backbone conformations.

For side chain demangling tasks where clashing side chains are placed in more favourable rotamers, information about the probabilities for conformations has to be available. Therefore new backbone independent rotamer libraries were calculated. The independent probability was calculated using a language model, which approximates the independent probability of the

whole angle conformation with the dependent probabilities of different angle conformations. Because the χ angle distribution was shown to be dependent on the previous χ angle, this approximation can be used as model. Zero probabilities for unseen events (rotamer combinations which are not seen in a finite set of structures) are avoided by a procedure in which the probability is redistributed depending on the probability of the unseen combination without the last angle so that a probability is assigned for unseen events. The data used for the compilation of libraries are the whole test set, residues in helices, sheets or random coils for unbound PDB structures and complexes. Difference probabilities for rotamer combinations could be seen. For evaluating the different libraries, the relative saving in searching for a given conformation compared to full search is calculated. All rotamer libraries led to equal or higher savings compared to the established Dunbrack library. The highest savings could be seen for longer, less bulky amino acids like ARG, GLU and GLN. Because in the backbone independent approach the backbone conformation and the influence by neighbouring side chains was not taken into account, the libraries showed less saving for bulky amino acids with ring systems or for residues with just one χ angle where the influence of the backbone conformation is higher. The specificities of the different libraries was tested with the residues of test proteins in Secondary Structures. The library compiled on special Secondary Structure element did not necessarily led to the highest saving on this data type. Especially the random coil library (which is compiled on residues in no Secondary Structure) shows higher saving for some residues in helices or sheets than the original helix or sheet library. For strands, the specificity was higher compared to the helix library. Here the diversity of steric constraints within these structure leads to a broadening of the distribution, so that the random coil library showed higher savings for some residues in sheets.

For a dynamisation of the penalisation of steric clashes (see above) information about the flexibility of a residues upon complex formation not given in static rotamer models is needed. The flexibility of residues where given as probability of rotamer changes. First this probability for single χ angles and for combinations of χ angles were calculated. An increased flexibility for higher χ angles has been noticed because a rotation in these angles is less restricted by the backbone atoms. The flexibility of different amino acids are given as flexibility scale. Flexible residues are ARG, SER and GLN residues. It was shown that the flexibility of the residues is influenced by the structure of the residue (branches, length) and the polarity of the side chain (charges, polar groups). Long, not bulky side chains have a higher flexibility compared to shorter ones because they often have to move away upon complex formation due to steric clashes and they are less bulky so that rotation is not restricted. Branches influence the flexibility as well: branched side chains are more voluminous so that a branch hinders rotation. Especially branched side chains with charges (ASP, GLU) showed a reduced flexibility, because the charge at the end of the side chain has to be neutralized in the inner part of the protein to avoid destabilisation or the branch is involved in electrostatic interactions and therefore non-flexible. For branches with polar groups (GLN, ASN) a higher flexibility compared to charges is allowed. In the over all data set some flexibility which would not be seen in the core of the protein are observed (e.g. a rotation of charged side chain), the flexibility was calculated for exposed residues. Because of the less steric hindrance exposed residues were shown to be more flexible. In contrast the exposed residues, it could be shown that the flexibility of exposed interface residues is reduced for some angles while it is higher for other angles. For active sites, the specificity of ligand binding must be guaranteed. Another factor which was shown to influence the flexibility is the rotamericity of the residue. For side chains with a larger distance to the

median of the rotamer a higher flexibility could be observed due to the higher energy niveau these side chains are in. The flexibility measurements are stored in a database backend, so that docking algorithm could access the data easily in a preprocessing phase.

8.1. outlook

Although flexibility parameters for different environments were calculated and preliminary docking results show an improvement for same cases, the prediction if a side chain is flexible can be improved by taking into account more than one factor at a time. For example the flexibility is not only dependent on the χ angle, but also on the rotamer of the χ angle itself. Side chains in unfavourable rotamers are shown to move away more often compared to side chains in energetically favoured rotamers. A dependency of the flexibility on the backbone conformation has been observed as well. To integrate all factors which influence the flexibility and calculate their weight a Bayesian network can be used. Bayesian networks are expert system derived from causal networks. The random variables are connected by acyclic graphs. The state of the random variable A can be modelled in dependency of other random variables. If variable A is not influenced by parent variables, the states of the variable get the absolute probabilities. If the variable has parents, the dependent probabilities of the different states for variable A given the state of variable B is chosen as probability. If some information about the state of one random variable is known, they may become independent. With a Baysian network, the probability for a rotamer change depending on the amino acid, rotamericity, Secondary Structure, SAS area, interface membership and the rotamer set of the side chain can be modelled with the calculated probabilities. Expert knowledge like reaction mechanism or knowledge about the behaviour of special important amino acids can be modelled in the networks as well.

The probability for rotamer changes could be used to improve the ranking of hypotheses calculated by docking algorithms. If a side chain is assigned as flexible, it is assumed to move away if a steric clash occurs. Therefore the penalty for the steric clash of this residue can be lowered compared to a steric clash for an inflexible residue. In side chain demangling tasks, side chains with steric clashes are placed in more favourable rotamers. For this placement, the probability calculated in the rotamer library can be used and the next favourable conformation without steric clashes can be chosen, or the side chains could be placed according to the preferred directions of movements.

During the statistical investigation some special conformational features or special flexibilities of important amino acids in the protein families may be missed because of the diversity of the test set. The test set used contains different kinds of bound and unbound proteins like enzymes and their inhibitors, structural proteins, viral proteins, receptor proteins and antibody-antigen complexes and their unbound forms (see EC numbers in section 6.1.1). Further improvement of the conformation and flexibility prediction may be achieved by the classification of the test data in protein families. With different parameter sets calculated on different protein families, the specificity of the prediction may be improved. Enough data for these calculations will be available because of the exponential growing of the available structures in the PDB.

A. Side chain structures

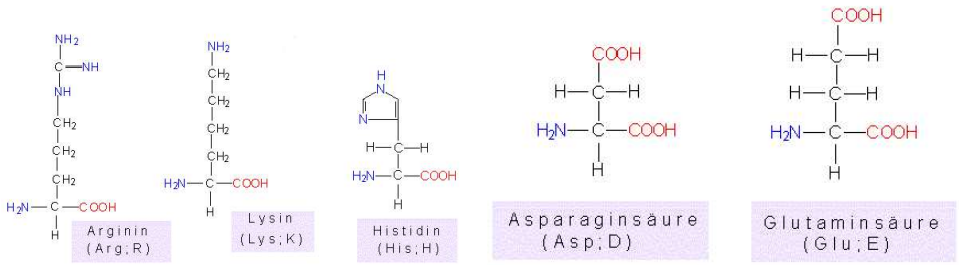


Figure A.1.: Side chain structures of basic and charged aa, from [47]

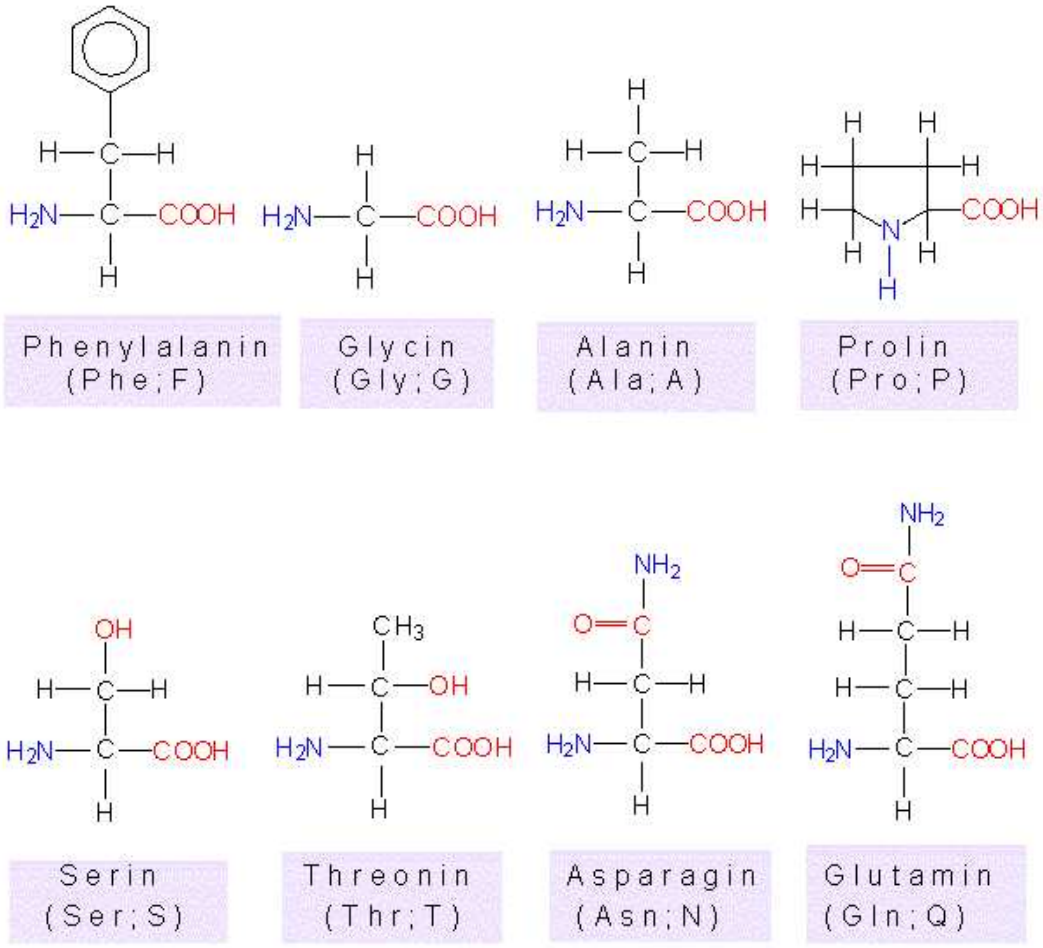


Figure A.2.: Side chain structures of aromatic and hydrophobic aa, from [47]

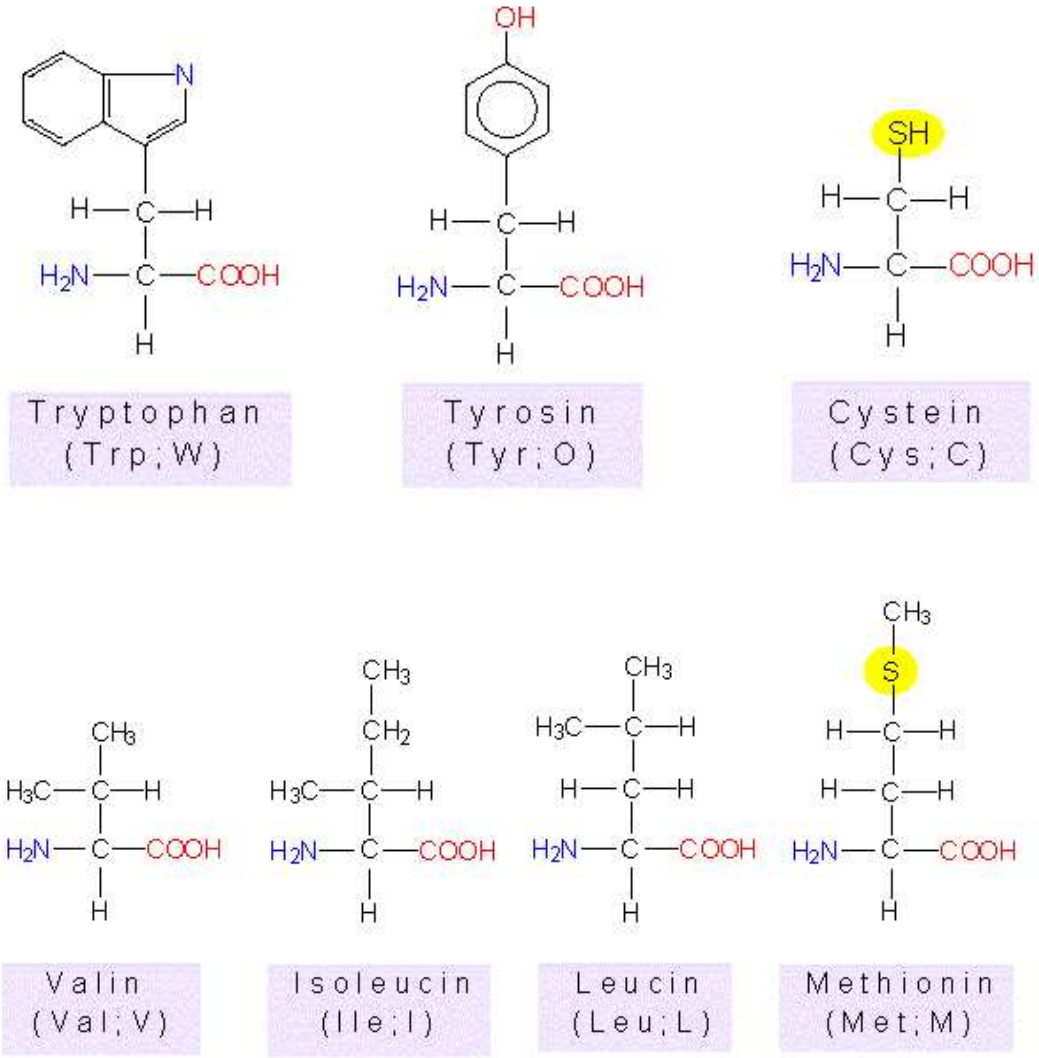


Figure A.3.: Side chain structures of hydrophobic, aromatic and polar aa, from [47]

B. Histogramms

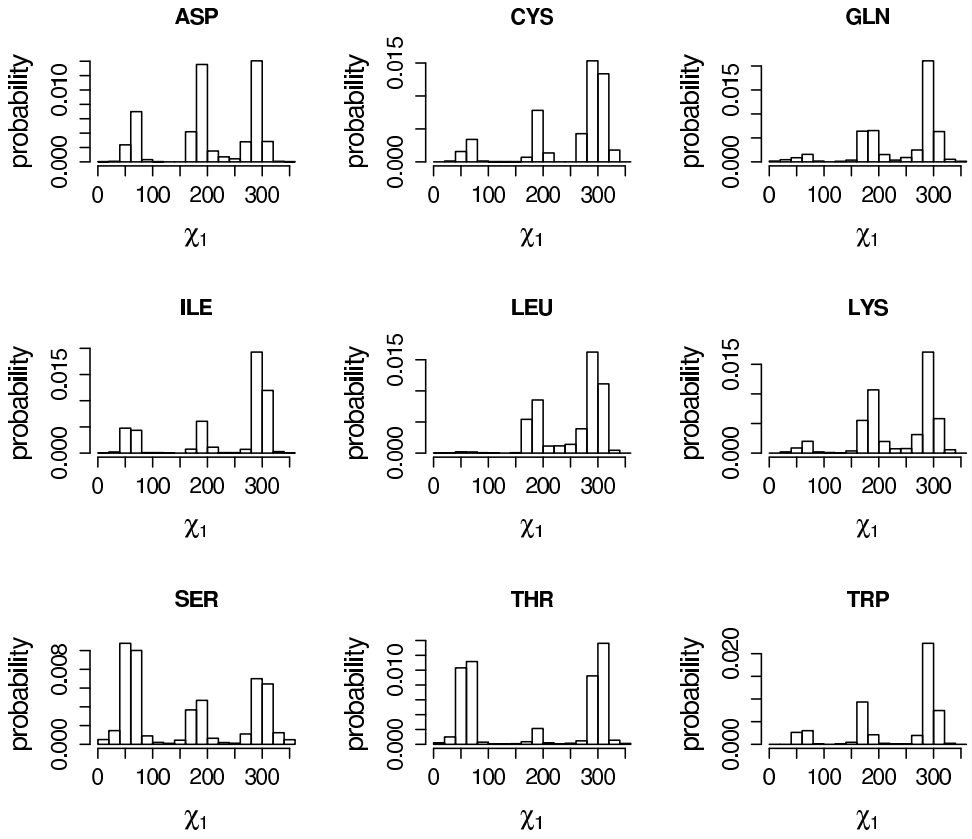


Figure B.1.: χ_1 rotamer Distribution for all residues (Part I)

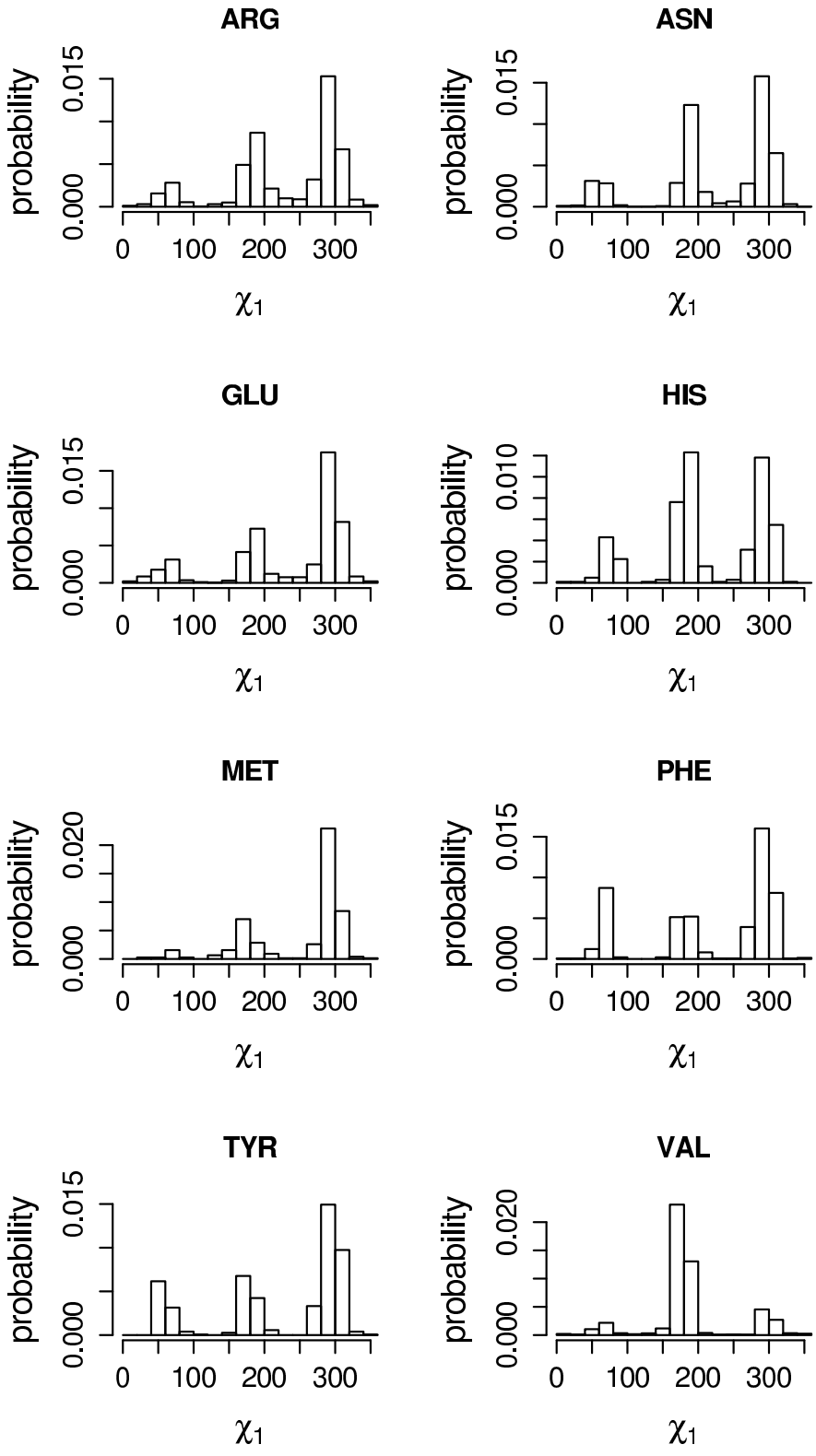


Figure B.2.: χ_1 rotamer Distribution for all residues (Part II)

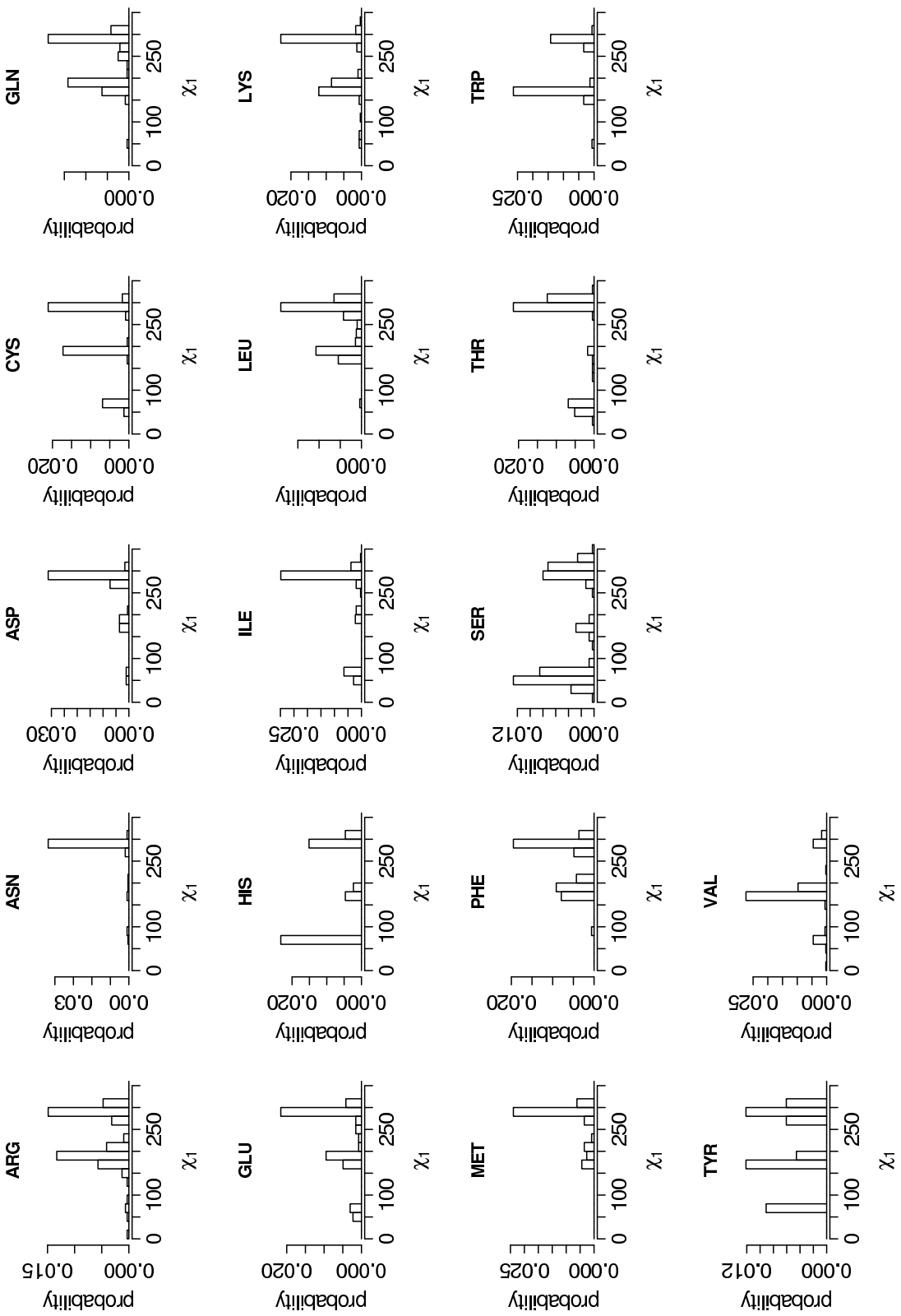


Figure B.3.: χ_1 rotamer Distribution for all helix types

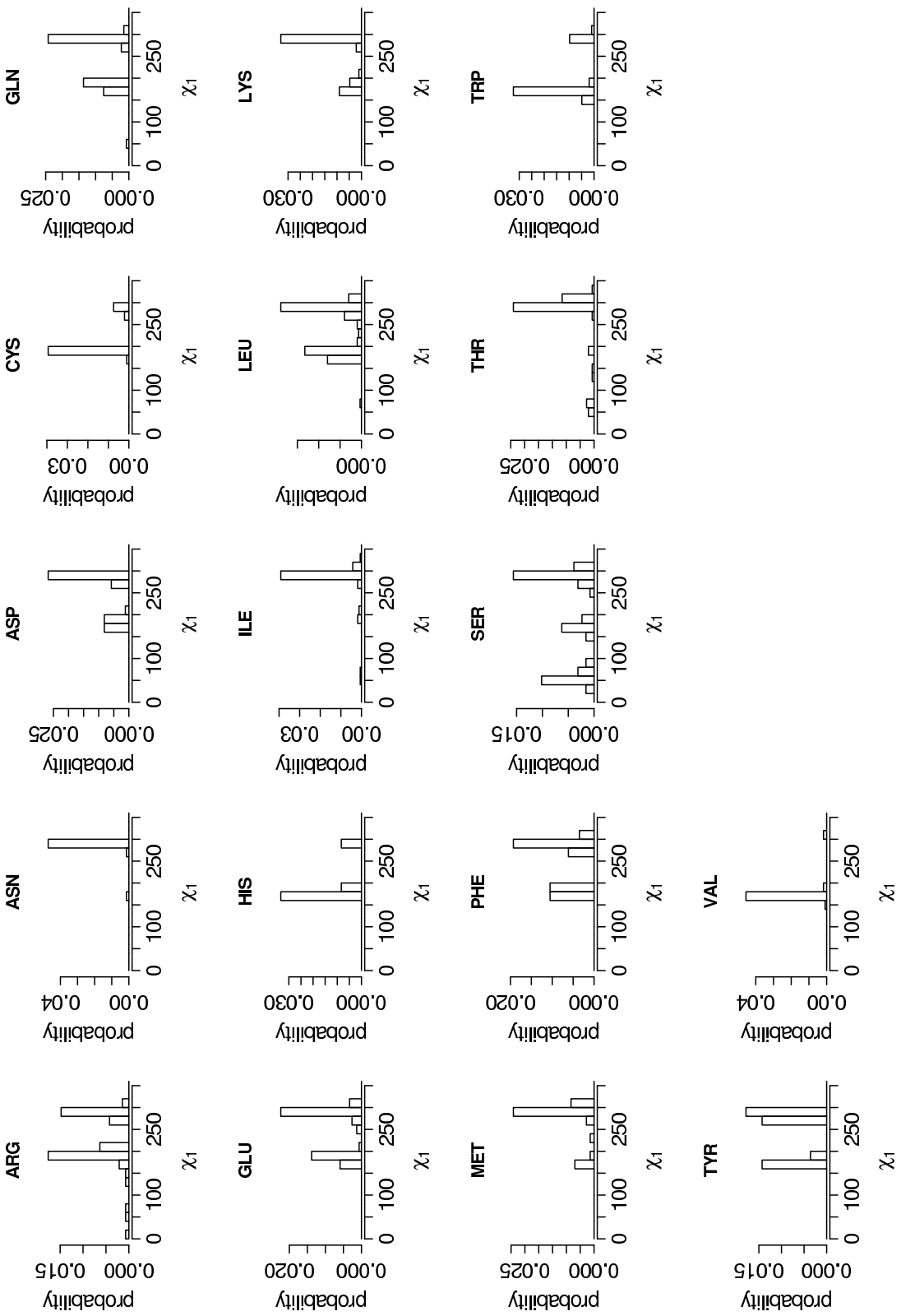


Figure B.4.: χ_1 rotamer Distribution for α-Helices

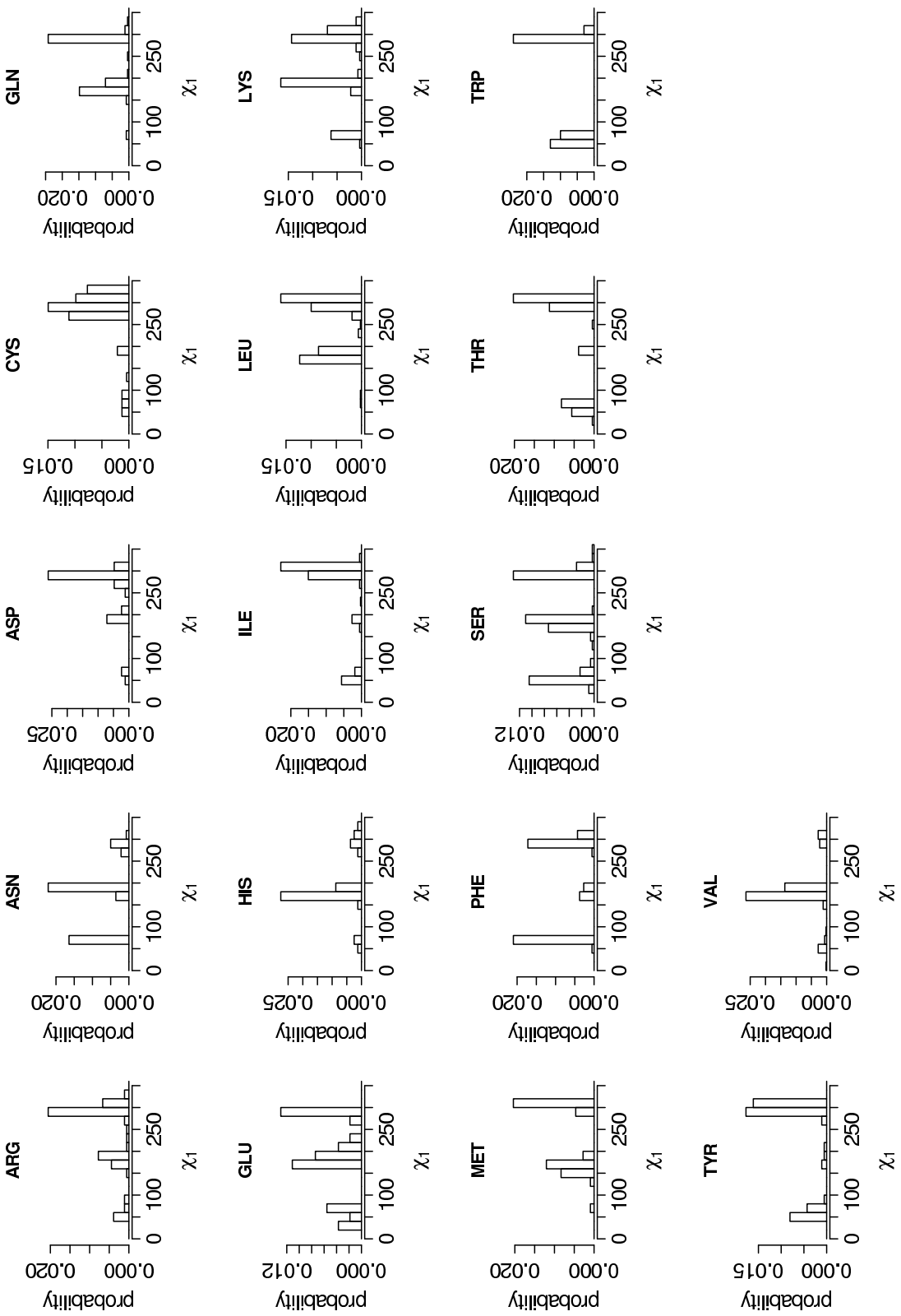


Figure B.5.: χ_1 rotamer Distribution for β-Sheets

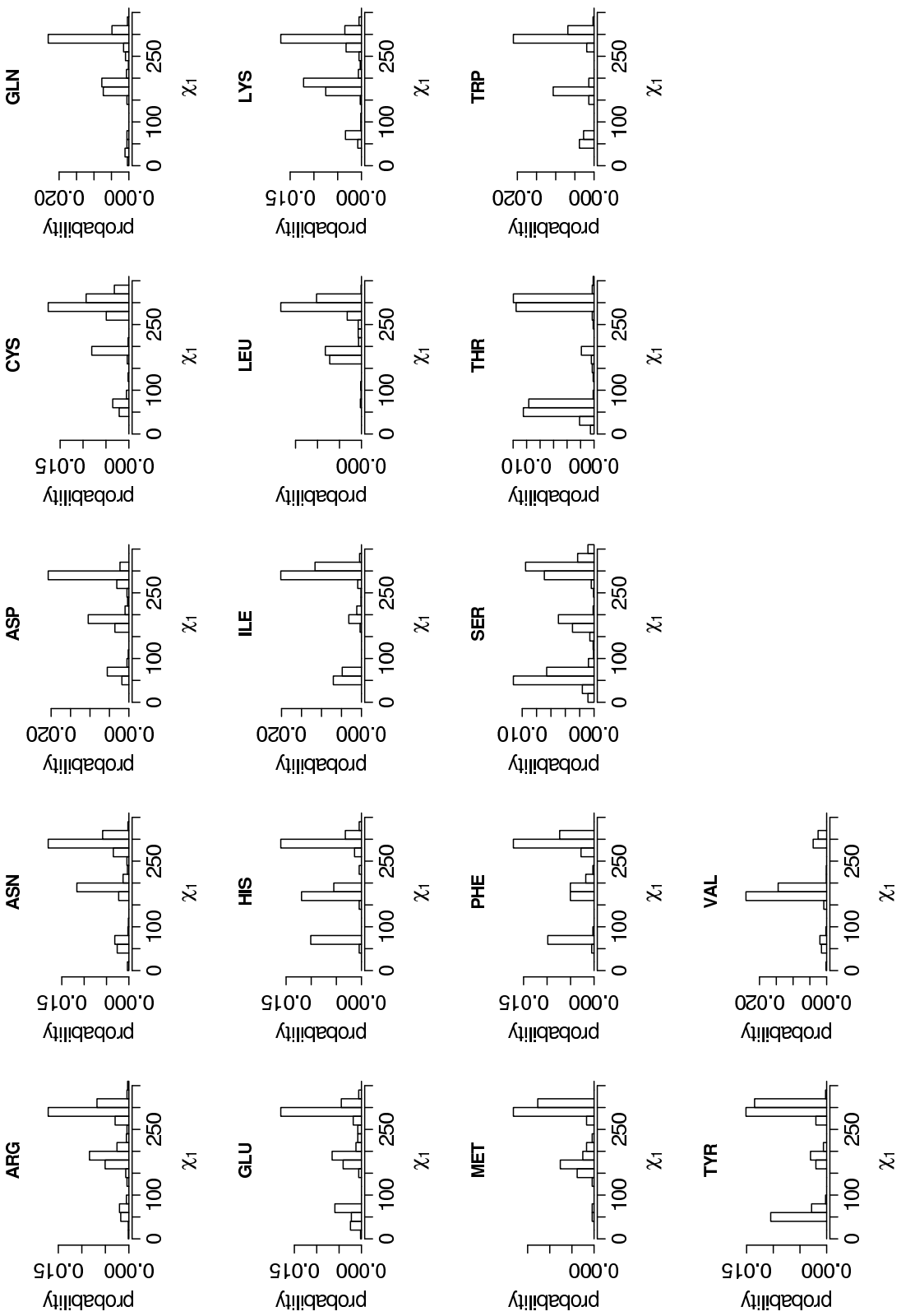


Figure B.6.: χ_1 rotamer Distribution for random coils

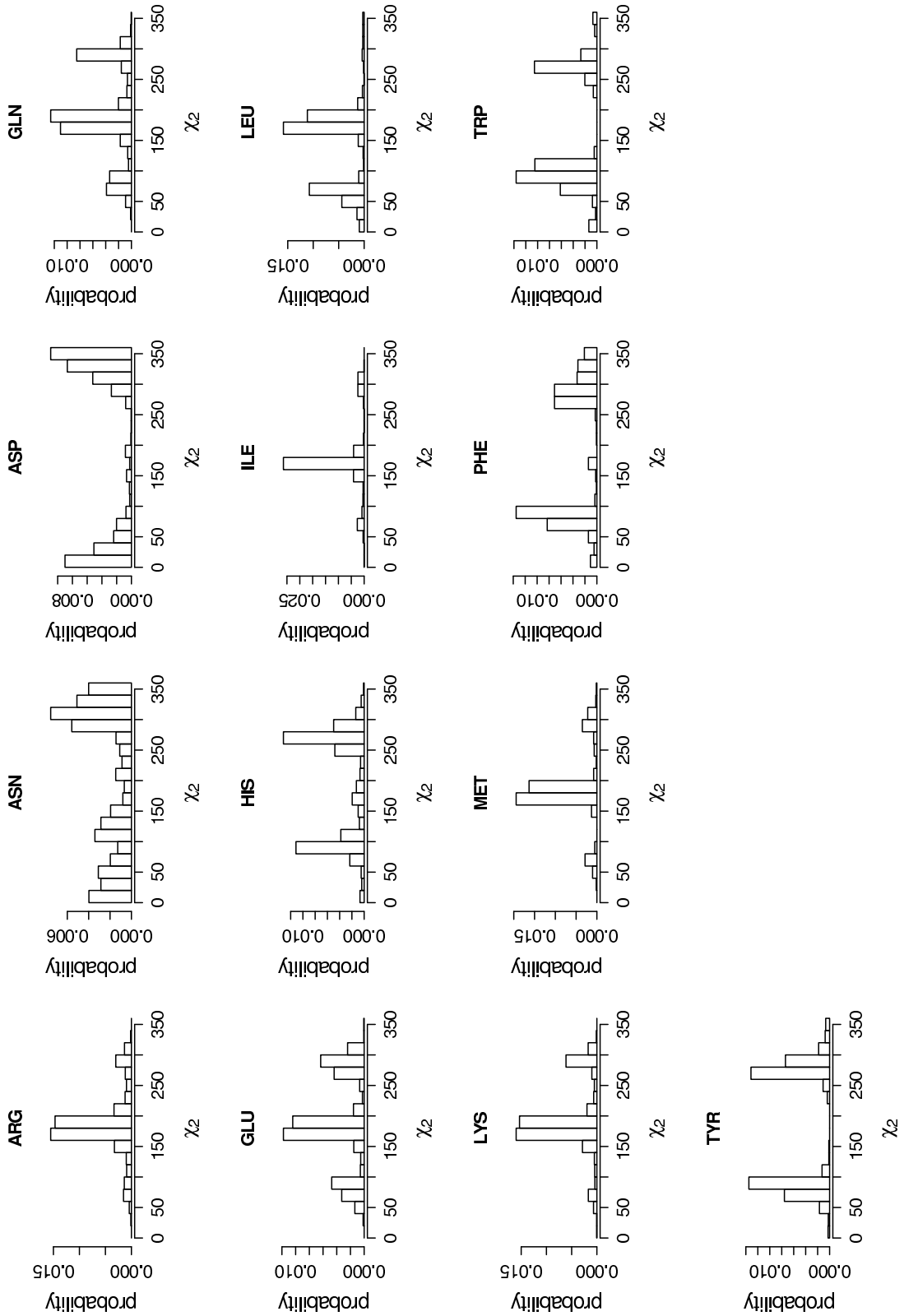


Figure B.7.: χ_2 rotamer Distribution for all residues

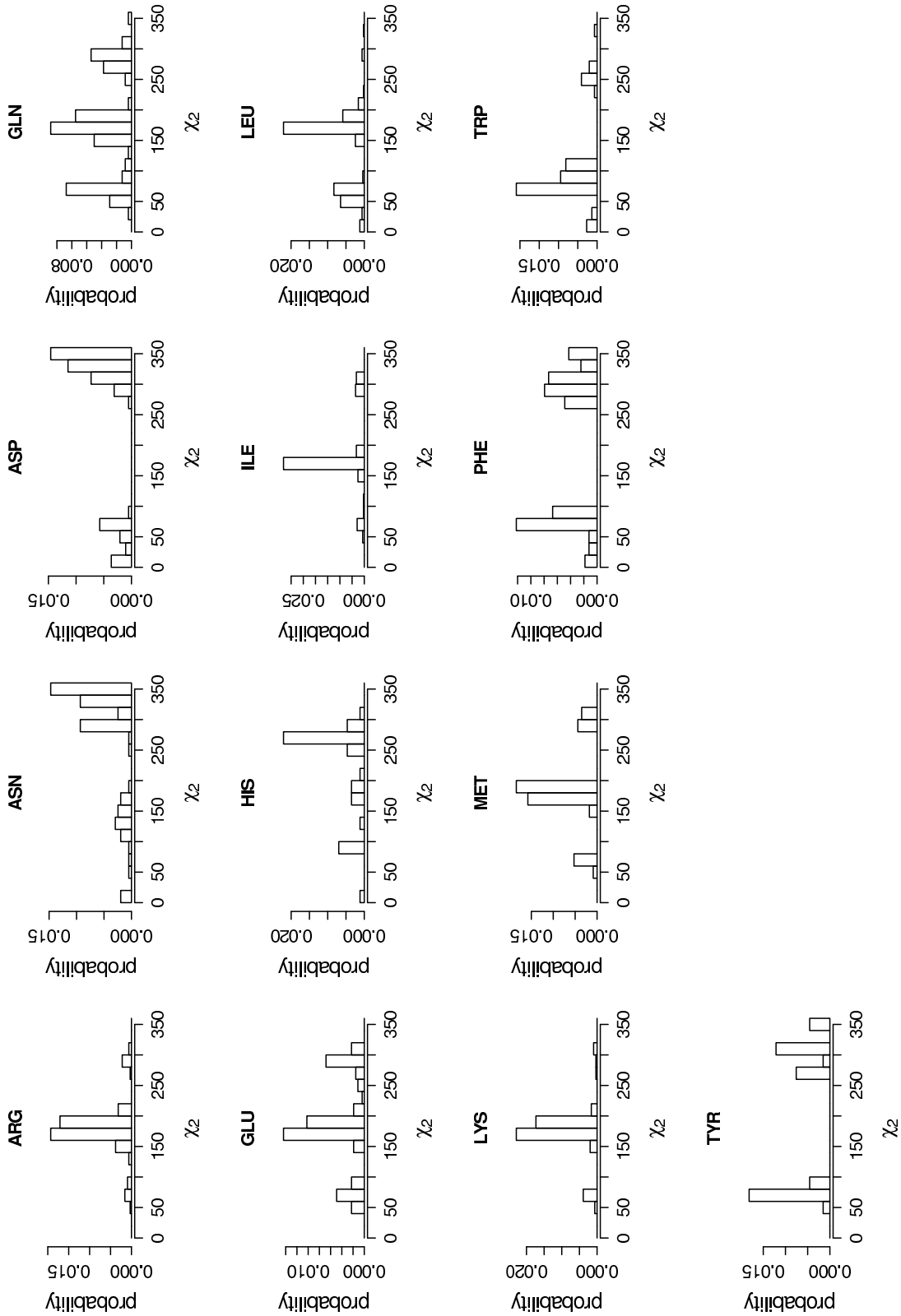


Figure B.8.: χ_2 rotamer Distribution for all helix types

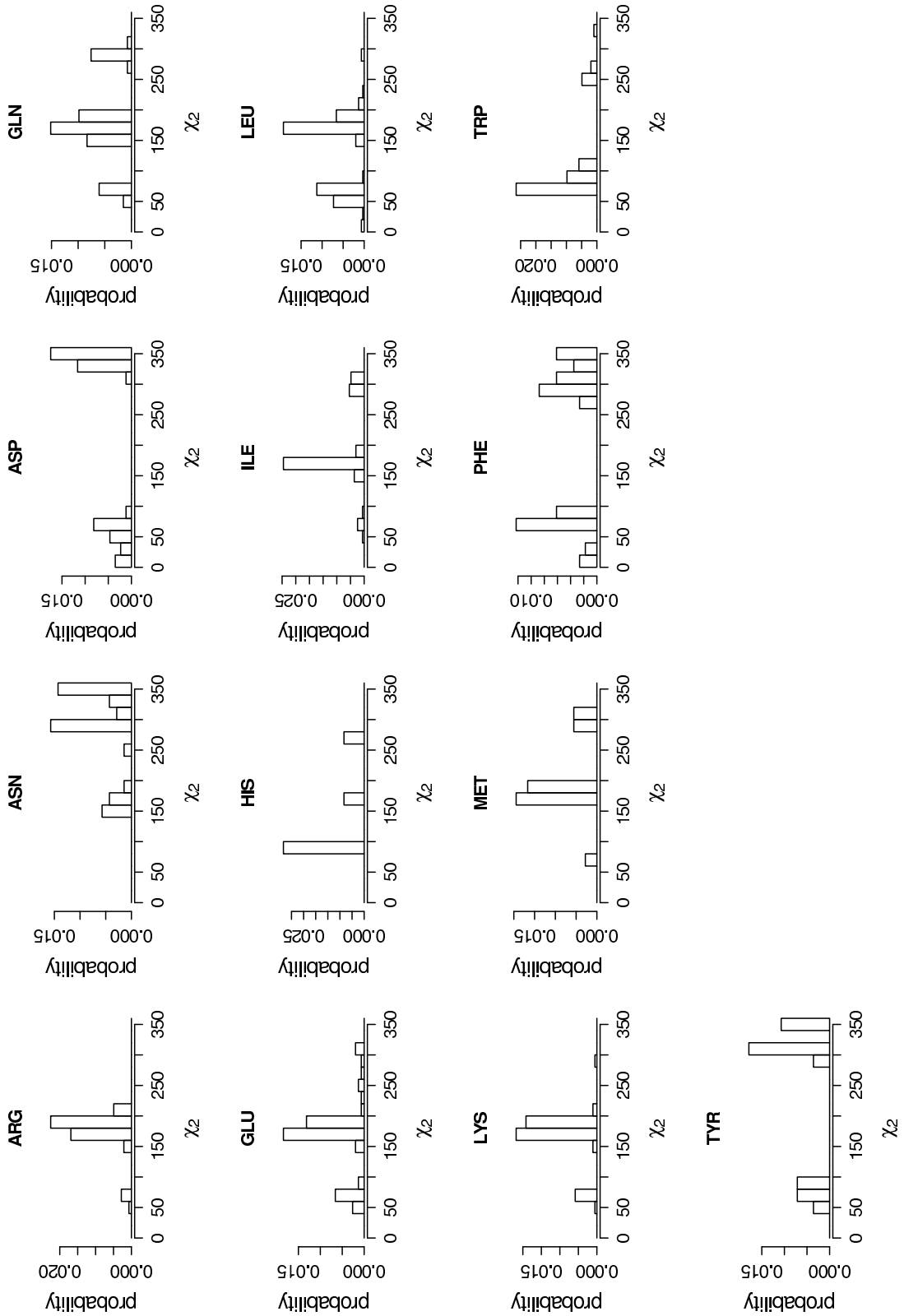


Figure B.9.: χ_2 rotamer Distribution for α-Helices

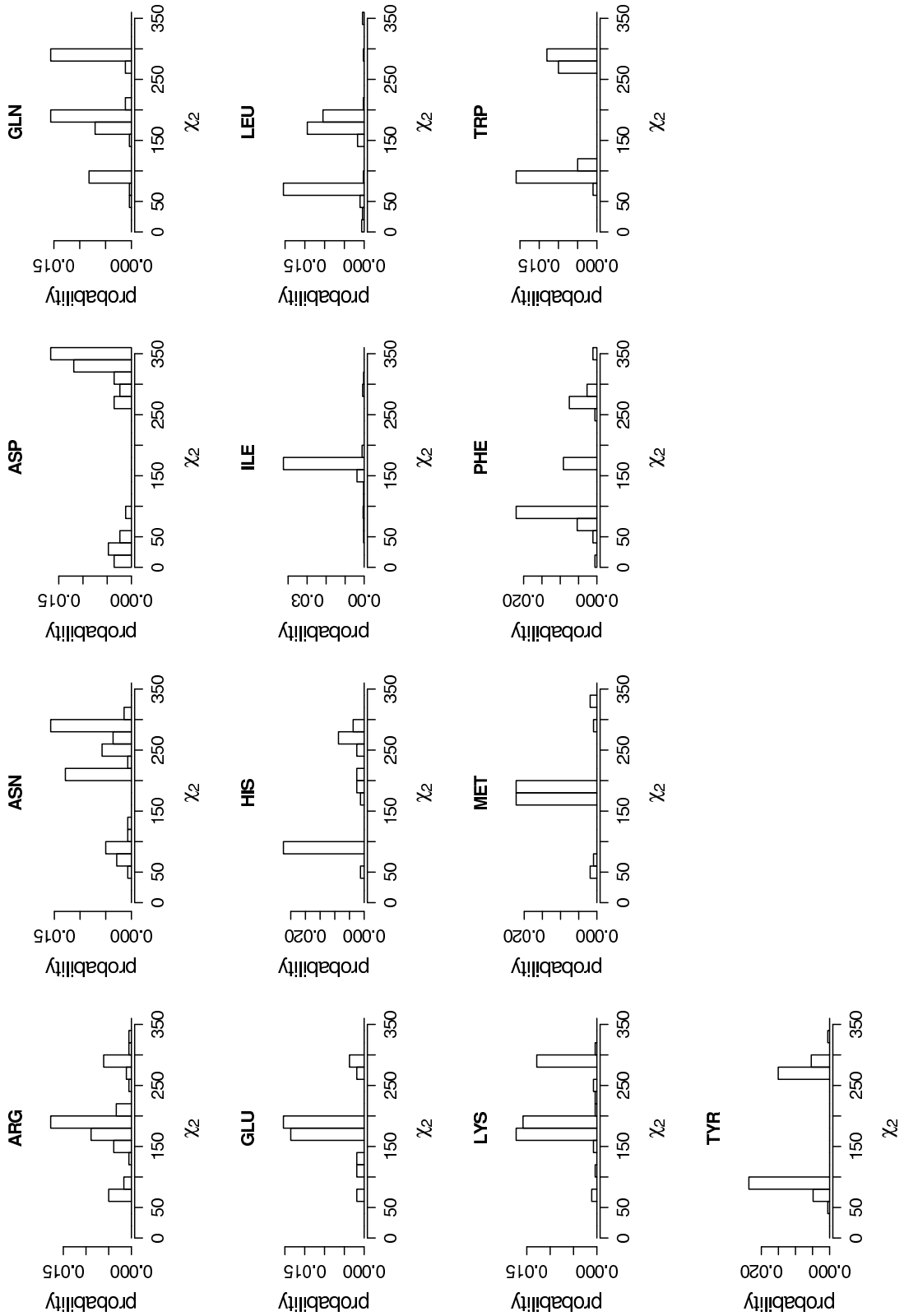


Figure B.10.: χ_2 rotamer Distribution for β-Sheets

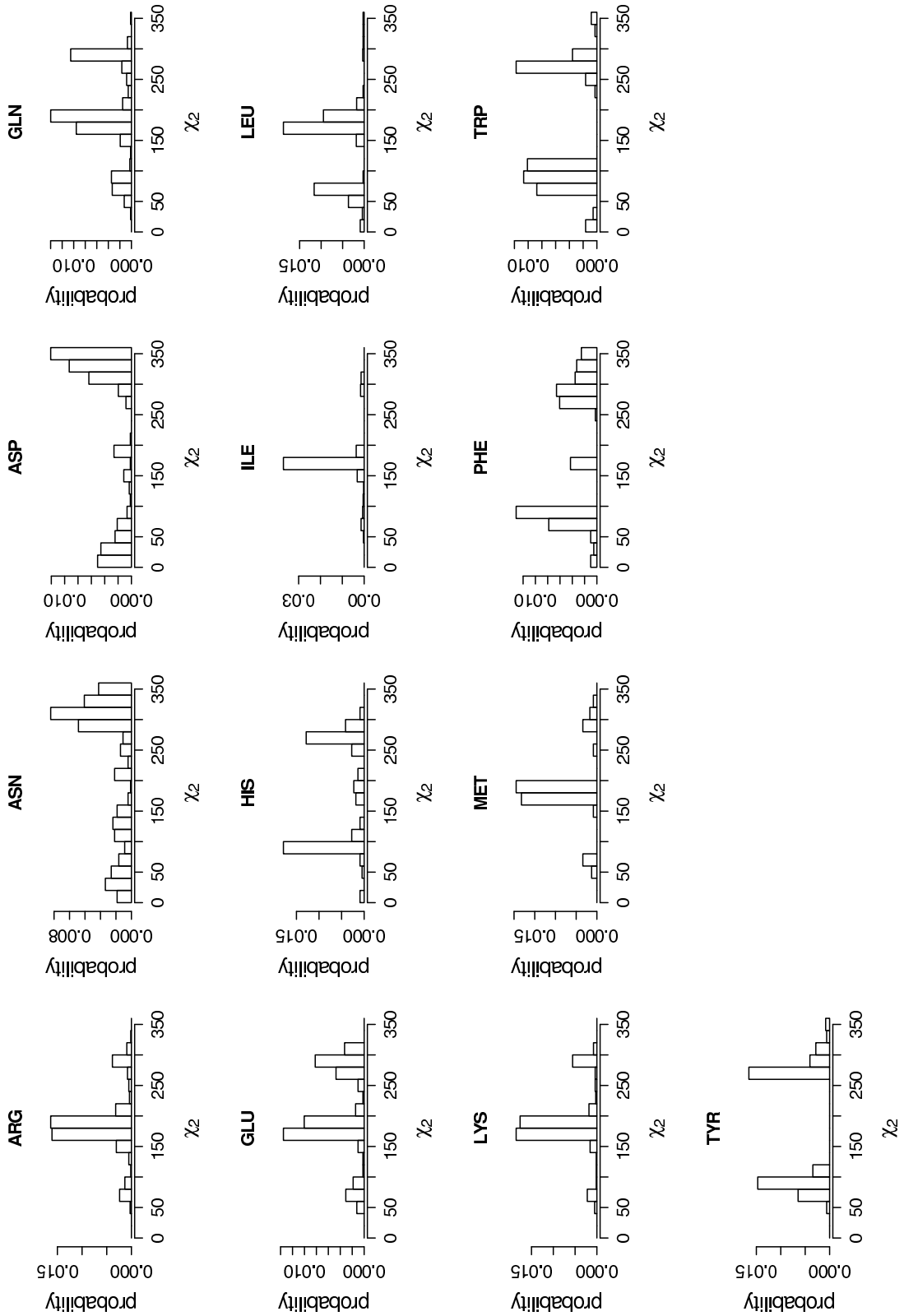


Figure B.11.: χ_2 rotamer Distribution for random coils

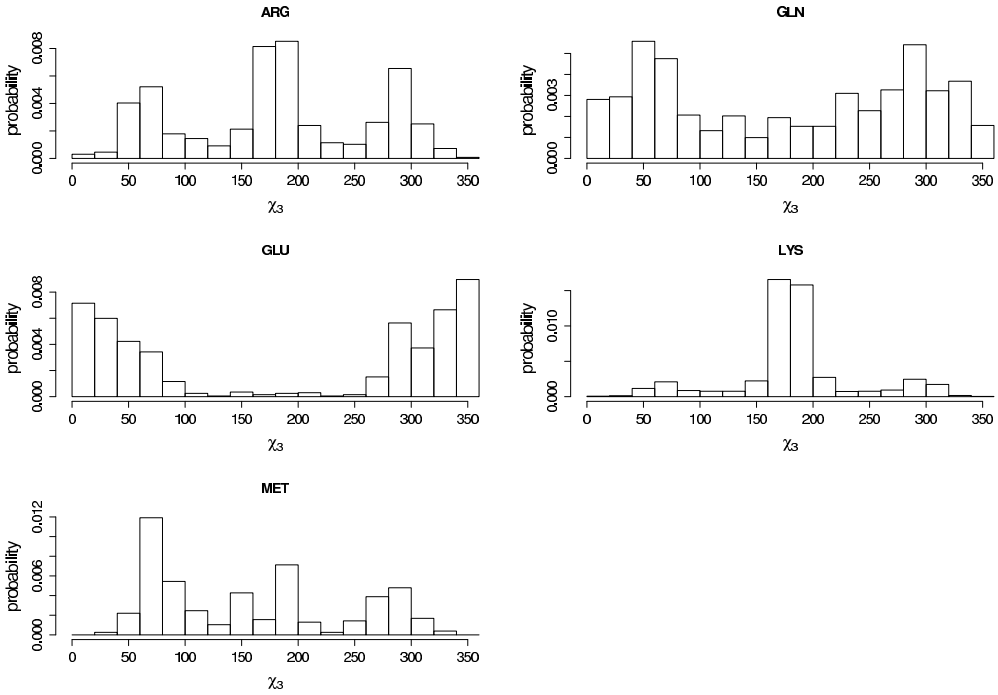


Figure B.12.: χ_3 rotamer Distribution for all residues

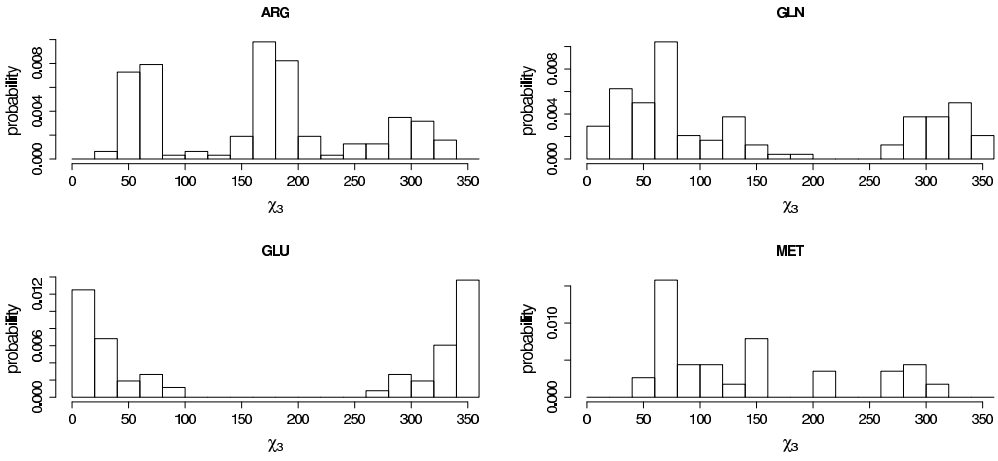


Figure B.13.: χ_3 rotamer Distribution for all helix types

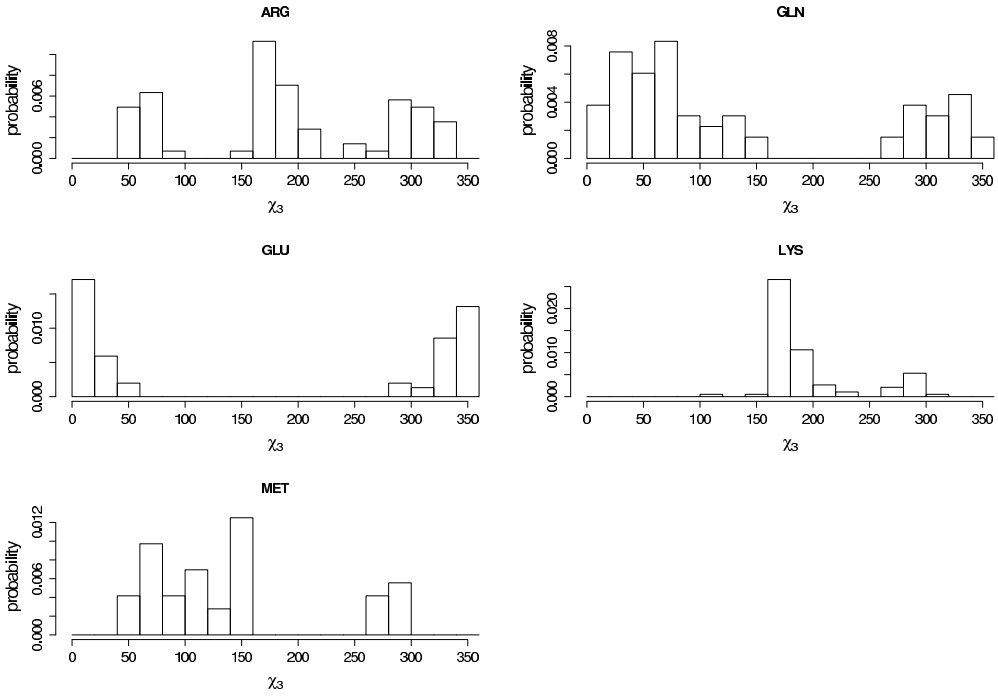


Figure B.14.: χ_3 rotamer Distribution for α-Helices

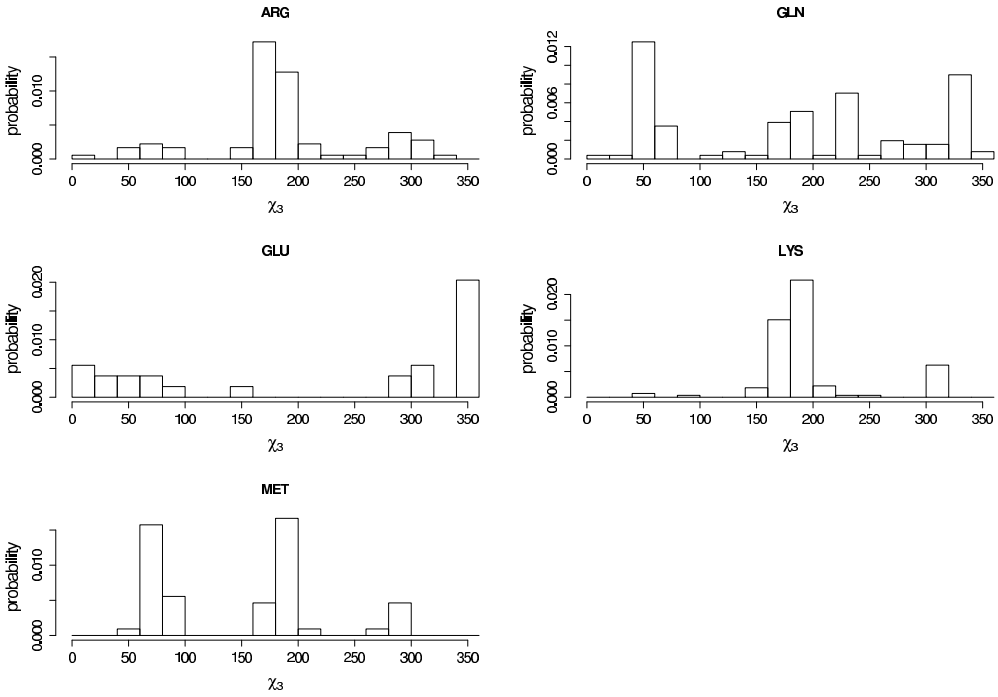


Figure B.15.: χ_3 rotamer Distribution for β-Sheets

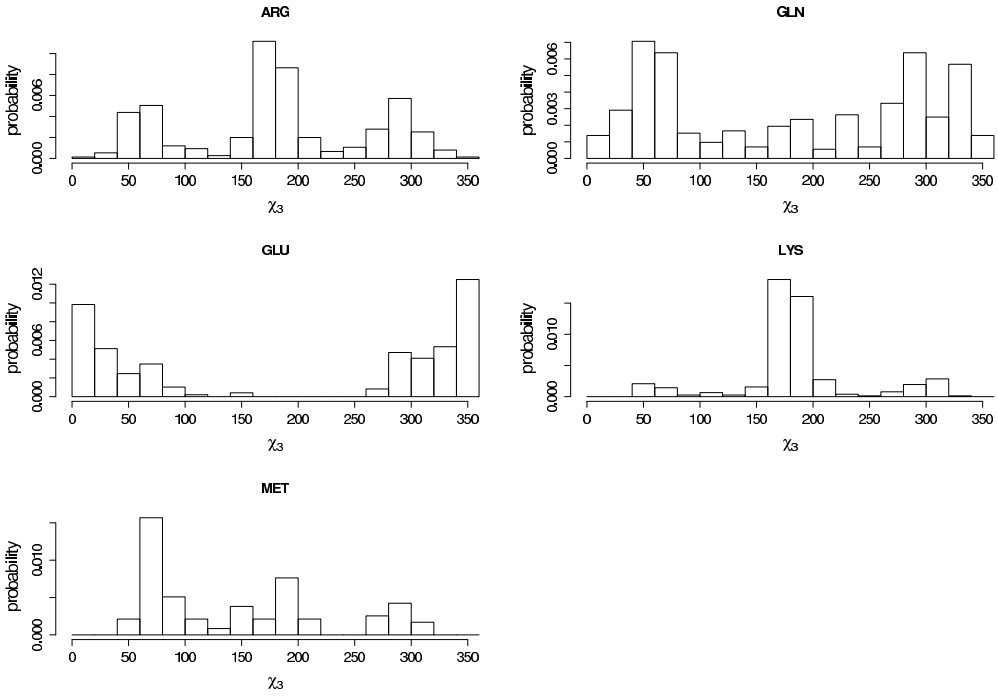


Figure B.16.: χ_3 rotamer Distribution for random coils

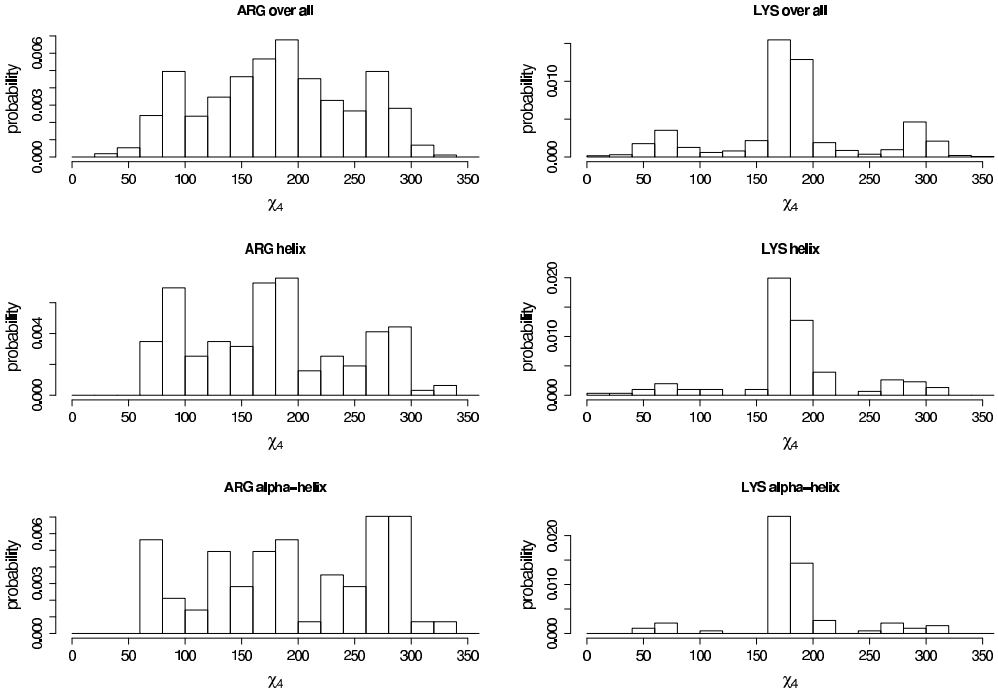


Figure B.17.: χ_4 rotamer Distribution for all residues, helices and alpha helices

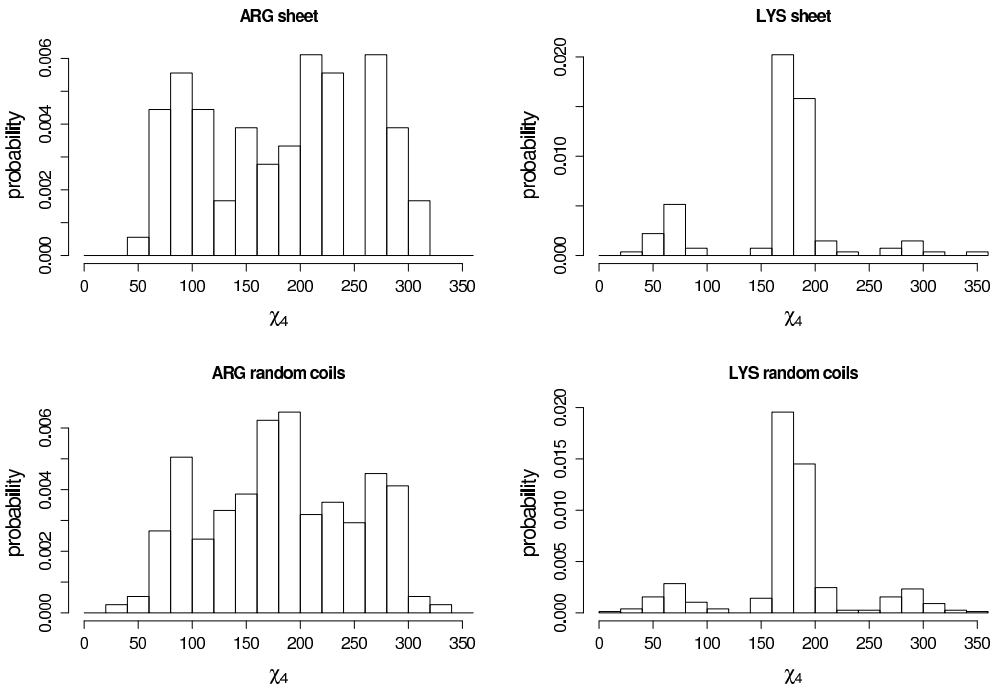


Figure B.18.: χ_4 rotamer Distribution for sheets and random coils

C. Rotamer Libraries

AName	r1	r2	r3	r4	avg Chi1	sig Chi1	avg Chi2	sig Chi2	agv Chi3	sig Chi3	avg Chi4	sig Chi4	pr234iAr1
ARG	1	1	1	1	62.854	15.831	81.463	19.742	68.406	21.862	86.493	17.299	0.0008
ARG	1	1	1	2	62.854	15.831	81.463	19.742	68.406	21.862	179.906	30.856	0.0042
ARG	1	1	1	3	62.854	15.831	81.463	19.742	68.406	21.862	273.157	18.412	0.0011
ARG	1	1	2	1	62.854	15.831	81.463	19.742	181.277	20.855	86.493	17.299	0.0057
ARG	1	1	2	2	62.854	15.831	81.463	19.742	181.277	20.855	179.906	30.856	0.0097
ARG	1	1	2	3	62.854	15.831	81.463	19.742	181.277	20.855	273.157	18.412	0.0048
ARG	1	1	3	1	62.854	15.831	81.463	19.742	289.018	18.960	86.493	17.299	0.0004
ARG	1	1	3	2	62.854	15.831	81.463	19.742	289.018	18.960	179.906	30.856	0.0089
ARG	1	1	3	3	62.854	15.831	81.463	19.742	289.018	18.960	273.157	18.412	0.0034
ARG	1	2	1	1	62.854	15.831	179.967	18.559	68.406	21.862	86.493	17.299	0.0040
ARG	1	2	1	2	62.854	15.831	179.967	18.559	68.406	21.862	179.906	30.856	0.0101
ARG	1	2	1	3	62.854	15.831	179.967	18.559	68.406	21.862	273.157	18.412	0.0030
ARG	1	2	2	1	62.854	15.831	179.967	18.559	181.277	20.855	86.493	17.299	0.0091
ARG	1	2	2	2	62.854	15.831	179.967	18.559	181.277	20.855	179.906	30.856	0.0256
ARG	1	2	2	3	62.854	15.831	179.967	18.559	181.277	20.855	273.157	18.412	0.0137
ARG	1	2	3	1	62.854	15.831	179.967	18.559	289.018	18.960	86.493	17.299	0.0015
ARG	1	2	3	2	62.854	15.831	179.967	18.559	289.018	18.960	179.906	30.856	0.0175
ARG	1	2	3	3	62.854	15.831	179.967	18.559	289.018	18.960	273.157	18.412	0.0067
ARG	1	3	1	1	62.854	15.831	285.634	18.865	68.406	21.862	86.493	17.299	0.0004
ARG	1	3	1	2	62.854	15.831	285.634	18.865	68.406	21.862	179.906	30.856	0.0023
ARG	1	3	1	3	62.854	15.831	285.634	18.865	68.406	21.862	273.157	18.412	0.0006
ARG	1	3	2	1	62.854	15.831	285.634	18.865	181.277	20.855	86.493	17.299	0.0055
ARG	1	3	2	2	62.854	15.831	285.634	18.865	181.277	20.855	179.906	30.856	0.0109
ARG	1	3	2	3	62.854	15.831	285.634	18.865	181.277	20.855	273.157	18.412	0.0063
ARG	1	3	3	1	62.854	15.831	285.634	18.865	289.018	18.960	86.493	17.299	0.0013
ARG	1	3	3	2	62.854	15.831	285.634	18.865	289.018	18.960	179.906	30.856	0.0076
ARG	1	3	3	3	62.854	15.831	285.634	18.865	289.018	18.960	273.157	18.412	0.0025
ARG	2	1	1	1	186.732	17.955	81.463	19.742	68.406	21.862	86.493	17.299	0.0048
ARG	2	1	1	2	186.732	17.955	81.463	19.742	68.406	21.862	179.906	30.856	0.0133
ARG	2	1	1	3	186.732	17.955	81.463	19.742	68.406	21.862	273.157	18.412	0.0110
ARG	2	1	2	1	186.732	17.955	81.463	19.742	181.277	20.855	86.493	17.299	0.0095
ARG	2	1	2	2	186.732	17.955	81.463	19.742	181.277	20.855	179.906	30.856	0.0301
ARG	2	1	2	3	186.732	17.955	81.463	19.742	181.277	20.855	273.157	18.412	0.0186
ARG	2	1	3	1	186.732	17.955	81.463	19.742	289.018	18.960	86.493	17.299	0.0015
ARG	2	1	3	2	186.732	17.955	81.463	19.742	289.018	18.960	179.906	30.856	0.0108
ARG	2	1	3	3	186.732	17.955	81.463	19.742	289.018	18.960	273.157	18.412	0.0070
ARG	2	2	1	1	186.732	17.955	179.967	18.559	68.406	21.862	86.493	17.299	0.0196
ARG	2	2	1	2	186.732	17.955	179.967	18.559	68.406	21.862	179.906	30.856	0.0232
ARG	2	2	1	3	186.732	17.955	179.967	18.559	68.406	21.862	273.157	18.412	0.0124
ARG	2	2	2	1	186.732	17.955	179.967	18.559	181.277	20.855	86.493	17.299	0.0276
ARG	2	2	2	2	186.732	17.955	179.967	18.559	181.277	20.855	179.906	30.856	0.0629
ARG	2	2	2	3	186.732	17.955	179.967	18.559	181.277	20.855	273.157	18.412	0.0462
ARG	2	2	3	1	186.732	17.955	179.967	18.559	289.018	18.960	86.493	17.299	0.0059
ARG	2	2	3	2	186.732	17.955	179.967	18.559	289.018	18.960	179.906	30.856	0.0470
ARG	2	2	3	3	186.732	17.955	179.967	18.559	289.018	18.960	273.157	18.412	0.0272
ARG	2	3	1	1	186.732	17.955	285.634	18.865	68.406	21.862	86.493	17.299	0.0010
ARG	2	3	1	2	186.732	17.955	285.634	18.865	68.406	21.862	179.906	30.856	0.0059
ARG	2	3	1	3	186.732	17.955	285.634	18.865	68.406	21.862	273.157	18.412	0.0029
ARG	2	3	2	1	186.732	17.955	285.634	18.865	181.277	20.855	86.493	17.299	0.0194
ARG	2	3	2	2	186.732	17.955	285.634	18.865	181.277	20.855	179.906	30.856	0.0449
ARG	2	3	2	3	186.732	17.955	285.634	18.865	181.277	20.855	273.157	18.412	0.0377
ARG	2	3	3	1	186.732	17.955	285.634	18.865	289.018	18.960	86.493	17.299	0.0034
ARG	2	3	3	2	186.732	17.955	285.634	18.865	289.018	18.960	179.906	30.856	0.0272
ARG	2	3	3	3	186.732	17.955	285.634	18.865	289.018	18.960	273.157	18.412	0.0188
ARG	3	1	1	1	293.018	15.389	81.463	19.742	68.406	21.862	86.493	17.299	0.0006
ARG	3	1	1	2	293.018	15.389	81.463	19.742	68.406	21.862	179.906	30.856	0.0027
ARG	3	1	1	3	293.018	15.389	81.463	19.742	68.406	21.862	273.157	18.412	0.0008
ARG	3	1	2	1	293.018	15.389	81.463	19.742	181.277	20.855	86.493	17.299	0.0019
ARG	3	1	2	2	293.018	15.389	81.463	19.742	181.277	20.855	179.906	30.856	0.0086
ARG	3	1	2	3	293.018	15.389	81.463	19.742	181.277	20.855	273.157	18.412	0.0023
ARG	3	1	3	1	293.018	15.389	81.463	19.742	289.018	18.960	86.493	17.299	0.0006
ARG	3	1	3	2	293.018	15.389	81.463	19.742	289.018	18.960	179.906	30.856	0.0029
ARG	3	1	3	3	293.018	15.389	81.463	19.742	289.018	18.960	273.157	18.412	0.0010

AName	r1	r2	r3	r4	avg Chi1	sig Chi1	avg Chi2	sig Chi2	agv Chi3	sig Chi3	avg Chi4	sig Chi4	pr234iAr1
ARG	3	2	1	1	293.018	15.389	179.967	18.559	68.406	21.862	86.493	17.299	0.0055
ARG	3	2	1	2	293.018	15.389	179.967	18.559	68.406	21.862	179.906	30.856	0.0249
ARG	3	2	1	3	293.018	15.389	179.967	18.559	68.406	21.862	273.157	18.412	0.0042
ARG	3	2	2	1	293.018	15.389	179.967	18.559	181.277	20.855	86.493	17.299	0.0185
ARG	3	2	2	2	293.018	15.389	179.967	18.559	181.277	20.855	179.906	30.856	0.0483
ARG	3	2	2	3	293.018	15.389	179.967	18.559	181.277	20.855	273.157	18.412	0.0202
ARG	3	2	3	1	293.018	15.389	179.967	18.559	289.018	18.960	86.493	17.299	0.0027
ARG	3	2	3	2	293.018	15.389	179.967	18.559	289.018	18.960	179.906	30.856	0.0377
ARG	3	2	3	3	293.018	15.389	179.967	18.559	289.018	18.960	273.157	18.412	0.0156
ARG	3	3	1	1	293.018	15.389	285.634	18.865	68.406	21.862	86.493	17.299	0.0027
ARG	3	3	1	2	293.018	15.389	285.634	18.865	68.406	21.862	179.906	30.856	0.0046
ARG	3	3	1	3	293.018	15.389	285.634	18.865	68.406	21.862	273.157	18.412	0.0006
ARG	3	3	2	1	293.018	15.389	285.634	18.865	181.277	20.855	86.493	17.299	0.0097
ARG	3	3	2	2	293.018	15.389	285.634	18.865	181.277	20.855	179.906	30.856	0.0310
ARG	3	3	2	3	293.018	15.389	285.634	18.865	181.277	20.855	273.157	18.412	0.0120
ARG	3	3	3	1	293.018	15.389	285.634	18.865	289.018	18.960	86.493	17.299	0.0034
ARG	3	3	3	2	293.018	15.389	285.634	18.865	289.018	18.960	179.906	30.856	0.0179
ARG	3	3	3	3	293.018	15.389	285.634	18.865	289.018	18.960	273.157	18.412	0.0120
ASN	1	1	0	0	60.467	12.120	56.211	17.028					0.0390
ASN	1	2	0	0	60.467	12.120	178.556	17.178					0.0611
ASN	1	3	0	0	60.467	12.120	124.045	14.930					0.0708
ASN	2	1	0	0	189.099	12.005	56.211	17.028					0.0900
ASN	2	2	0	0	189.099	12.005	178.556	17.178					0.1057
ASN	2	3	0	0	189.099	12.005	124.045	14.930					0.1197
ASN	3	1	0	0	291.939	13.241	56.211	17.028					0.0419
ASN	3	2	0	0	291.939	13.241	178.556	17.178					0.1484
ASN	3	3	0	0	291.939	13.241	124.045	14.930					0.3234
ASP	1	1	0	0	64.247	9.243	53.736	17.171					0.0217
ASP	1	2	0	0	64.247	9.243	177.893	15.886					0.0956
ASP	1	3	0	0	64.247	9.243	130.087	15.240					0.0543
ASP	2	1	0	0	188.557	12.389	53.736	17.171					0.1064
ASP	2	2	0	0	188.557	12.389	177.893	15.886					0.2510
ASP	2	3	0	0	188.557	12.389	130.087	15.240					0.1277
ASP	3	1	0	0	289.847	12.003	53.736	17.171					0.0439
ASP	3	2	0	0	289.847	12.003	177.893	15.886					0.1383
ASP	3	3	0	0	289.847	12.003	130.087	15.240					0.1611
CYS	1	0	0	0	61.131	9.154							0.1057
CYS	2	0	0	0	189.678	8.581							0.1991
CYS	3	0	0	0	296.542	13.099							0.6952
HIS	1	1	0	0	75.007	12.349	91.559	18.271					0.3232
HIS	1	2	0	0	75.007	12.349	177.160	15.584					0.1960
HIS	2	1	0	0	183.085	13.636	91.559	18.271					0.2383
HIS	2	2	0	0	183.085	13.636	177.160	15.584					0.0323
HIS	3	1	0	0	292.362	13.100	91.559	18.271					0.1670
HIS	3	2	0	0	292.362	13.100	177.160	15.584					0.0431
GLN	1	1	1	0	55.947	17.839	75.723	14.686	59.816	15.012			0.0025
GLN	1	1	2	0	55.947	17.839	75.723	14.686	181.093	17.559			0.0135
GLN	1	1	3	0	55.947	17.839	75.723	14.686	118.337	16.572			0.0270
GLN	1	2	1	0	55.947	17.839	180.946	16.999	59.816	15.012			0.0245
GLN	1	2	2	0	55.947	17.839	180.946	16.999	181.093	17.559			0.0372
GLN	1	2	3	0	55.947	17.839	180.946	16.999	118.337	16.572			0.0157
GLN	1	3	1	0	55.947	17.839	288.727	13.715	59.816	15.012			0.0096
GLN	1	3	2	0	55.947	17.839	288.727	13.715	181.093	17.559			0.0560
GLN	1	3	3	0	55.947	17.839	288.727	13.715	118.337	16.572			0.0287
GLN	2	1	1	0	183.335	15.483	75.723	14.686	59.816	15.012			0.0314
GLN	2	1	2	0	183.335	15.483	75.723	14.686	181.093	17.559			0.0526
GLN	2	1	3	0	183.335	15.483	75.723	14.686	118.337	16.572			0.0593
GLN	2	2	1	0	183.335	15.483	180.946	16.999	59.816	15.012			0.0438
GLN	2	2	2	0	183.335	15.483	180.946	16.999	181.093	17.559			0.0457
GLN	2	2	3	0	183.335	15.483	180.946	16.999	118.337	16.572			0.0400
GLN	2	3	1	0	183.335	15.483	288.727	13.715	59.816	15.012			0.0091
GLN	2	3	2	0	183.335	15.483	288.727	13.715	181.093	17.559			0.0458
GLN	2	3	3	0	183.335	15.483	288.727	13.715	118.337	16.572			0.0482
GLN	3	1	1	0	292.702	13.047	75.723	14.686	59.816	15.012			0.0094
GLN	3	1	2	0	292.702	13.047	75.723	14.686	181.093	17.559			0.0113
GLN	3	1	3	0	292.702	13.047	75.723	14.686	118.337	16.572			0.0080
GLN	3	2	1	0	292.702	13.047	180.946	16.999	59.816	15.012			0.0749
GLN	3	2	2	0	292.702	13.047	180.946	16.999	181.093	17.559			0.0466
GLN	3	2	3	0	292.702	13.047	180.946	16.999	118.337	16.572			0.0474
GLN	3	3	1	0	292.702	13.047	288.727	13.715	59.816	15.012			0.0099
GLN	3	3	2	0	292.702	13.047	288.727	13.715	181.093	17.559			0.0671
GLN	3	3	3	0	292.702	13.047	288.727	13.715	118.337	16.572			0.1348
GLU	1	1	1	0	58.952	18.678	78.546	16.288	56.787	16.987			0.0027
GLU	1	1	2	0	58.952	18.678	78.546	16.288	179.406	17.427			0.0107
GLU	1	1	3	0	58.952	18.678	78.546	16.288	120.992	16.901			0.0165
GLU	1	2	1	0	58.952	18.678	178.710	14.849	56.787	16.987			0.0151
GLU	1	2	2	0	58.952	18.678	178.710	14.849	179.406	17.427			0.0289
GLU	1	2	3	0	58.952	18.678	178.710	14.849	120.992	16.901			0.0333

AName	r1	r2	r3	r4	avg Chi1	sig Chi1	avg Chi2	sig Chi2	agv Chi3	sig Chi3	avg Chi4	sig Chi4	pr234iAr1
GLU	1	3	1	0	58.952	18.678	286.022	15.320	56.787	16.987			0.0155
GLU	1	3	2	0	58.952	18.678	286.022	15.320	179.406	17.427			0.0322
GLU	1	3	3	0	58.952	18.678	286.022	15.320	120.992	16.901			0.0302
GLU	2	1	1	0	186.349	15.779	78.546	16.288	56.787	16.987			0.0168
GLU	2	1	2	0	186.349	15.779	78.546	16.288	179.406	17.427			0.0302
GLU	2	1	3	0	186.349	15.779	78.546	16.288	120.992	16.901			0.0443
GLU	2	2	1	0	186.349	15.779	178.710	14.849	56.787	16.987			0.0299
GLU	2	2	2	0	186.349	15.779	178.710	14.849	179.406	17.427			0.0950
GLU	2	2	3	0	186.349	15.779	178.710	14.849	120.992	16.901			0.0635
GLU	2	3	1	0	186.349	15.779	286.022	15.320	56.787	16.987			0.0229
GLU	2	3	2	0	186.349	15.779	286.022	15.320	179.406	17.427			0.0577
GLU	2	3	3	0	186.349	15.779	286.022	15.320	120.992	16.901			0.0685
GLU	3	1	1	0	294.291	14.242	78.546	16.288	56.787	16.987			0.0104
GLU	3	1	2	0	294.291	14.242	78.546	16.288	179.406	17.427			0.0363
GLU	3	1	3	0	294.291	14.242	78.546	16.288	120.992	16.901			0.0171
GLU	3	2	1	0	294.291	14.242	178.710	14.849	56.787	16.987			0.0463
GLU	3	2	2	0	294.291	14.242	178.710	14.849	179.406	17.427			0.0645
GLU	3	2	3	0	294.291	14.242	178.710	14.849	120.992	16.901			0.0524
GLU	3	3	1	0	294.291	14.242	286.022	15.320	56.787	16.987			0.0253
GLU	3	3	2	0	294.291	14.242	286.022	15.320	179.406	17.427			0.0735
GLU	3	3	3	0	294.291	14.242	286.022	15.320	120.992	16.901			0.0601
ILE	1	1	0	0	59.630	10.940	74.214	17.489					0.0131
ILE	1	2	0	0	59.630	10.940	170.527	9.750					0.0883
ILE	1	3	0	0	59.630	10.940	298.127	9.237					0.0385
ILE	2	1	0	0	190.112	10.368	74.214	17.489					0.1165
ILE	2	2	0	0	190.112	10.368	170.527	9.750					0.1246
ILE	2	3	0	0	190.112	10.368	298.127	9.237					0.2428
ILE	3	1	0	0	297.721	8.680	74.214	17.489					0.0103
ILE	3	2	0	0	297.721	8.680	170.527	9.750					0.2711
ILE	3	3	0	0	297.721	8.680	298.127	9.237					0.0946
LEU	1	1	0	0	63.060	23.189	59.896	17.684					0.0061
LEU	1	2	0	0	63.060	23.189	178.542	12.810					0.0711
LEU	1	3	0	0	63.060	23.189	312.866	27.741					0.1152
LEU	2	1	0	0	187.560	15.395	59.896	17.684					0.1505
LEU	2	2	0	0	187.560	15.395	178.542	12.810					0.1026
LEU	2	3	0	0	187.560	15.395	312.866	27.741					0.2097
LEU	3	1	0	0	292.797	15.365	59.896	17.684					0.0356
LEU	3	2	0	0	292.797	15.365	178.542	12.810					0.2893
LEU	3	3	0	0	292.797	15.365	312.866	27.741					0.0198
LYS	1	1	1	1	64.594	14.598	72.985	19.487	73.808	20.741	68.459	19.372	0.0003
LYS	1	1	1	2	64.594	14.598	72.985	19.487	73.808	20.741	178.934	16.499	0.0010
LYS	1	1	1	3	64.594	14.598	72.985	19.487	73.808	20.741	292.303	16.546	0.0002
LYS	1	1	2	1	64.594	14.598	72.985	19.487	180.046	16.074	68.459	19.372	0.0007
LYS	1	1	2	2	64.594	14.598	72.985	19.487	180.046	16.074	178.934	16.499	0.0028
LYS	1	1	2	3	64.594	14.598	72.985	19.487	180.046	16.074	292.303	16.546	0.0013
LYS	1	1	3	1	64.594	14.598	72.985	19.487	288.811	20.575	68.459	19.372	0.0002
LYS	1	1	3	2	64.594	14.598	72.985	19.487	288.811	20.575	178.934	16.499	0.0017
LYS	1	1	3	3	64.594	14.598	72.985	19.487	288.811	20.575	292.303	16.546	0.0015
LYS	1	2	1	1	64.594	14.598	179.309	15.392	73.808	20.741	68.459	19.372	0.0012
LYS	1	2	1	2	64.594	14.598	179.309	15.392	73.808	20.741	178.934	16.499	0.0070
LYS	1	2	1	3	64.594	14.598	179.309	15.392	73.808	20.741	292.303	16.546	0.0008
LYS	1	2	2	1	64.594	14.598	179.309	15.392	180.046	16.074	68.459	19.372	0.0032
LYS	1	2	2	2	64.594	14.598	179.309	15.392	180.046	16.074	178.934	16.499	0.0246
LYS	1	2	2	3	64.594	14.598	179.309	15.392	180.046	16.074	292.303	16.546	0.0052
LYS	1	2	3	1	64.594	14.598	179.309	15.392	288.811	20.575	68.459	19.372	0.0014
LYS	1	2	3	2	64.594	14.598	179.309	15.392	288.811	20.575	178.934	16.499	0.0121
LYS	1	2	3	3	64.594	14.598	179.309	15.392	288.811	20.575	292.303	16.546	0.0033
LYS	1	3	1	1	64.594	14.598	289.689	14.175	73.808	20.741	68.459	19.372	0.0005
LYS	1	3	1	2	64.594	14.598	289.689	14.175	73.808	20.741	178.934	16.499	0.0010
LYS	1	3	1	3	64.594	14.598	289.689	14.175	73.808	20.741	292.303	16.546	0.0003
LYS	1	3	2	1	64.594	14.598	289.689	14.175	180.046	16.074	68.459	19.372	0.0008
LYS	1	3	2	2	64.594	14.598	289.689	14.175	180.046	16.074	178.934	16.499	0.0151
LYS	1	3	2	3	64.594	14.598	289.689	14.175	180.046	16.074	292.303	16.546	0.0018
LYS	1	3	3	1	64.594	14.598	289.689	14.175	288.811	20.575	68.459	19.372	0.0005
LYS	1	3	3	2	64.594	14.598	289.689	14.175	288.811	20.575	178.934	16.499	0.0068
LYS	1	3	3	3	64.594	14.598	289.689	14.175	288.811	20.575	292.303	16.546	0.0020
LYS	2	1	1	1	187.012	14.519	72.985	19.487	73.808	20.741	68.459	19.372	0.0007
LYS	2	1	1	2	187.012	14.519	72.985	19.487	73.808	20.741	178.934	16.499	0.0025
LYS	2	1	1	3	187.012	14.519	72.985	19.487	73.808	20.741	292.303	16.546	0.0028
LYS	2	1	2	1	187.012	14.519	72.985	19.487	180.046	16.074	68.459	19.372	0.0065
LYS	2	1	2	2	187.012	14.519	72.985	19.487	180.046	16.074	178.934	16.499	0.0263
LYS	2	1	2	3	187.012	14.519	72.985	19.487	180.046	16.074	292.303	16.546	0.0127
LYS	2	1	3	1	187.012	14.519	72.985	19.487	288.811	20.575	68.459	19.372	0.0007
LYS	2	1	3	2	187.012	14.519	72.985	19.487	288.811	20.575	178.934	16.499	0.0126
LYS	2	1	3	3	187.012	14.519	72.985	19.487	288.811	20.575	292.303	16.546	0.0073
LYS	2	2	1	1	187.012	14.519	179.309	15.392	73.808	20.741	68.459	19.372	0.0033
LYS	2	2	1	2	187.012	14.519	179.309	15.392	73.808	20.741	178.934	16.499	0.0278
LYS	2	2	1	3	187.012	14.519	179.309	15.392	73.808	20.741	292.303	16.546	0.0158

AName	r1	r2	r3	r4	avg Chi1	sig Chi1	avg Chi2	sig Chi2	agv Chi3	sig Chi3	avg Chi4	sig Chi4	pr234iAr1
LYS	2	2	2	1	187.012	14.519	179.309	15.392	180.046	16.074	68.459	19.372	0.0288
LYS	2	2	2	2	187.012	14.519	179.309	15.392	180.046	16.074	178.934	16.499	0.1749
LYS	2	2	2	3	187.012	14.519	179.309	15.392	180.046	16.074	292.303	16.546	0.0701
LYS	2	2	3	1	187.012	14.519	179.309	15.392	288.811	20.575	68.459	19.372	0.0051
LYS	2	2	3	2	187.012	14.519	179.309	15.392	288.811	20.575	178.934	16.499	0.0582
LYS	2	2	3	3	187.012	14.519	179.309	15.392	288.811	20.575	292.303	16.546	0.0379
LYS	2	3	1	1	187.012	14.519	289.689	14.175	73.808	20.741	68.459	19.372	0.0012
LYS	2	3	1	2	187.012	14.519	289.689	14.175	73.808	20.741	178.934	16.499	0.0048
LYS	2	3	1	3	187.012	14.519	289.689	14.175	73.808	20.741	292.303	16.546	0.0033
LYS	2	3	2	1	187.012	14.519	289.689	14.175	180.046	16.074	68.459	19.372	0.0096
LYS	2	3	2	2	187.012	14.519	289.689	14.175	180.046	16.074	178.934	16.499	0.0626
LYS	2	3	2	3	187.012	14.519	289.689	14.175	180.046	16.074	292.303	16.546	0.0175
LYS	2	3	3	1	187.012	14.519	289.689	14.175	288.811	20.575	68.459	19.372	0.0020
LYS	2	3	3	2	187.012	14.519	289.689	14.175	288.811	20.575	178.934	16.499	0.0429
LYS	2	3	3	3	187.012	14.519	289.689	14.175	288.811	20.575	292.303	16.546	0.0083
LYS	3	1	1	1	291.832	13.667	72.985	19.487	73.808	20.741	68.459	19.372	0.0003
LYS	3	1	1	2	291.832	13.667	72.985	19.487	73.808	20.741	178.934	16.499	0.0013
LYS	3	1	1	3	291.832	13.667	72.985	19.487	73.808	20.741	292.303	16.546	0.0005
LYS	3	1	2	1	291.832	13.667	72.985	19.487	180.046	16.074	68.459	19.372	0.0018
LYS	3	1	2	2	291.832	13.667	72.985	19.487	180.046	16.074	178.934	16.499	0.0038
LYS	3	1	2	3	291.832	13.667	72.985	19.487	180.046	16.074	292.303	16.546	0.0028
LYS	3	1	3	1	291.832	13.667	72.985	19.487	288.811	20.575	68.459	19.372	0.0003
LYS	3	1	3	2	291.832	13.667	72.985	19.487	288.811	20.575	178.934	16.499	0.0035
LYS	3	1	3	3	291.832	13.667	72.985	19.487	288.811	20.575	292.303	16.546	0.0005
LYS	3	2	1	1	291.832	13.667	179.309	15.392	73.808	20.741	68.459	19.372	0.0018
LYS	3	2	1	2	291.832	13.667	179.309	15.392	73.808	20.741	178.934	16.499	0.0106
LYS	3	2	1	3	291.832	13.667	179.309	15.392	73.808	20.741	292.303	16.546	0.0043
LYS	3	2	2	1	291.832	13.667	179.309	15.392	180.046	16.074	68.459	19.372	0.0150
LYS	3	2	2	2	291.832	13.667	179.309	15.392	180.046	16.074	178.934	16.499	0.0744
LYS	3	2	2	3	291.832	13.667	179.309	15.392	180.046	16.074	292.303	16.546	0.0260
LYS	3	2	3	1	291.832	13.667	179.309	15.392	288.811	20.575	68.459	19.372	0.0032
LYS	3	2	3	2	291.832	13.667	179.309	15.392	288.811	20.575	178.934	16.499	0.0193
LYS	3	2	3	3	291.832	13.667	179.309	15.392	288.811	20.575	292.303	16.546	0.0120
LYS	3	3	1	1	291.832	13.667	289.689	14.175	73.808	20.741	68.459	19.372	0.0003
LYS	3	3	1	2	291.832	13.667	289.689	14.175	73.808	20.741	178.934	16.499	0.0027
LYS	3	3	1	3	291.832	13.667	289.689	14.175	73.808	20.741	292.303	16.546	0.0002
LYS	3	3	2	1	291.832	13.667	289.689	14.175	180.046	16.074	68.459	19.372	0.0063
LYS	3	3	2	2	291.832	13.667	289.689	14.175	180.046	16.074	178.934	16.499	0.0376
LYS	3	3	2	3	291.832	13.667	289.689	14.175	180.046	16.074	292.303	16.546	0.0152
LYS	3	3	3	1	291.832	13.667	289.689	14.175	288.811	20.575	68.459	19.372	0.0010
LYS	3	3	3	2	291.832	13.667	289.689	14.175	288.811	20.575	178.934	16.499	0.0093
LYS	3	3	3	3	291.832	13.667	289.689	14.175	288.811	20.575	292.303	16.546	0.0023
MET	1	1	1	0	65.359	13.901	66.817	12.936	77.949	15.521			0.0026
MET	1	1	2	0	65.359	13.901	66.817	12.936	175.170	23.516			0.0095
MET	1	1	3	0	65.359	13.901	66.817	12.936	284.020	18.928			0.0147
MET	1	2	1	0	65.359	13.901	179.049	11.331	77.949	15.521			0.0104
MET	1	2	2	0	65.359	13.901	179.049	11.331	175.170	23.516			0.0164
MET	1	2	3	0	65.359	13.901	179.049	11.331	284.020	18.928			0.0182
MET	1	3	1	0	65.359	13.901	292.287	18.082	77.949	15.521			0.0026
MET	1	3	2	0	65.359	13.901	292.287	18.082	175.170	23.516			0.0971
MET	1	3	3	0	65.359	13.901	292.287	18.082	284.020	18.928			0.0225
MET	2	1	1	0	174.424	17.007	66.817	12.936	77.949	15.521			0.0216
MET	2	1	2	0	174.424	17.007	66.817	12.936	175.170	23.516			0.0328
MET	2	1	3	0	174.424	17.007	66.817	12.936	284.020	18.928			0.0892
MET	2	2	1	0	174.424	17.007	179.049	11.331	77.949	15.521			0.0596
MET	2	2	2	0	174.424	17.007	179.049	11.331	175.170	23.516			0.0743
MET	2	2	3	0	174.424	17.007	179.049	11.331	284.020	18.928			0.0545
MET	2	3	1	0	174.424	17.007	292.287	18.082	77.949	15.521			0.0043
MET	2	3	2	0	174.424	17.007	292.287	18.082	175.170	23.516			0.0633
MET	2	3	3	0	174.424	17.007	292.287	18.082	284.020	18.928			0.0441
MET	3	1	1	0	292.958	11.162	66.817	12.936	77.949	15.521			0.0026
MET	3	1	2	0	292.958	11.162	66.817	12.936	175.170	23.516			0.0026
MET	3	1	3	0	292.958	11.162	66.817	12.936	284.020	18.928			0.0182
MET	3	2	1	0	292.958	11.162	179.049	11.331	77.949	15.521			0.0743
MET	3	2	2	0	292.958	11.162	179.049	11.331	175.170	23.516			0.0976
MET	3	2	3	0	292.958	11.162	179.049	11.331	284.020	18.928			0.0390
MET	3	3	1	0	292.958	11.162	292.287	18.082	77.949	15.521			0.0164
MET	3	3	2	0	292.958	11.162	292.287	18.082	175.170	23.516			0.0503
MET	3	3	3	0	292.958	11.162	292.287	18.082	284.020	18.928			0.0615
PHE	1	1	0	0	66.567	8.553	92.908	20.752					0.4291
PHE	1	2	0	0	66.567	8.553	170.784	14.449					0.1072
PHE	2	1	0	0	181.848	11.992	92.908	20.752					0.1689
PHE	2	2	0	0	181.848	11.992	170.784	14.449					0.0102
PHE	3	1	0	0	293.755	11.812	92.908	20.752					0.2223
PHE	3	2	0	0	293.755	11.812	170.784	14.449					0.0624
SER	1	0	0	0	58.330	14.840							0.4772
SER	2	0	0	0	181.625	15.165							0.1948
SER	3	0	0	0	301.092	16.652							0.3280

AName	r1	r2	r3	r4	avg Chi1	sig Chi1	avg Chi2	sig Chi2	agv Chi3	sig Chi3	avg Chi4	sig Chi4	pr234iAr1
THR	1	0	0	0	59.067	10.947							0.4599
THR	2	0	0	0	186.878	13.716							0.0574
THR	3	0	0	0	301.295	10.187							0.4827
TRP	1	1	0	0	60.035	8.441	91.573	12.274					0.0550
TRP	1	2	0	0	60.035	8.441	208.711	42.324					0.0226
TRP	1	3	0	0	60.035	8.441	92.626	13.241					0.1380
TRP	2	1	0	0	175.171	10.209	91.573	12.274					0.0346
TRP	2	2	0	0	175.171	10.209	208.711	42.324					0.0181
TRP	2	3	0	0	175.171	10.209	92.626	13.241					0.1040
TRP	3	1	0	0	293.898	10.429	91.573	12.274					0.1265
TRP	3	2	0	0	293.898	10.429	208.711	42.324					0.1160
TRP	3	3	0	0	293.898	10.429	92.626	13.241					0.3853
TYR	1	1	0	0	58.091	10.527	89.526	16.781					0.4655
TYR	1	2	0	0	58.091	10.527	169.522	14.207					0.1170
TYR	2	1	0	0	178.173	10.738	89.526	16.781					0.1301
TYR	2	2	0	0	178.173	10.738	169.522	14.207					0.0022
TYR	3	1	0	0	295.220	11.204	89.526	16.781					0.2716
TYR	3	2	0	0	295.220	11.204	169.522	14.207					0.0136
VAL	1	0	0	0	64.302	18.935							0.0793
VAL	2	0	0	0	176.793	9.825							0.7612
VAL	3	0	0	0	300.354	13.942							0.1595

Table C.1.: Rotamer library calculated on unbound proteins of the whole test set

AS	r1	r2	r3	r4	$P(r1234)$ ALL	$P(r1234)$ HELIX	$P(r1234)$ SHEET	$P(r1234)$ RND
ASN	1	1	0	0	0.0390	0.0098	0.0929	0.0326
ASN	1	2	0	0	0.0611	0.0098	0.1714	0.0598
ASN	1	3	0	0	0.0708	0.0296	0.1023	0.0220
ASN	2	1	0	0	0.0900	0.0098	0.1071	0.0707
ASN	2	2	0	0	0.1057	0.0098	0.0571	0.2174
ASN	2	3	0	0	0.1197	0.2468	0.1534	0.1596
ASN	3	1	0	0	0.0419	0.0098	0.1643	0.0109
ASN	3	2	0	0	0.1484	0.2549	0.0857	0.1685
ASN	3	3	0	0	0.3234	0.4196	0.0657	0.2586
ASP	1	1	0	0	0.0217	0.0156	0.0357	0.0183
ASP	1	2	0	0	0.0956	0.0104	0.0245	0.1155
ASP	1	3	0	0	0.0543	0.0947	0.0595	0.0092
ASP	2	1	0	0	0.1064	0.0729	0.0119	0.1009
ASP	2	2	0	0	0.2510	0.0781	0.0981	0.4296
ASP	2	3	0	0	0.1277	0.1999	0.2738	0.1147
ASP	3	1	0	0	0.0439	0.0208	0.0476	0.0138
ASP	3	2	0	0	0.1383	0.2708	0.2821	0.1063
ASP	3	3	0	0	0.1611	0.2367	0.1667	0.0917
CYS	1	0	0	0	0.1057	0.1638	0.0769	0.1628
CYS	2	0	0	0	0.1991	0.3621	0.0513	0.2791
CYS	3	0	0	0	0.6952	0.4741	0.8718	0.5581
HIS	1	1	0	0	0.3232	0.3089	0.0750	0.0796
HIS	1	2	0	0	0.1960	0.0233	0.2574	0.0784
HIS	2	1	0	0	0.2383	0.1475	0.3625	0.3695
HIS	2	2	0	0	0.0323	0.0116	0.0772	0.0588
HIS	3	1	0	0	0.1670	0.3808	0.1250	0.1195
HIS	3	2	0	0	0.0431	0.1279	0.1029	0.2941
GLN	1	1	1	0	0.0025	0.0083	0.0022	0.0063
GLN	1	1	2	0	0.0135	0.0224	0.0079	0.0044
GLN	1	1	3	0	0.0270	0.0556	0.0443	0.0159
GLN	1	2	1	0	0.0245	0.0255	0.0513	0.0238
GLN	1	2	2	0	0.0372	0.0504	0.0560	0.0273
GLN	1	2	3	0	0.0157	0.0139	0.0052	0.0238
GLN	1	3	1	0	0.0096	0.0056	0.0015	0.0159
GLN	1	3	2	0	0.0560	0.0364	0.0105	0.0573
GLN	1	3	3	0	0.0287	0.0389	0.0573	0.0238
GLN	2	1	1	0	0.0314	0.0528	0.0522	0.0175
GLN	2	1	2	0	0.0526	0.0616	0.0917	0.0120
GLN	2	1	3	0	0.0593	0.0167	0.0182	0.0635
GLN	2	2	1	0	0.0438	0.0369	0.0539	0.0238
GLN	2	2	2	0	0.0457	0.0336	0.0614	0.0091
GLN	2	2	3	0	0.0400	0.0417	0.0286	0.0238
GLN	2	3	1	0	0.0091	0.0083	0.0021	0.0048
GLN	2	3	2	0	0.0458	0.0392	0.0341	0.0409
GLN	2	3	3	0	0.0482	0.0667	0.0573	0.0476
GLN	3	1	1	0	0.0094	0.0222	0.0003	0.0079
GLN	3	1	2	0	0.0113	0.0056	0.0079	0.0082
GLN	3	1	3	0	0.0080	0.0083	0.0052	0.0238
GLN	3	2	1	0	0.0749	0.0709	0.0641	0.0397
GLN	3	2	2	0	0.0466	0.0280	0.0294	0.0455
GLN	3	2	3	0	0.0474	0.0583	0.0599	0.0476
GLN	3	3	1	0	0.0099	0.0250	0.0043	0.0190
GLN	3	3	2	0	0.0671	0.0812	0.1127	0.0573
GLN	3	3	3	0	0.1348	0.0861	0.0807	0.3095
GLU	1	1	1	0	0.0027	0.0008	0.0046	0.0012
GLU	1	1	2	0	0.0107	0.0102	0.0252	0.0296

AS	r1	r2	r3	r4	$P(r1234)$ ALL	$P(r1234)$ HELIX	$P(r1234)$ SHEET	$P(r1234)$ RND
GLU	1	1	3	0	0.0165	0.0101	0.0049	0.0112
GLU	1	2	1	0	0.0151	0.0278	0.0370	0.0222
GLU	1	2	2	0	0.0289	0.0282	0.0378	0.0185
GLU	1	2	3	0	0.0333	0.0253	0.0272	0.0598
GLU	1	3	1	0	0.0155	0.0101	0.0049	0.0025
GLU	1	3	2	0	0.0322	0.0355	0.0252	0.0151
GLU	1	3	3	0	0.0302	0.0126	0.0123	0.0262
GLU	2	1	1	0	0.0168	0.0177	0.0247	0.0333
GLU	2	1	2	0	0.0302	0.0431	0.0378	0.0111
GLU	2	1	3	0	0.0443	0.0354	0.0115	0.0224
GLU	2	2	1	0	0.0299	0.0505	0.0123	0.0259
GLU	2	2	2	0	0.0950	0.1230	0.2646	0.0963
GLU	2	2	3	0	0.0635	0.0884	0.0815	0.0112
GLU	2	3	1	0	0.0229	0.0177	0.0494	0.0667
GLU	2	3	2	0	0.0577	0.0660	0.0630	0.0681
GLU	2	3	3	0	0.0685	0.0732	0.0194	0.0673
GLU	3	1	1	0	0.0104	0.0017	0.0077	0.0025
GLU	3	1	2	0	0.0363	0.0279	0.0630	0.0593
GLU	3	1	3	0	0.0171	0.0126	0.0082	0.0112
GLU	3	2	1	0	0.0463	0.0202	0.0370	0.0185
GLU	3	2	2	0	0.0645	0.1102	0.0882	0.0222
GLU	3	2	3	0	0.0524	0.0429	0.0272	0.1346
GLU	3	3	1	0	0.0253	0.0126	0.0074	0.0086
GLU	3	3	2	0	0.0735	0.0736	0.0126	0.0908
GLU	3	3	3	0	0.0601	0.0227	0.0053	0.0635
ILE	1	1	0	0	0.0131	0.0097	0.0054	0.0029
ILE	1	2	0	0	0.0883	0.0877	0.0753	0.1345
ILE	1	3	0	0	0.0385	0.0393	0.0081	0.0095
ILE	2	1	0	0	0.1165	0.1039	0.0780	0.1286
ILE	2	2	0	0	0.1246	0.0584	0.0618	0.1922
ILE	2	3	0	0	0.2428	0.2683	0.3708	0.2095
ILE	3	1	0	0	0.0103	0.0227	0.0054	0.0066
ILE	3	2	0	0	0.2711	0.2857	0.3737	0.2162
ILE	3	3	0	0	0.0946	0.1243	0.0217	0.1000
LEU	1	1	0	0	0.0061	0.0061	0.0042	0.0059
LEU	1	2	0	0	0.0711	0.0772	0.1004	0.0417
LEU	1	3	0	0	0.1152	0.0891	0.1271	0.1243
LEU	2	1	0	0	0.1505	0.1457	0.2146	0.1391
LEU	2	2	0	0	0.1026	0.1443	0.1151	0.1012
LEU	2	3	0	0	0.2097	0.2206	0.1500	0.2189
LEU	3	1	0	0	0.0356	0.0223	0.0125	0.0266
LEU	3	2	0	0	0.2893	0.2886	0.2657	0.3186
LEU	3	3	0	0	0.0198	0.0061	0.0104	0.0237
LYS	1	1	1	1	0.0003	0.0001	0.0000	0.0000
LYS	1	1	1	2	0.0010	0.0007	0.0002	0.0003
LYS	1	1	1	3	0.0002	0.0007	0.0000	0.0003
LYS	1	1	2	1	0.0007	0.0007	0.0005	0.0007
LYS	1	1	2	2	0.0028	0.0033	0.0025	0.0047
LYS	1	1	2	3	0.0013	0.0007	0.0025	0.0013
LYS	1	1	3	1	0.0002	0.0002	0.0000	0.0001
LYS	1	1	3	2	0.0017	0.0033	0.0009	0.0004
LYS	1	1	3	3	0.0015	0.0005	0.0009	0.0044
LYS	1	2	1	1	0.0012	0.0012	0.0002	0.0003
LYS	1	2	1	2	0.0070	0.0051	0.0022	0.0009
LYS	1	2	1	3	0.0008	0.0005	0.0035	0.0042
LYS	1	2	2	1	0.0032	0.0034	0.0036	0.0047
LYS	1	2	2	2	0.0246	0.0262	0.0092	0.0094
LYS	1	2	2	3	0.0052	0.0049	0.0037	0.0092
LYS	1	2	3	1	0.0014	0.0006	0.0005	0.0042
LYS	1	2	3	2	0.0121	0.0066	0.0313	0.0070
LYS	1	2	3	3	0.0033	0.0030	0.0035	0.0046
LYS	1	3	1	1	0.0005	0.0001	0.0000	0.0001
LYS	1	3	1	2	0.0010	0.0013	0.0003	0.0010
LYS	1	3	1	3	0.0003	0.0007	0.0002	0.0006
LYS	1	3	2	1	0.0008	0.0017	0.0001	0.0001
LYS	1	3	2	2	0.0151	0.0164	0.0092	0.0094
LYS	1	3	2	3	0.0018	0.0033	0.0004	0.0046
LYS	1	3	3	1	0.0005	0.0002	0.0005	0.0009
LYS	1	3	3	2	0.0068	0.0033	0.0295	0.0376
LYS	1	3	3	3	0.0020	0.0026	0.0021	0.0018
LYS	2	1	1	1	0.0007	0.0005	0.0001	0.0001
LYS	2	1	1	2	0.0025	0.0016	0.0034	0.0020
LYS	2	1	1	3	0.0028	0.0016	0.0008	0.0020
LYS	2	1	2	1	0.0065	0.0017	0.0048	0.0023
LYS	2	1	2	2	0.0263	0.0262	0.0055	0.0047
LYS	2	1	2	3	0.0127	0.0049	0.0294	0.0025
LYS	2	1	3	1	0.0007	0.0009	0.0006	0.0005
LYS	2	1	3	2	0.0126	0.0131	0.0074	0.0070

AS	r1	r2	r3	r4	$P(r1234)$ ALL	$P(r1234)$ HELIX	$P(r1234)$ SHEET	$P(r1234)$ RND
LYS	2	1	3	3	0.0073	0.0022	0.0277	0.0088
LYS	2	2	1	1	0.0033	0.0012	0.0030	0.0013
LYS	2	2	1	2	0.0278	0.0275	0.0350	0.0070
LYS	2	2	1	3	0.0158	0.0114	0.0334	0.0105
LYS	2	2	2	1	0.0288	0.0320	0.0376	0.0117
LYS	2	2	2	2	0.1749	0.2162	0.2323	0.0352
LYS	2	2	2	3	0.0701	0.0997	0.0478	0.0710
LYS	2	2	3	1	0.0051	0.0025	0.0012	0.0042
LYS	2	2	3	2	0.0582	0.1032	0.0369	0.0634
LYS	2	2	3	3	0.0379	0.0151	0.0352	0.0456
LYS	2	3	1	1	0.0012	0.0005	0.0001	0.0021
LYS	2	3	1	2	0.0048	0.0016	0.0009	0.0078
LYS	2	3	1	3	0.0033	0.0016	0.0005	0.0236
LYS	2	3	2	1	0.0096	0.0034	0.0011	0.0047
LYS	2	3	2	2	0.0626	0.1081	0.0700	0.0282
LYS	2	3	2	3	0.0175	0.0114	0.0033	0.0341
LYS	2	3	3	1	0.0020	0.0013	0.0007	0.0022
LYS	2	3	3	2	0.0429	0.0147	0.0387	0.0962
LYS	2	3	3	3	0.0083	0.0026	0.0030	0.0048
LYS	3	1	1	1	0.0003	0.0001	0.0000	0.0001
LYS	3	1	1	2	0.0013	0.0009	0.0002	0.0024
LYS	3	1	1	3	0.0005	0.0009	0.0000	0.0024
LYS	3	1	2	1	0.0018	0.0010	0.0002	0.0017
LYS	3	1	2	2	0.0038	0.0049	0.0012	0.0094
LYS	3	1	2	3	0.0028	0.0010	0.0012	0.0032
LYS	3	1	3	1	0.0003	0.0002	0.0000	0.0006
LYS	3	1	3	2	0.0035	0.0033	0.0009	0.0020
LYS	3	1	3	3	0.0005	0.0005	0.0009	0.0242
LYS	3	2	1	1	0.0018	0.0025	0.0002	0.0004
LYS	3	2	1	2	0.0106	0.0034	0.0015	0.0014
LYS	3	2	1	3	0.0043	0.0011	0.0035	0.0063
LYS	3	2	2	1	0.0150	0.0101	0.0305	0.0047
LYS	3	2	2	2	0.0744	0.1048	0.0811	0.0775
LYS	3	2	2	3	0.0260	0.0229	0.0276	0.0390
LYS	3	2	3	1	0.0032	0.0018	0.0001	0.0270
LYS	3	2	3	2	0.0193	0.0147	0.0037	0.0023
LYS	3	2	3	3	0.0120	0.0015	0.0035	0.0479
LYS	3	3	1	1	0.0003	0.0002	0.0000	0.0002
LYS	3	3	1	2	0.0027	0.0020	0.0006	0.0029
LYS	3	3	1	3	0.0002	0.0010	0.0003	0.0016
LYS	3	3	2	1	0.0063	0.0017	0.0025	0.0022
LYS	3	3	2	2	0.0376	0.0131	0.0645	0.0868
LYS	3	3	2	3	0.0152	0.0049	0.0018	0.0501
LYS	3	3	3	1	0.0010	0.0004	0.0001	0.0002
LYS	3	3	3	2	0.0093	0.0016	0.0055	0.0070
LYS	3	3	3	3	0.0023	0.0013	0.0004	0.0003
MET	1	1	1	0	0.0026	0.0022	0.0009	0.0110
MET	1	1	2	0	0.0095	0.0175	0.0062	0.0494
MET	1	1	3	0	0.0147	0.0119	0.0123	0.0165
MET	1	2	1	0	0.0104	0.0351	0.0185	0.0494
MET	1	2	2	0	0.0164	0.0175	0.0125	0.0556
MET	1	2	3	0	0.0182	0.0238	0.0134	0.0741
MET	1	3	1	0	0.0026	0.0027	0.0056	0.0165
MET	1	3	2	0	0.0971	0.0877	0.1310	0.0247
MET	1	3	3	0	0.0225	0.0313	0.0056	0.0247
MET	2	1	1	0	0.0216	0.0117	0.0101	0.0219
MET	2	1	2	0	0.0328	0.0526	0.0125	0.0988
MET	2	1	3	0	0.0892	0.0951	0.1111	0.0329
MET	2	2	1	0	0.0596	0.0292	0.1173	0.0247
MET	2	2	2	0	0.0743	0.0117	0.2370	0.0278
MET	2	2	3	0	0.0545	0.0653	0.0134	0.0370
MET	2	3	1	0	0.0043	0.0031	0.0123	0.0329
MET	2	3	2	0	0.0633	0.0994	0.0499	0.0494
MET	2	3	3	0	0.0441	0.0375	0.0062	0.0494
MET	3	1	1	0	0.0026	0.0037	0.0013	0.0165
MET	3	1	2	0	0.0026	0.0117	0.0125	0.0741
MET	3	1	3	0	0.0182	0.0119	0.0062	0.0247
MET	3	2	1	0	0.0743	0.1111	0.0062	0.0247
MET	3	2	2	0	0.0976	0.0643	0.1310	0.0278
MET	3	2	3	0	0.0390	0.0475	0.0535	0.0370
MET	3	3	1	0	0.0164	0.0234	0.0005	0.0247
MET	3	3	2	0	0.0503	0.0409	0.0125	0.0370
MET	3	3	3	0	0.0615	0.0500	0.0005	0.0370
PHE	1	1	0	0	0.4291	0.1294	0.4758	0.3230
PHE	1	2	0	0	0.1072	0.1067	0.0296	0.0122
PHE	2	1	0	0	0.1689	0.2138	0.1562	0.1225
PHE	2	2	0	0	0.0102	0.0061	0.0054	0.0244
PHE	3	1	0	0	0.2223	0.4434	0.2121	0.4569

AS	r1	r2	r3	r4	$P(r1234)$ ALL	$P(r1234)$ HELIX	$P(r1234)$ SHEET	$P(r1234)$ RND
PHE	3	2	0	0	0.0624	0.1006	0.1210	0.0610
SER	1	0	0	0	0.4772	0.5155	0.2825	0.5721
SER	2	0	0	0	0.1948	0.0928	0.3898	0.1674
SER	3	0	0	0	0.3280	0.3918	0.3277	0.2605
THR	1	0	0	0	0.4599	0.2479	0.2845	0.5833
THR	2	0	0	0	0.0574	0.0598	0.0776	0.0152
THR	3	0	0	0	0.4827	0.6923	0.6379	0.4015
TRP	1	1	0	0	0.0550	0.0023	0.2200	0.0769
TRP	1	2	0	0	0.0226	0.0275	0.0090	0.1154
TRP	1	3	0	0	0.1380	0.0405	0.2159	0.1709
TRP	2	1	0	0	0.0346	0.0541	0.0200	0.0385
TRP	2	2	0	0	0.0181	0.0482	0.0004	0.1538
TRP	2	3	0	0	0.1040	0.2500	0.0196	0.1282
TRP	3	1	0	0	0.1265	0.0179	0.2100	0.1538
TRP	3	2	0	0	0.1160	0.2824	0.0106	0.0769
TRP	3	3	0	0	0.3853	0.2770	0.2944	0.0855
TYR	1	1	0	0	0.4655	0.2576	0.4130	0.2056
TYR	1	2	0	0	0.1170	0.0583	0.0026	0.0204
TYR	2	1	0	0	0.1301	0.1667	0.0272	0.1845
TYR	2	2	0	0	0.0022	0.0167	0.0001	0.0714
TYR	3	1	0	0	0.2716	0.4091	0.5384	0.3548
TYR	3	2	0	0	0.0136	0.0917	0.0188	0.1633
VAL	1	0	0	0	0.0793	0.1143	0.0749	0.0610
VAL	2	0	0	0	0.7612	0.7600	0.8241	0.6829
VAL	3	0	0	0	0.1595	0.1257	0.1010	0.2561

Table C.2.: Probabilities for different rotamers from new compiled unbound rotamer libraries

AS	r1	r2	r3	r4	$P(r1234)$ ALL	$P(r1234)$ HELIX	$P(r1234)$ SHEET	$P(r1234)$ RND
ARG	1	1	1	1	0.0017	0.0009	0.0006	0.0005
ARG	1	1	1	2	0.0060	0.0060	0.0025	0.0041
ARG	1	1	1	3	0.0013	0.0017	0.0051	0.0020
ARG	1	1	2	1	0.0056	0.0039	0.0024	0.0043
ARG	1	1	2	2	0.0090	0.0335	0.0169	0.0044
ARG	1	1	2	3	0.0047	0.0060	0.0129	0.0133
ARG	1	1	3	1	0.0002	0.0023	0.0020	0.0014
ARG	1	1	3	2	0.0073	0.0090	0.0042	0.0065
ARG	1	1	3	3	0.0021	0.0030	0.0134	0.0065
ARG	1	2	1	1	0.0060	0.0060	0.0075	0.0037
ARG	1	2	1	2	0.0068	0.0060	0.0085	0.0131
ARG	1	2	1	3	0.0030	0.0089	0.0129	0.0086
ARG	1	2	2	1	0.0107	0.0149	0.0169	0.0086
ARG	1	2	2	2	0.0155	0.0152	0.0169	0.0174
ARG	1	2	2	3	0.0141	0.0298	0.0129	0.0133
ARG	1	2	3	1	0.0025	0.0055	0.0029	0.0072
ARG	1	2	3	2	0.0197	0.0120	0.0127	0.0087
ARG	1	2	3	3	0.0103	0.0089	0.0134	0.0140
ARG	1	3	1	1	0.0001	0.0040	0.0010	0.0010
ARG	1	3	1	2	0.0009	0.0040	0.0085	0.0087
ARG	1	3	1	3	0.0009	0.0020	0.0043	0.0022
ARG	1	3	2	1	0.0056	0.0082	0.0085	0.0086
ARG	1	3	2	2	0.0086	0.0091	0.0212	0.0174
ARG	1	3	2	3	0.0073	0.0149	0.0086	0.0133
ARG	1	3	3	1	0.0008	0.0002	0.0008	0.0019
ARG	1	3	3	2	0.0115	0.0060	0.0181	0.0218
ARG	1	3	3	3	0.0034	0.0060	0.0089	0.0129
ARG	2	1	1	1	0.0073	0.0030	0.0040	0.0030
ARG	2	1	1	2	0.0103	0.0210	0.0169	0.0131
ARG	2	1	1	3	0.0068	0.0089	0.0043	0.0043
ARG	2	1	2	1	0.0068	0.0030	0.0085	0.0049
ARG	2	1	2	2	0.0240	0.0152	0.0254	0.0218
ARG	2	1	2	3	0.0226	0.0149	0.0043	0.0044
ARG	2	1	3	1	0.0004	0.0027	0.0046	0.0057
ARG	2	1	3	2	0.0107	0.0120	0.0169	0.0087
ARG	2	1	3	3	0.0085	0.0030	0.0045	0.0172
ARG	2	2	1	1	0.0137	0.0298	0.0188	0.0111
ARG	2	2	1	2	0.0240	0.0030	0.0297	0.0131
ARG	2	2	1	3	0.0107	0.0060	0.0043	0.0216
ARG	2	2	2	1	0.0188	0.0119	0.0381	0.0302
ARG	2	2	2	2	0.0422	0.0152	0.0508	0.0392
ARG	2	2	2	3	0.0397	0.0298	0.0516	0.0133
ARG	2	2	3	1	0.0117	0.0218	0.0064	0.0072
ARG	2	2	3	2	0.0432	0.0210	0.0254	0.0435
ARG	2	2	3	3	0.0295	0.0298	0.0313	0.0093
ARG	2	3	1	1	0.0026	0.0089	0.0015	0.0025
ARG	2	3	1	2	0.0098	0.0150	0.0042	0.0087
ARG	2	3	1	3	0.0064	0.0030	0.0043	0.0043
ARG	2	3	2	1	0.0115	0.0027	0.0085	0.0086
ARG	2	3	2	2	0.0467	0.0335	0.0424	0.0305
ARG	2	3	2	3	0.0299	0.0149	0.0043	0.0398
ARG	2	3	3	1	0.0028	0.0014	0.0020	0.0035

AS	r1	r2	r3	r4	$P(r1234)$ ALL	$P(r1234)$ HELIX	$P(r1234)$ SHEET	$P(r1234)$ RND
ARG	2	3	3	2	0.0376	0.0330	0.0362	0.0174
ARG	2	3	3	3	0.0244	0.0417	0.0134	0.0302
ARG	3	1	1	1	0.0009	0.0021	0.0004	0.0005
ARG	3	1	1	2	0.0030	0.0120	0.0017	0.0046
ARG	3	1	1	3	0.0013	0.0043	0.0035	0.0023
ARG	3	1	2	1	0.0034	0.0021	0.0019	0.0037
ARG	3	1	2	2	0.0073	0.0122	0.0085	0.0131
ARG	3	1	2	3	0.0051	0.0060	0.0043	0.0133
ARG	3	1	3	1	0.0002	0.0009	0.0020	0.0014
ARG	3	1	3	2	0.0034	0.0060	0.0127	0.0065
ARG	3	1	3	3	0.0051	0.0060	0.0045	0.0065
ARG	3	2	1	1	0.0056	0.0030	0.0075	0.0111
ARG	3	2	1	2	0.0227	0.0210	0.0085	0.0087
ARG	3	2	1	3	0.0056	0.0030	0.0129	0.0043
ARG	3	2	2	1	0.0188	0.0179	0.0085	0.0216
ARG	3	2	2	2	0.0430	0.0274	0.0763	0.0261
ARG	3	2	2	3	0.0312	0.0179	0.0129	0.0354
ARG	3	2	3	1	0.0054	0.0055	0.0035	0.0072
ARG	3	2	3	2	0.0252	0.0210	0.0127	0.0348
ARG	3	2	3	3	0.0252	0.0387	0.0134	0.0327
ARG	3	3	1	1	0.0008	0.0020	0.0010	0.0010
ARG	3	3	1	2	0.0051	0.0020	0.0042	0.0087
ARG	3	3	1	3	0.0013	0.0010	0.0043	0.0022
ARG	3	3	2	1	0.0167	0.0218	0.0212	0.0086
ARG	3	3	2	2	0.0377	0.0213	0.0254	0.0348
ARG	3	3	2	3	0.0188	0.0357	0.0129	0.0221
ARG	3	3	3	1	0.0036	0.0014	0.0056	0.0032
ARG	3	3	3	2	0.0240	0.0449	0.0136	0.0348
ARG	3	3	3	3	0.0115	0.0298	0.0089	0.0086
ASN	1	1	0	0	0.0357	0.0130	0.0417	0.0612
ASN	1	2	0	0	0.0702	0.0195	0.2346	0.1033
ASN	1	3	0	0	0.0649	0.0656	0.0417	0.0561
ASN	2	1	0	0	0.0984	0.0325	0.2396	0.1020
ASN	2	2	0	0	0.1322	0.0065	0.0853	0.1653
ASN	2	3	0	0	0.1346	0.2296	0.1250	0.1378
ASN	3	1	0	0	0.0366	0.0455	0.0312	0.0561
ASN	3	2	0	0	0.1625	0.2468	0.1280	0.1395
ASN	3	3	0	0	0.2649	0.3411	0.0729	0.1786
ASP	1	1	0	0	0.0265	0.0233	0.0369	0.0139
ASP	1	2	0	0	0.0911	0.0698	0.0645	0.1111
ASP	1	3	0	0	0.0647	0.1073	0.0484	0.0236
ASP	2	1	0	0	0.1043	0.1047	0.0369	0.1019
ASP	2	2	0	0	0.2724	0.0465	0.1129	0.3611
ASP	2	3	0	0	0.1174	0.1789	0.1452	0.1225
ASP	3	1	0	0	0.0521	0.0698	0.0553	0.0324
ASP	3	2	0	0	0.1302	0.2209	0.1452	0.1157
ASP	3	3	0	0	0.1414	0.1789	0.3548	0.1178
CYS	1	0	0	0	0.1161	0.0532	0.0361	0.1552
CYS	2	0	0	0	0.1726	0.4255	0.0361	0.2069
CYS	3	0	0	0	0.7113	0.5213	0.9277	0.6379
HIS	1	1	0	0	0.1922	0.3576	0.1833	0.0606
HIS	1	2	0	0	0.0752	0.0312	0.1579	0.0787
HIS	2	1	0	0	0.2351	0.1273	0.3000	0.3485
HIS	2	2	0	0	0.0301	0.0208	0.0351	0.0629
HIS	3	1	0	0	0.3078	0.3692	0.1833	0.1818
HIS	3	2	0	0	0.1597	0.0938	0.1404	0.2675
GLN	1	1	1	0	0.0017	0.0067	0.0110	0.0196
GLN	1	1	2	0	0.0196	0.0203	0.0183	0.0294
GLN	1	1	3	0	0.0185	0.0067	0.0185	0.0201
GLN	1	2	1	0	0.0257	0.0210	0.0256	0.0294
GLN	1	2	2	0	0.0433	0.0541	0.0403	0.0490
GLN	1	2	3	0	0.0201	0.0267	0.0111	0.0201
GLN	1	3	1	0	0.0028	0.0069	0.0021	0.0115
GLN	1	3	2	0	0.0521	0.0203	0.0449	0.0415
GLN	1	3	3	0	0.0313	0.0333	0.0259	0.0401
GLN	2	1	1	0	0.0268	0.0067	0.0403	0.0098
GLN	2	1	2	0	0.0526	0.0744	0.0440	0.0098
GLN	2	1	3	0	0.0587	0.0333	0.0444	0.0401
GLN	2	2	1	0	0.0464	0.0630	0.0513	0.0392
GLN	2	2	2	0	0.0551	0.0676	0.0659	0.0392
GLN	2	2	3	0	0.0503	0.0800	0.0629	0.0502
GLN	2	3	1	0	0.0190	0.0067	0.0073	0.0179
GLN	2	3	2	0	0.0526	0.0811	0.0524	0.0519
GLN	2	3	3	0	0.0492	0.0467	0.0518	0.0803
GLN	3	1	1	0	0.0112	0.0200	0.0073	0.0098
GLN	3	1	2	0	0.0168	0.0068	0.0147	0.0196
GLN	3	1	3	0	0.0089	0.0200	0.0074	0.0301
GLN	3	2	1	0	0.0666	0.0560	0.0586	0.0294

AS	r1	r2	r3	r4	$P(r1234)$ ALL	$P(r1234)$ HELIX	$P(r1234)$ SHEET	$P(r1234)$ RND
GLN	3	2	2	0	0.0534	0.0879	0.0733	0.0588
GLN	3	2	3	0	0.0503	0.0333	0.0407	0.0703
GLN	3	3	1	0	0.0151	0.0131	0.0052	0.0294
GLN	3	3	2	0	0.0655	0.0744	0.0749	0.0830
GLN	3	3	3	0	0.0861	0.0333	0.0999	0.0703
GLU	1	1	1	0	0.0048	0.0008	0.0183	0.0010
GLU	1	1	2	0	0.0110	0.0051	0.0283	0.0128
GLU	1	1	3	0	0.0199	0.0156	0.0067	0.0131
GLU	1	2	1	0	0.0171	0.0154	0.0275	0.0192
GLU	1	2	2	0	0.0364	0.0205	0.0965	0.0577
GLU	1	2	3	0	0.0316	0.0156	0.0175	0.0131
GLU	1	3	1	0	0.0117	0.0045	0.0105	0.0071
GLU	1	3	2	0	0.0358	0.0410	0.0175	0.0449
GLU	1	3	3	0	0.0282	0.0312	0.0100	0.0392
GLU	2	1	1	0	0.0151	0.0044	0.0183	0.0064
GLU	2	1	2	0	0.0370	0.0359	0.0378	0.0449
GLU	2	1	3	0	0.0412	0.0416	0.0175	0.0523
GLU	2	2	1	0	0.0370	0.0513	0.0275	0.0641
GLU	2	2	2	0	0.1049	0.1436	0.2193	0.1090
GLU	2	2	3	0	0.0680	0.0885	0.0877	0.0327
GLU	2	3	1	0	0.0412	0.0293	0.0549	0.0494
GLU	2	3	2	0	0.0599	0.0410	0.0702	0.0449
GLU	2	3	3	0	0.0502	0.0677	0.0088	0.0327
GLU	3	1	1	0	0.0158	0.0103	0.0092	0.0054
GLU	3	1	2	0	0.0370	0.0103	0.0567	0.0321
GLU	3	1	3	0	0.0144	0.0208	0.0108	0.0327
GLU	3	2	1	0	0.0391	0.0205	0.0366	0.0385
GLU	3	2	2	0	0.0686	0.1179	0.0263	0.0385
GLU	3	2	3	0	0.0515	0.0364	0.0439	0.0849
GLU	3	3	1	0	0.0144	0.0072	0.0078	0.0141
GLU	3	3	2	0	0.0641	0.0974	0.0263	0.0705
GLU	3	3	3	0	0.0440	0.0260	0.0075	0.0392
ILE	1	1	0	0	0.0230	0.0174	0.0280	0.0049
ILE	1	2	0	0	0.0813	0.0349	0.0600	0.1439
ILE	1	3	0	0	0.0464	0.0412	0.0484	0.0155
ILE	2	1	0	0	0.1040	0.0640	0.0680	0.1364
ILE	2	2	0	0	0.1030	0.0465	0.0800	0.1818
ILE	2	3	0	0	0.2482	0.3060	0.2826	0.2018
ILE	3	1	0	0	0.0242	0.0174	0.0400	0.0103
ILE	3	2	0	0	0.2705	0.3314	0.2920	0.2045
ILE	3	3	0	0	0.0994	0.1412	0.1009	0.1009
LEU	1	1	0	0	0.0122	0.0100	0.0355	0.0085
LEU	1	2	0	0	0.0661	0.1000	0.0955	0.0256
LEU	1	3	0	0	0.1347	0.1528	0.1242	0.1295
LEU	2	1	0	0	0.1528	0.2100	0.1677	0.0897
LEU	2	2	0	0	0.1022	0.1500	0.1019	0.0684
LEU	2	3	0	0	0.1830	0.1019	0.1624	0.2590
LEU	3	1	0	0	0.0478	0.0400	0.0516	0.0556
LEU	3	2	0	0	0.2699	0.2200	0.2357	0.3291
LEU	3	3	0	0	0.0314	0.0153	0.0255	0.0345
LYS	1	1	1	1	0.0004	0.0006	0.0001	0.0001
LYS	1	1	1	2	0.0010	0.0023	0.0011	0.0008
LYS	1	1	1	3	0.0007	0.0023	0.0004	0.0004
LYS	1	1	2	1	0.0010	0.0006	0.0033	0.0009
LYS	1	1	2	2	0.0046	0.0078	0.0017	0.0099
LYS	1	1	2	3	0.0023	0.0013	0.0058	0.0067
LYS	1	1	3	1	0.0005	0.0010	0.0001	0.0008
LYS	1	1	3	2	0.0020	0.0053	0.0015	0.0025
LYS	1	1	3	3	0.0010	0.0016	0.0029	0.0017
LYS	1	2	1	1	0.0010	0.0042	0.0032	0.0010
LYS	1	2	1	2	0.0053	0.0052	0.0083	0.0090
LYS	1	2	1	3	0.0026	0.0006	0.0036	0.0006
LYS	1	2	2	1	0.0026	0.0052	0.0056	0.0066
LYS	1	2	2	2	0.0262	0.0417	0.0345	0.0197
LYS	1	2	2	3	0.0106	0.0132	0.0058	0.0200
LYS	1	2	3	1	0.0033	0.0045	0.0009	0.0032
LYS	1	2	3	2	0.0109	0.0053	0.0176	0.0167
LYS	1	2	3	3	0.0050	0.0046	0.0087	0.0067
LYS	1	3	1	1	0.0007	0.0010	0.0001	0.0005
LYS	1	3	1	2	0.0023	0.0017	0.0005	0.0058
LYS	1	3	1	3	0.0007	0.0035	0.0003	0.0020
LYS	1	3	2	1	0.0010	0.0035	0.0005	0.0013
LYS	1	3	2	2	0.0079	0.0241	0.0059	0.0101
LYS	1	3	2	3	0.0033	0.0079	0.0087	0.0067
LYS	1	3	3	1	0.0010	0.0007	0.0004	0.0010
LYS	1	3	3	2	0.0053	0.0040	0.0059	0.0067
LYS	1	3	3	3	0.0017	0.0010	0.0087	0.0033
LYS	2	1	1	1	0.0007	0.0020	0.0008	0.0005

AS	r1	r2	r3	r4	$P(r1234)$ ALL	$P(r1234)$ HELIX	$P(r1234)$ SHEET	$P(r1234)$ RND
LYS	2	1	1	2	0.0050	0.0026	0.0065	0.0066
LYS	2	1	1	3	0.0017	0.0026	0.0022	0.0017
LYS	2	1	2	1	0.0053	0.0041	0.0140	0.0033
LYS	2	1	2	2	0.0265	0.0469	0.0431	0.0164
LYS	2	1	2	3	0.0116	0.0026	0.0087	0.0133
LYS	2	1	3	1	0.0020	0.0024	0.0005	0.0033
LYS	2	1	3	2	0.0083	0.0186	0.0059	0.0033
LYS	2	1	3	3	0.0060	0.0037	0.0029	0.0067
LYS	2	2	1	1	0.0040	0.0042	0.0030	0.0022
LYS	2	2	1	2	0.0232	0.0365	0.0580	0.0090
LYS	2	2	1	3	0.0122	0.0158	0.0070	0.0106
LYS	2	2	2	1	0.0225	0.0443	0.0391	0.0164
LYS	2	2	2	2	0.1562	0.1849	0.2241	0.0559
LYS	2	2	2	3	0.0820	0.0686	0.0554	0.0798
LYS	2	2	3	1	0.0073	0.0067	0.0030	0.0113
LYS	2	2	3	2	0.0787	0.0689	0.0528	0.0770
LYS	2	2	3	3	0.0278	0.0115	0.0117	0.0333
LYS	2	3	1	1	0.0010	0.0029	0.0007	0.0015
LYS	2	3	1	2	0.0103	0.0052	0.0036	0.0146
LYS	2	3	1	3	0.0020	0.0079	0.0018	0.0060
LYS	2	3	2	1	0.0162	0.0130	0.0028	0.0164
LYS	2	3	2	2	0.0755	0.0588	0.0622	0.0777
LYS	2	3	2	3	0.0159	0.0132	0.0204	0.0200
LYS	2	3	3	1	0.0030	0.0017	0.0014	0.0044
LYS	2	3	3	2	0.0291	0.0133	0.0235	0.0368
LYS	2	3	3	3	0.0099	0.0013	0.0029	0.0033
LYS	3	1	1	1	0.0006	0.0008	0.0001	0.0002
LYS	3	1	1	2	0.0013	0.0029	0.0011	0.0025
LYS	3	1	1	3	0.0010	0.0030	0.0004	0.0013
LYS	3	1	2	1	0.0026	0.0006	0.0022	0.0024
LYS	3	1	2	2	0.0083	0.0052	0.0011	0.0132
LYS	3	1	2	3	0.0053	0.0013	0.0058	0.0067
LYS	3	1	3	1	0.0005	0.0017	0.0001	0.0025
LYS	3	1	3	2	0.0020	0.0133	0.0015	0.0075
LYS	3	1	3	3	0.0010	0.0026	0.0029	0.0050
LYS	3	2	1	1	0.0033	0.0021	0.0002	0.0034
LYS	3	2	1	2	0.0152	0.0130	0.0055	0.0181
LYS	3	2	1	3	0.0017	0.0021	0.0010	0.0021
LYS	3	2	2	1	0.0142	0.0130	0.0279	0.0132
LYS	3	2	2	2	0.0786	0.0703	0.0632	0.0822
LYS	3	2	2	3	0.0212	0.0158	0.0058	0.0266
LYS	3	2	3	1	0.0030	0.0045	0.0015	0.0052
LYS	3	2	3	2	0.0179	0.0159	0.0205	0.0268
LYS	3	2	3	3	0.0093	0.0023	0.0117	0.0133
LYS	3	3	1	1	0.0007	0.0005	0.0003	0.0005
LYS	3	3	1	2	0.0040	0.0009	0.0016	0.0058
LYS	3	3	1	3	0.0007	0.0018	0.0008	0.0020
LYS	3	3	2	1	0.0030	0.0017	0.0023	0.0053
LYS	3	3	2	2	0.0305	0.0160	0.0296	0.0372
LYS	3	3	2	3	0.0109	0.0079	0.0058	0.0166
LYS	3	3	3	1	0.0017	0.0002	0.0006	0.0012
LYS	3	3	3	2	0.0099	0.0013	0.0088	0.0134
LYS	3	3	3	3	0.0036	0.0003	0.0029	0.0100
MET	1	1	1	0	0.0048	0.0038	0.0036	0.0104
MET	1	1	2	0	0.0112	0.0025	0.0108	0.0208
MET	1	1	3	0	0.0112	0.0161	0.0044	0.0312
MET	1	2	1	0	0.0096	0.0030	0.0022	0.0347
MET	1	2	2	0	0.0159	0.0021	0.0323	0.0260
MET	1	2	3	0	0.0144	0.0161	0.0111	0.0391
MET	1	3	1	0	0.0048	0.0019	0.0042	0.1172
MET	1	3	2	0	0.0766	0.0952	0.0459	0.0391
MET	1	3	3	0	0.0258	0.0250	0.0222	0.0391
MET	2	1	1	0	0.0175	0.0188	0.0125	0.0417
MET	2	1	2	0	0.0128	0.0125	0.0215	0.0208
MET	2	1	3	0	0.0736	0.0968	0.0333	0.0312
MET	2	2	1	0	0.0431	0.0193	0.0659	0.0208
MET	2	2	2	0	0.0734	0.0138	0.1075	0.0156
MET	2	2	3	0	0.0576	0.0887	0.0999	0.0234
MET	2	3	1	0	0.0191	0.0020	0.0135	0.0703
MET	2	3	2	0	0.0510	0.0794	0.0803	0.0234
MET	2	3	3	0	0.0452	0.0501	0.0444	0.0234
MET	3	1	1	0	0.0048	0.0013	0.0054	0.0312
MET	3	1	2	0	0.0159	0.0008	0.0215	0.0625
MET	3	1	3	0	0.0224	0.0081	0.0067	0.0938
MET	3	2	1	0	0.0526	0.0967	0.0072	0.0278
MET	3	2	2	0	0.0845	0.0873	0.1613	0.0208
MET	3	2	3	0	0.0432	0.0242	0.0333	0.0313
MET	3	3	1	0	0.0191	0.0040	0.0146	0.0625

AS	r1	r2	r3	r4	$P(r1234)$ ALL	$P(r1234)$ HELIX	$P(r1234)$ SHEET	$P(r1234)$ RND
MET	3	3	2	0	0.0526	0.0635	0.0459	0.0208
MET	3	3	3	0	0.1371	0.1669	0.0888	0.0208
PHE	1	1	0	0	0.3430	0.2416	0.4631	0.1939
PHE	1	2	0	0	0.0538	0.1298	0.0100	0.0266
PHE	2	1	0	0	0.1228	0.2167	0.0539	0.0753
PHE	2	2	0	0	0.0012	0.0096	0.0067	0.0106
PHE	3	1	0	0	0.4115	0.3301	0.4296	0.6563
PHE	3	2	0	0	0.0677	0.0721	0.0367	0.0372
SER	1	0	0	0	0.4801	0.5347	0.2970	0.5235
SER	2	0	0	0	0.2250	0.1386	0.4950	0.2282
SER	3	0	0	0	0.2949	0.3267	0.2079	0.2483
THR	1	0	0	0	0.4427	0.2388	0.2080	0.4694
THR	2	0	0	0	0.0513	0.0597	0.1680	0.0510
THR	3	0	0	0	0.5060	0.7015	0.6240	0.4796
TRP	1	1	0	0	0.0664	0.0273	0.1261	0.0750
TRP	1	2	0	0	0.0201	0.0094	0.0087	0.0500
TRP	1	3	0	0	0.1408	0.0364	0.1471	0.2912
TRP	2	1	0	0	0.0185	0.0182	0.0280	0.0750
TRP	2	2	0	0	0.0262	0.0283	0.0009	0.0163
TRP	2	3	0	0	0.1021	0.2182	0.0294	0.0529
TRP	3	1	0	0	0.1435	0.0182	0.1401	0.2750
TRP	3	2	0	0	0.1003	0.2168	0.0199	0.0587
TRP	3	3	0	0	0.3821	0.4273	0.5000	0.1059
TYR	1	1	0	0	0.2702	0.1421	0.3000	0.1395
TYR	1	2	0	0	0.0598	0.1379	0.0664	0.1570
TYR	2	1	0	0	0.1288	0.1995	0.1000	0.2384
TYR	2	2	0	0	0.0083	0.0172	0.0066	0.0725
TYR	3	1	0	0	0.4640	0.4343	0.4937	0.3081
TYR	3	2	0	0	0.0689	0.0690	0.0332	0.0845
VAL	1	0	0	0	0.0783	0.0571	0.0950	0.1279
VAL	2	0	0	0	0.7331	0.7810	0.7285	0.6512
VAL	3	0	0	0	0.1886	0.1619	0.1765	0.2209

Table C.3.: Probabilities for different rotamers from new compiled complex rotamer libraries

D. Flexibility

AA	probabilities for change in direction						prob
	r1→2	r1→3	r2→1	r2→3	r3→1	r3→2	all
ARG	0.8276	0.0690	0.2426	0.3069	0.0000	0.0357	0.4226
ASN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0012	0.0012
ASP	0.0000	0.2000	0.0000	0.1667	0.0000	0.7651	0.7288
CYS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GLN	0.0000	0.0000	0.0000	0.0861	0.0233	0.7733	0.3645
GLU	0.0000	1.0000	0.0063	0.0125	0.0000	0.0124	0.0307
HIS	0.0000	0.0000	0.0000	0.0000	0.1111	0.0000	0.0038
ILE	0.0030	0.6647	0.0000	0.2222	0.0021	0.0043	0.2822
LEU	0.0030	0.0000	0.0000	0.0248	0.0000	0.0245	0.0247
LYS	0.6000	0.2000	0.0055	0.0801	0.0000	0.1157	0.1069
MET	0.6000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PHE	0.6000	0.0000	0.0000	0.1290	0.0000	0.0063	0.0265
SER	0.0228	0.2066	0.3333	0.1667	0.3569	0.0216	0.2918
THR	0.0833	0.2500	0.0000	1.0000	0.0000	0.0000	0.0189
TRP	0.0000	0.0000	0.0025	0.0025	0.0036	0.0000	0.0044
TYR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
VAL	0.7945	0.0274	0.0063	0.0021	0.0000	0.0134	0.1041

Table D.1.: Direction of change for χ_1 rotamers in helices

AA	probability for change in direction						prob
	r1→2	r1→3	r2→	r2→3	r3→1	r3→2	all
ARG	0.0000	0.0000	0.0183	0.4404	0.0035	0.0456	0.1612
ASN	0.3571	0.0000	0.0000	0.0140	0.0087	0.7739	0.2340
ASP	0.0000	0.0000	0.0000	0.1429	0.0000	0.0278	0.0323
CYS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GLN	0.0000	0.0000	0.0038	0.0038	0.0000	0.0630	0.0368
GLU	0.2727	0.0000	0.1429	0.0000	0.0000	0.7500	0.3182
HIS	0.2727	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ILE	0.0000	0.0000	0.0000	0.0000	0.0021	0.0007	0.0024
LEU	0.0000	0.0000	0.0000	0.0039	0.0000	0.0172	0.0106
LYS	0.0714	0.8571	0.0023	0.0700	0.0124	0.0041	0.2242
MET	0.0714	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PHE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
SER	0.0000	0.0000	0.0013	0.0340	0.0543	0.0231	0.0421
THR	0.0667	0.6133	0.0000	0.1250	0.0045	0.0030	0.1309
TRP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TYR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
VAL	0.0968	0.0121	0.0078	0.0959	0.0328	0.0820	0.1045

Table D.2.: Directions of change for χ_1 rotamers in sheets

AA	r1	r2	r3	r4	# combination	# change	P(change)
ARG	1	1	1	2	10	2	0.2000
ARG	1	2	1	2	4	0	0.0000
ARG	1	2	1	3	1	1	1.0000
ARG	1	2	2	1	16	14	0.8750
ARG	1	2	2	2	130	76	0.5846
ARG	1	2	2	3	70	65	0.9286
ARG	1	2	3	1	24	14	0.5833
ARG	1	2	3	2	2	1	0.5000
ARG	1	2	3	3	4	0	0.0000
ARG	2	1	1	1	24	12	0.5000
ARG	2	1	1	2	72	10	0.1389
ARG	2	1	2	1	25	15	0.6000
ARG	2	1	2	2	1	0	0.0000
ARG	2	1	2	3	74	30	0.4054
ARG	2	1	3	2	35	16	0.4571
ARG	2	1	3	3	2	1	0.5000
ARG	2	2	1	1	264	165	0.6250
ARG	2	2	1	2	120	74	0.6167
ARG	2	2	1	3	3	3	1.0000
ARG	2	2	2	1	37	14	0.3784
ARG	2	2	2	2	45	21	0.4667
ARG	2	2	2	3	33	25	0.7576
ARG	2	2	3	1	3	2	0.6667
ARG	2	2	3	2	163	49	0.3006
ARG	2	2	3	3	65	9	0.1385
ARG	2	3	2	1	1	0	0.0000
ARG	2	3	2	2	4	0	0.0000
ARG	3	1	1	2	2	2	1.0000
ARG	3	1	2	1	1	0	0.0000
ARG	3	1	2	2	9	1	0.1111
ARG	3	1	3	2	18	1	0.0556
ARG	3	1	3	3	4	1	0.2500
ARG	3	2	1	1	64	14	0.2188
ARG	3	2	1	2	160	8	0.0500
ARG	3	2	1	3	21	3	0.1429
ARG	3	2	2	1	22	5	0.2273
ARG	3	2	2	2	422	7	0.0166
ARG	3	2	2	3	79	19	0.2405
ARG	3	2	3	1	16	9	0.5625
ARG	3	2	3	2	68	12	0.1765
ARG	3	2	3	3	75	19	0.2533
ARG	3	3	1	1	11	6	0.5455
ARG	3	3	1	2	44	15	0.3409
ARG	3	3	2	1	21	2	0.0952
ARG	3	3	2	2	91	4	0.0440
ARG	3	3	2	3	49	7	0.1429
ARG	3	3	3	2	44	0	0.0000
ARG	3	3	3	3	83	7	0.0843
ASN	1	1	0	0	385	94	0.2442
ASN	1	2	0	0	196	63	0.3214
ASN	1	3	0	0	41	27	0.6585
ASN	2	1	0	0	689	47	0.0682
ASN	2	2	0	0	1165	35	0.0300
ASN	2	3	0	0	1261	0	0.0000
ASN	3	1	0	0	22	9	0.4091
ASN	3	2	0	0	829	26	0.0314
ASN	3	3	0	0	3309	377	0.1139
ASP	1	1	0	0	10	3	0.3000
ASP	1	2	0	0	379	7	0.0185
ASP	1	3	0	0	51	30	0.5882
ASP	2	1	0	0	464	82	0.1767
ASP	2	2	0	0	1401	1	0.0007
ASP	2	3	0	0	6	4	0.6667
ASP	3	1	0	0	17	14	0.8235
ASP	3	2	0	0	176	43	0.2443
ASP	3	3	0	0	1112	181	0.1628
CYS	1	0	0	0	580	0	0.0000
CYS	2	0	0	0	1172	6	0.0051
CYS	3	0	0	0	4569	1	0.0002
GLN	1	1	2	0	14	14	1.0000
GLN	1	2	1	0	162	155	0.9568
GLN	1	2	2	0	56	56	1.0000
GLN	1	2	3	0	7	5	0.7143
GLN	1	3	2	0	14	14	1.0000
GLN	1	3	3	0	29	29	1.0000
GLN	2	1	1	0	426	5	0.0117
GLN	2	1	2	0	22	11	0.5000
GLN	2	1	3	0	7	1	0.1429

AA	r1	r2	r3	r4	# combination	# change	P(change)
GLN	2	2	1	0	578	24	0.0415
GLN	2	2	2	0	99	50	0.5051
GLN	2	2	3	0	45	5	0.1111
GLN	2	3	1	0	15	15	1.0000
GLN	2	3	2	0	28	2	0.0714
GLN	2	3	3	0	57	4	0.0702
GLN	3	1	1	0	74	28	0.3784
GLN	3	1	2	0	39	15	0.3846
GLN	3	1	3	0	4	2	0.5000
GLN	3	2	1	0	802	141	0.1758
GLN	3	2	2	0	146	67	0.4589
GLN	3	2	3	0	748	224	0.2995
GLN	3	3	1	0	63	4	0.0635
GLN	3	3	2	0	449	10	0.0223
GLN	3	3	3	0	566	60	0.1060
GLU	1	2	1	0	25	11	0.4400
GLU	1	2	2	0	2	2	1.0000
GLU	1	2	3	0	9	3	0.3333
GLU	1	3	1	0	32	18	0.5625
GLU	1	3	2	0	28	13	0.4643
GLU	1	3	3	0	66	40	0.6061
GLU	2	1	1	0	7	3	0.4286
GLU	2	1	2	0	1	0	0.0000
GLU	2	1	3	0	14	0	0.0000
GLU	2	2	1	0	2	0	0.0000
GLU	2	2	2	0	155	1	0.0065
GLU	2	2	3	0	25	0	0.0000
GLU	2	3	1	0	3	1	0.3333
GLU	2	3	3	0	1	1	1.0000
GLU	3	1	1	0	6	4	0.6667
GLU	3	1	2	0	360	15	0.0417
GLU	3	1	3	0	61	0	0.0000
GLU	3	2	1	0	40	2	0.0500
GLU	3	2	2	0	12	3	0.2500
GLU	3	2	3	0	385	3	0.0078
GLU	3	3	1	0	15	0	0.0000
GLU	3	3	2	0	571	3	0.0053
GLU	3	3	3	0	20	11	0.5500
HIS	1	1	0	0	415	0	0.0000
HIS	2	1	0	0	859	3	0.0035
HIS	2	2	0	0	9	1	0.1111
HIS	3	1	0	0	205	3	0.0146
HIS	3	2	0	0	8	2	0.2500
ILE	1	1	0	0	22	1	0.0455
ILE	1	2	0	0	1497	324	0.2164
ILE	2	1	0	0	511	4	0.0078
ILE	2	2	0	0	568	7	0.0123
ILE	2	3	0	0	2	2	1.0000
ILE	3	1	0	0	51	5	0.0980
ILE	3	2	0	0	3813	3	0.0008
ILE	3	3	0	0	446	3	0.0067
LEU	1	1	0	0	1	1	1.0000
LEU	1	2	0	0	1	0	0.0000
LEU	2	1	0	0	2448	41	0.0167
LEU	2	2	0	0	60	3	0.0500
LEU	2	3	0	0	1	0	0.0000
LEU	3	1	0	0	246	3	0.0122
LEU	3	2	0	0	4443	45	0.0101
LEU	3	3	0	0	49	0	0.0000
LYS	1	2	1	1	1	1	1.0000
LYS	1	2	1	2	51	50	0.9804
LYS	1	2	2	2	6	3	0.5000
LYS	1	2	2	3	2	0	0.0000
LYS	1	2	3	1	14	13	0.9286
LYS	1	2	3	2	274	256	0.9343
LYS	1	2	3	3	10	10	1.0000
LYS	2	1	1	2	14	5	0.3571
LYS	2	1	2	1	3	1	0.3333
LYS	2	1	2	2	173	1	0.0058
LYS	2	1	2	3	4	2	0.5000
LYS	2	2	1	1	3	3	1.0000
LYS	2	2	1	2	16	15	0.9375
LYS	2	2	1	3	31	17	0.5484
LYS	2	2	2	1	99	8	0.0808
LYS	2	2	2	2	1139	138	0.1212
LYS	2	2	2	3	49	2	0.0408
LYS	2	2	3	1	15	6	0.4000
LYS	2	2	3	2	138	56	0.4058

AA	r1	r2	r3	r4	# combination	# change	P(change)
LYS	2	2	3	3	13	3	0.2308
LYS	2	3	2	2	1	1	1.0000
LYS	2	3	3	1	2	0	0.0000
LYS	3	1	1	2	2	1	0.5000
LYS	3	1	2	2	36	0	0.0000
LYS	3	1	2	3	1	0	0.0000
LYS	3	2	1	1	4	0	0.0000
LYS	3	2	1	2	259	170	0.6564
LYS	3	2	1	3	16	0	0.0000
LYS	3	2	2	1	126	18	0.1429
LYS	3	2	2	2	811	160	0.1973
LYS	3	2	2	3	87	45	0.5172
LYS	3	2	3	1	50	1	0.0200
LYS	3	2	3	2	28	3	0.1071
LYS	3	2	3	3	4	2	0.5000
LYS	3	3	1	2	13	0	0.0000
LYS	3	3	2	1	33	1	0.0303
LYS	3	3	2	2	877	3	0.0034
LYS	3	3	2	3	140	3	0.0214
LYS	3	3	3	3	6	2	0.3333
MET	1	2	1	0	4	0	0.0000
MET	1	2	3	0	14	14	1.0000
MET	2	1	1	0	2	0	0.0000
MET	2	2	1	0	400	1	0.0025
MET	2	2	3	0	2	0	0.0000
MET	3	2	1	0	175	0	0.0000
MET	3	2	2	0	582	0	0.0000
MET	3	2	3	0	9	0	0.0000
MET	3	3	1	0	6	0	0.0000
MET	3	3	3	0	5	0	0.0000
PHE	1	1	0	0	830	0	0.0000
PHE	2	1	0	0	47	6	0.1277
PHE	2	2	0	0	4	0	0.0000
PHE	3	1	0	0	551	5	0.0091
PHE	3	2	0	0	449	0	0.0000
SER	1	0	0	0	6454	1080	0.1673
SER	2	0	0	0	2604	459	0.1763
SER	3	0	0	0	4853	1728	0.3561
THR	1	0	0	0	2399	275	0.1146
THR	2	0	0	0	71	51	0.7183
THR	3	0	0	0	2735	127	0.0464
TRP	1	1	0	0	411	0	0.0000
TRP	2	1	0	0	13	1	0.0769
TRP	2	2	0	0	4	0	0.0000
TRP	2	3	0	0	584	1	0.0017
TRP	3	1	0	0	614	0	0.0000
TRP	3	2	0	0	20	1	0.0500
TRP	3	3	0	0	1071	0	0.0000
TYR	1	1	0	0	1240	4	0.0032
TYR	2	1	0	0	1027	0	0.0000
TYR	3	1	0	0	2475	4	0.0016
TYR	3	2	0	0	12	0	0.0000
VAL	1	0	0	0	446	162	0.3632
VAL	2	0	0	0	7094	798	0.1125
VAL	3	0	0	0	785	42	0.0535

Table D.3.: Probabilities for χ_1 rotamer changes depending on the rotamer set

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
ARG	1	1	1	2	2	2	1	2	10	1	0.1000
ARG	1	1	1	2	3	3	3	1	10	1	0.1000
ARG	1	2	1	3	3	2	2	3	1	1	1.0000
ARG	1	2	2	1	1	2	2	3	16	1	0.0625
ARG	1	2	2	1	2	2	1	2	16	1	0.0625
ARG	1	2	2	1	2	2	3	1	16	9	0.5625
ARG	1	2	2	1	2	2	3	3	16	2	0.1250
ARG	1	2	2	1	3	1	1	2	16	1	0.0625
ARG	1	2	2	1	3	2	3	3	16	1	0.0625
ARG	1	2	2	2	1	2	1	3	130	1	0.0077
ARG	1	2	2	2	1	2	2	1	130	9	0.0692
ARG	1	2	2	2	1	2	3	2	130	3	0.0231
ARG	1	2	2	2	1	2	3	3	130	18	0.1385
ARG	1	2	2	2	2	2	1	2	130	1	0.0077
ARG	1	2	2	2	2	2	2	3	130	18	0.1385
ARG	1	2	2	2	2	2	3	2	130	2	0.0154
ARG	1	2	2	2	2	3	2	2	130	9	0.0692
ARG	1	2	2	2	3	2	1	1	130	9	0.0692
ARG	1	2	2	2	3	2	1	3	130	9	0.0692
ARG	1	2	2	2	3	2	3	1	130	9	0.0692
ARG	1	2	2	2	3	2	3	3	130	10	0.0769
ARG	1	2	2	2	3	3	2	2	130	9	0.0692
ARG	1	2	2	3	1	2	1	3	70	3	0.0429
ARG	1	2	2	3	2	2	1	2	70	5	0.0714
ARG	1	2	2	3	2	2	3	1	70	18	0.2571
ARG	1	2	2	3	2	2	3	2	70	6	0.0857
ARG	1	2	2	3	2	2	3	3	70	4	0.0571
ARG	1	2	2	3	3	2	3	3	70	32	0.4571
ARG	1	2	3	1	1	2	2	1	24	2	0.0833
ARG	1	2	3	1	1	2	2	2	24	4	0.1667
ARG	1	2	3	1	1	2	3	3	24	4	0.1667
ARG	1	2	3	1	2	2	2	3	24	4	0.1667
ARG	1	2	3	1	2	3	2	2	24	2	0.0833
ARG	1	2	3	1	3	2	1	1	24	2	0.0833
ARG	1	2	3	1	3	2	1	3	24	2	0.0833
ARG	1	2	3	1	3	2	3	1	24	2	0.0833
ARG	1	2	3	1	3	3	2	2	24	2	0.0833
ARG	1	2	3	2	1	3	2	3	2	1	0.5000
ARG	1	2	3	2	3	2	1	1	2	1	0.5000
ARG	2	1	1	1	2	2	1	1	24	2	0.0833
ARG	2	1	1	1	2	2	2	3	24	2	0.0833
ARG	2	1	1	1	2	2	3	2	24	2	0.0833
ARG	2	1	1	1	2	2	3	3	24	2	0.0833
ARG	2	1	1	1	2	3	3	2	24	2	0.0833
ARG	2	1	1	1	3	2	1	3	24	2	0.0833
ARG	2	1	1	1	3	2	2	1	24	2	0.0833
ARG	2	1	1	1	3	2	2	2	24	2	0.0833
ARG	2	1	1	1	3	3	3	1	24	6	0.2500
ARG	2	1	1	2	2	1	1	1	72	5	0.0694
ARG	2	1	1	2	2	2	1	1	72	1	0.0139
ARG	2	1	1	2	2	2	2	3	72	1	0.0139
ARG	2	1	1	2	2	2	3	1	72	12	0.1667
ARG	2	1	1	2	2	2	3	2	72	25	0.3472
ARG	2	1	1	2	2	2	3	3	72	13	0.1806
ARG	2	1	1	2	2	3	2	2	72	4	0.0556
ARG	2	1	1	2	2	3	3	2	72	1	0.0139
ARG	2	1	1	2	3	2	1	3	72	1	0.0139
ARG	2	1	1	2	3	2	2	1	72	1	0.0139
ARG	2	1	1	2	3	2	2	2	72	1	0.0139
ARG	2	1	1	2	3	3	2	2	72	2	0.0278
ARG	2	1	1	2	3	3	2	3	72	2	0.0278
ARG	2	1	1	2	3	3	3	1	72	3	0.0417
ARG	2	1	2	1	1	2	1	3	25	1	0.0400
ARG	2	1	2	1	2	1	1	1	25	1	0.0400
ARG	2	1	2	1	2	2	1	2	25	5	0.2000
ARG	2	1	2	1	2	2	1	3	25	1	0.0400
ARG	2	1	2	1	2	2	3	1	25	1	0.0400
ARG	2	1	2	1	2	2	3	2	25	2	0.0800
ARG	2	1	2	1	3	1	2	3	25	1	0.0400
ARG	2	1	2	1	3	2	3	3	25	10	0.4000
ARG	2	1	2	1	3	3	1	2	25	1	0.0400
ARG	2	1	2	1	3	3	2	2	25	1	0.0400
ARG	2	1	2	1	3	3	3	3	25	1	0.0400
ARG	2	1	2	2	2	2	1	2	1	1	1.0000
ARG	2	1	2	3	2	1	1	1	74	6	0.0811
ARG	2	1	2	3	2	2	1	1	74	5	0.0676
ARG	2	1	2	3	2	2	2	3	74	5	0.0676
ARG	2	1	2	3	2	2	3	1	74	3	0.0405

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
ARG	2	1	2	3	2	2	3	2	74	11	0.1486
ARG	2	1	2	3	2	2	3	3	74	8	0.1081
ARG	2	1	2	3	2	3	2	2	74	1	0.0135
ARG	2	1	2	3	2	3	3	2	74	5	0.0676
ARG	2	1	2	3	3	2	1	3	74	5	0.0676
ARG	2	1	2	3	3	2	2	1	74	5	0.0676
ARG	2	1	2	3	3	2	2	2	74	5	0.0676
ARG	2	1	2	3	3	3	3	1	74	15	0.2027
ARG	2	1	3	2	2	1	1	1	35	1	0.0286
ARG	2	1	3	2	2	1	1	2	35	1	0.0286
ARG	2	1	3	2	2	1	2	1	35	1	0.0286
ARG	2	1	3	2	2	2	3	1	35	3	0.0857
ARG	2	1	3	2	2	2	3	2	35	9	0.2571
ARG	2	1	3	2	2	2	3	3	35	3	0.0857
ARG	2	1	3	2	2	3	2	2	35	1	0.0286
ARG	2	1	3	2	3	2	3	2	35	1	0.0286
ARG	2	1	3	2	3	3	1	2	35	1	0.0286
ARG	2	1	3	2	3	3	2	2	35	9	0.2571
ARG	2	1	3	2	3	3	2	3	35	3	0.0857
ARG	2	1	3	2	3	3	3	3	35	2	0.0571
ARG	2	1	3	3	2	2	2	2	2	1	0.5000
ARG	2	1	3	3	3	2	3	3	2	1	0.5000
ARG	2	2	1	1	1	2	2	3	264	109	0.4129
ARG	2	2	1	1	2	2	1	2	264	8	0.0303
ARG	2	2	1	1	2	2	3	1	264	10	0.0379
ARG	2	2	1	1	2	2	3	2	264	5	0.0189
ARG	2	2	1	1	2	2	3	3	264	7	0.0265
ARG	2	2	1	1	3	2	1	2	264	5	0.0189
ARG	2	2	1	1	3	2	2	2	264	5	0.0189
ARG	2	2	1	1	3	2	3	1	264	10	0.0379
ARG	2	2	1	1	3	2	3	3	264	36	0.1364
ARG	2	2	1	2	1	2	1	3	120	2	0.0167
ARG	2	2	1	2	2	2	1	3	120	4	0.0333
ARG	2	2	1	2	2	2	2	2	120	2	0.0167
ARG	2	2	1	2	2	2	3	1	120	5	0.0417
ARG	2	2	1	2	2	2	3	2	120	7	0.0583
ARG	2	2	1	2	2	2	3	3	120	3	0.0250
ARG	2	2	1	2	2	3	2	2	120	2	0.0167
ARG	2	2	1	2	3	1	2	3	120	4	0.0333
ARG	2	2	1	2	3	2	1	2	120	3	0.0250
ARG	2	2	1	2	3	2	2	1	120	1	0.0083
ARG	2	2	1	2	3	2	2	2	120	3	0.0250
ARG	2	2	1	2	3	2	3	1	120	6	0.0500
ARG	2	2	1	2	3	2	3	3	120	41	0.3417
ARG	2	2	1	2	3	3	1	2	120	4	0.0333
ARG	2	2	1	2	3	3	2	1	120	1	0.0083
ARG	2	2	1	2	3	3	2	2	120	4	0.0333
ARG	2	2	1	2	3	3	3	3	120	5	0.0417
ARG	2	2	1	3	3	2	2	2	3	2	0.6667
ARG	2	2	1	3	3	2	2	3	3	1	0.3333
ARG	2	2	2	1	1	1	2	1	37	1	0.0270
ARG	2	2	2	1	2	2	1	2	37	13	0.3514
ARG	2	2	2	1	2	2	1	3	37	3	0.0811
ARG	2	2	2	1	2	2	3	1	37	3	0.0811
ARG	2	2	2	1	3	1	2	3	37	3	0.0811
ARG	2	2	2	1	3	1	3	2	37	1	0.0270
ARG	2	2	2	1	3	3	1	2	37	3	0.0811
ARG	2	2	2	1	3	3	2	2	37	3	0.0811
ARG	2	2	2	1	3	3	3	3	37	3	0.0811
ARG	2	2	2	2	1	2	3	2	45	2	0.0444
ARG	2	2	2	2	2	1	1	1	45	2	0.0444
ARG	2	2	2	2	2	2	1	1	45	1	0.0222
ARG	2	2	2	2	2	2	1	2	45	1	0.0222
ARG	2	2	2	2	2	2	2	3	45	1	0.0222
ARG	2	2	2	2	2	2	3	1	45	4	0.0889
ARG	2	2	2	2	2	2	3	2	45	8	0.1778
ARG	2	2	2	2	2	2	3	3	45	5	0.1111
ARG	2	2	2	2	2	3	2	2	45	1	0.0222
ARG	2	2	2	2	2	3	3	2	45	1	0.0222
ARG	2	2	2	2	3	2	1	2	45	1	0.0222
ARG	2	2	2	2	3	2	1	3	45	1	0.0222
ARG	2	2	2	2	3	2	2	1	45	1	0.0222
ARG	2	2	2	2	3	2	2	2	45	3	0.0667
ARG	2	2	2	2	3	2	2	3	45	1	0.0222
ARG	2	2	2	2	3	2	3	1	45	2	0.0444
ARG	2	2	2	2	3	2	3	3	45	7	0.1556
ARG	2	2	2	2	3	3	3	1	45	3	0.0667
ARG	2	2	2	3	1	2	2	3	33	1	0.0303

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
ARG	2	2	2	3	1	2	3	2	33	1	0.0303
ARG	2	2	2	3	2	2	1	2	33	2	0.0606
ARG	2	2	2	3	2	2	3	1	33	1	0.0303
ARG	2	2	2	3	2	2	3	2	33	2	0.0606
ARG	2	2	2	3	2	2	3	3	33	2	0.0606
ARG	2	2	2	3	3	2	1	2	33	2	0.0606
ARG	2	2	2	3	3	2	2	2	33	2	0.0606
ARG	2	2	2	3	3	2	3	1	33	4	0.1212
ARG	2	2	2	3	3	2	3	3	33	15	0.4545
ARG	2	2	3	1	1	2	2	1	3	1	0.3333
ARG	2	2	3	1	2	2	1	1	3	1	0.3333
ARG	2	2	3	1	3	2	2	1	3	1	0.3333
ARG	2	2	3	2	1	1	2	1	163	1	0.0061
ARG	2	2	3	2	1	2	1	3	163	3	0.0184
ARG	2	2	3	2	1	2	2	3	163	7	0.0429
ARG	2	2	3	2	2	1	1	1	163	1	0.0061
ARG	2	2	3	2	2	2	1	2	163	11	0.0675
ARG	2	2	3	2	2	2	3	1	163	66	0.4049
ARG	2	2	3	2	2	2	3	3	163	17	0.1043
ARG	2	2	3	2	2	3	2	2	163	1	0.0061
ARG	2	2	3	2	3	1	3	2	163	1	0.0061
ARG	2	2	3	2	3	2	3	3	163	37	0.2270
ARG	2	2	3	3	1	2	2	3	65	1	0.0154
ARG	2	2	3	3	2	1	1	1	65	3	0.0462
ARG	2	2	3	3	2	2	1	1	65	1	0.0154
ARG	2	2	3	3	2	2	1	2	65	1	0.0154
ARG	2	2	3	3	2	2	3	1	65	18	0.2769
ARG	2	2	3	3	2	2	3	2	65	18	0.2769
ARG	2	2	3	3	2	3	2	2	65	4	0.0615
ARG	2	2	3	3	3	2	1	1	65	2	0.0308
ARG	2	2	3	3	3	2	1	2	65	1	0.0154
ARG	2	2	3	3	3	2	1	3	65	2	0.0308
ARG	2	2	3	3	3	2	2	1	65	1	0.0154
ARG	2	2	3	3	3	2	3	3	65	1	0.0154
ARG	2	2	3	3	3	3	3	2	65	1	0.0154
ARG	2	3	2	1	2	2	3	2	1	1	1.0000
ARG	2	3	2	2	2	2	1	2	4	3	0.7500
ARG	2	3	2	2	2	2	2	2	4	1	0.2500
ARG	3	1	1	2	1	1	2	2	2	1	0.5000
ARG	3	1	1	2	1	1	3	2	2	1	0.5000
ARG	3	1	2	1	3	3	2	2	1	1	1.0000
ARG	3	1	2	2	1	2	1	1	9	1	0.1111
ARG	3	1	2	2	3	2	2	1	9	2	0.2222
ARG	3	1	2	2	3	2	2	3	9	1	0.1111
ARG	3	1	2	2	3	3	2	2	9	2	0.2222
ARG	3	1	2	2	3	3	2	3	9	2	0.2222
ARG	3	1	2	2	3	3	3	2	9	1	0.1111
ARG	3	1	3	2	1	2	1	1	18	1	0.0556
ARG	3	1	3	2	3	1	2	2	18	2	0.1111
ARG	3	1	3	2	3	2	1	1	18	1	0.0556
ARG	3	1	3	2	3	2	2	1	18	1	0.0556
ARG	3	1	3	2	3	2	2	2	18	1	0.0556
ARG	3	1	3	2	3	3	2	2	18	2	0.1111
ARG	3	1	3	2	3	3	2	3	18	9	0.5000
ARG	3	1	3	2	3	3	3	3	18	1	0.0556
ARG	3	1	3	3	1	2	1	1	4	1	0.2500
ARG	3	1	3	3	3	2	2	1	4	1	0.2500
ARG	3	1	3	3	3	3	2	2	4	2	0.5000
ARG	3	2	1	1	1	2	2	1	64	1	0.0156
ARG	3	2	1	1	2	2	1	1	64	1	0.0156
ARG	3	2	1	1	2	2	1	2	64	8	0.1250
ARG	3	2	1	1	2	2	1	3	64	2	0.0313
ARG	3	2	1	1	2	2	3	1	64	2	0.0313
ARG	3	2	1	1	3	1	2	2	64	4	0.0625
ARG	3	2	1	1	3	1	2	3	64	2	0.0313
ARG	3	2	1	1	3	2	2	2	64	16	0.2500
ARG	3	2	1	1	3	3	1	2	64	2	0.0313
ARG	3	2	1	1	3	3	2	2	64	2	0.0313
ARG	3	2	1	1	3	3	2	3	64	18	0.2813
ARG	3	2	1	1	3	3	3	3	64	4	0.0625
ARG	3	2	1	2	1	2	2	1	160	1	0.0063
ARG	3	2	1	2	1	2	3	2	160	3	0.0188
ARG	3	2	1	2	2	1	2	2	160	2	0.0125
ARG	3	2	1	2	2	2	1	2	160	1	0.0063
ARG	3	2	1	2	2	3	3	2	160	1	0.0063
ARG	3	2	1	2	3	1	2	3	160	10	0.0625
ARG	3	2	1	2	3	2	1	1	160	1	0.0063
ARG	3	2	1	2	3	2	2	3	160	1	0.0063

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
ARG	3	2	1	2	3	2	3	2	160	5	0.0313
ARG	3	2	1	2	3	3	2	1	160	10	0.0625
ARG	3	2	1	2	3	3	2	2	160	10	0.0625
ARG	3	2	1	2	3	3	3	1	160	10	0.0625
ARG	3	2	1	2	3	3	3	3	160	10	0.0625
ARG	3	2	1	3	1	1	2	2	21	1	0.0476
ARG	3	2	1	3	1	1	3	2	21	1	0.0476
ARG	3	2	1	3	2	2	1	2	21	1	0.0476
ARG	3	2	1	3	3	1	2	3	21	1	0.0476
ARG	3	2	1	3	3	2	1	2	21	9	0.4286
ARG	3	2	1	3	3	2	2	3	21	4	0.1905
ARG	3	2	1	3	3	3	2	1	21	1	0.0476
ARG	3	2	1	3	3	3	2	2	21	1	0.0476
ARG	3	2	1	3	3	3	3	1	21	1	0.0476
ARG	3	2	1	3	3	3	3	3	21	1	0.0476
ARG	3	2	2	1	1	2	1	1	22	2	0.0909
ARG	3	2	2	1	1	2	3	1	22	1	0.0455
ARG	3	2	2	1	2	2	2	2	22	1	0.0455
ARG	3	2	2	1	2	3	3	2	22	1	0.0455
ARG	3	2	2	1	3	1	2	1	22	1	0.0455
ARG	3	2	2	1	3	2	2	2	22	2	0.0909
ARG	3	2	2	1	3	2	2	3	22	1	0.0455
ARG	3	2	2	1	3	2	3	1	22	4	0.1818
ARG	3	2	2	1	3	2	3	3	22	1	0.0455
ARG	3	2	2	1	3	3	2	2	22	1	0.0455
ARG	3	2	2	2	1	2	1	1	422	2	0.0047
ARG	3	2	2	2	2	1	3	3	422	1	0.0024
ARG	3	2	2	2	2	2	1	2	422	1	0.0024
ARG	3	2	2	2	2	2	1	3	422	2	0.0047
ARG	3	2	2	2	2	2	3	3	422	1	0.0024
ARG	3	2	2	2	3	1	2	2	422	2	0.0047
ARG	3	2	2	2	3	2	1	1	422	1	0.0024
ARG	3	2	2	2	3	2	1	2	422	2	0.0047
ARG	3	2	2	2	3	2	2	1	422	2	0.0047
ARG	3	2	2	2	3	2	3	2	422	2	0.0047
ARG	3	2	2	2	3	2	3	3	422	5	0.0118
ARG	3	2	2	2	3	3	2	2	422	4	0.0095
ARG	3	2	2	2	3	3	2	3	422	9	0.0213
ARG	3	2	2	2	3	3	3	3	422	1	0.0024
ARG	3	2	2	3	1	2	1	1	79	13	0.1646
ARG	3	2	2	3	2	1	1	1	79	1	0.0127
ARG	3	2	2	3	2	2	1	1	79	1	0.0127
ARG	3	2	2	3	2	2	2	3	79	1	0.0127
ARG	3	2	2	3	2	2	3	2	79	1	0.0127
ARG	3	2	2	3	2	2	3	3	79	1	0.0127
ARG	3	2	2	3	2	3	3	2	79	1	0.0127
ARG	3	2	2	3	3	2	1	3	79	1	0.0127
ARG	3	2	2	3	3	2	2	1	79	14	0.1772
ARG	3	2	2	3	3	2	2	2	79	7	0.0886
ARG	3	2	2	3	3	2	3	2	79	4	0.0506
ARG	3	2	2	3	3	2	3	3	79	1	0.0127
ARG	3	2	2	3	3	3	2	2	79	26	0.3291
ARG	3	2	2	3	3	3	3	1	79	3	0.0380
ARG	3	2	3	1	1	2	2	1	16	1	0.0625
ARG	3	2	3	1	1	2	2	2	16	2	0.1250
ARG	3	2	3	1	1	2	3	3	16	2	0.1250
ARG	3	2	3	1	2	2	1	2	16	1	0.0625
ARG	3	2	3	1	2	2	2	3	16	2	0.1250
ARG	3	2	3	1	2	3	2	2	16	1	0.0625
ARG	3	2	3	1	3	2	1	1	16	1	0.0625
ARG	3	2	3	1	3	2	1	2	16	1	0.0625
ARG	3	2	3	1	3	2	1	3	16	1	0.0625
ARG	3	2	3	1	3	2	2	2	16	2	0.1250
ARG	3	2	3	1	3	3	2	2	16	1	0.0625
ARG	3	2	3	2	1	2	1	1	68	10	0.1471
ARG	3	2	3	2	1	2	3	2	68	1	0.0147
ARG	3	2	3	2	2	2	1	2	68	1	0.0147
ARG	3	2	3	2	3	1	2	2	68	2	0.0294
ARG	3	2	3	2	3	2	1	1	68	1	0.0147
ARG	3	2	3	2	3	2	1	2	68	2	0.0294
ARG	3	2	3	2	3	2	2	1	68	11	0.1618
ARG	3	2	3	2	3	2	2	2	68	1	0.0147
ARG	3	2	3	2	3	2	3	3	68	4	0.0588
ARG	3	2	3	2	3	3	2	2	68	20	0.2941
ARG	3	2	3	2	3	3	2	3	68	9	0.1324
ARG	3	2	3	2	3	3	3	2	68	1	0.0147
ARG	3	2	3	2	3	3	3	3	68	1	0.0147
ARG	3	2	3	3	1	2	1	3	75	2	0.0267

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
ARG	3	2	3	3	1	2	2	2	75	1	0.0133
ARG	3	2	3	3	2	1	1	1	75	1	0.0133
ARG	3	2	3	3	2	2	1	2	75	7	0.0933
ARG	3	2	3	3	2	2	1	3	75	1	0.0133
ARG	3	2	3	3	2	2	3	1	75	1	0.0133
ARG	3	2	3	3	2	2	3	2	75	5	0.0667
ARG	3	2	3	3	2	2	3	3	75	1	0.0133
ARG	3	2	3	3	3	1	2	2	75	2	0.0267
ARG	3	2	3	3	3	1	2	3	75	1	0.0133
ARG	3	2	3	3	3	2	1	1	75	1	0.0133
ARG	3	2	3	3	3	2	1	2	75	2	0.0267
ARG	3	2	3	3	3	2	1	3	75	1	0.0133
ARG	3	2	3	3	3	2	2	2	75	2	0.0267
ARG	3	2	3	3	3	2	3	1	75	2	0.0267
ARG	3	2	3	3	3	2	3	2	75	1	0.0133
ARG	3	2	3	3	3	3	1	2	75	1	0.0133
ARG	3	2	3	3	3	3	2	2	75	1	0.0133
ARG	3	2	3	3	3	3	2	3	75	9	0.1200
ARG	3	2	3	3	3	3	3	3	75	2	0.0267
ARG	3	3	1	1	2	2	1	2	11	4	0.3636
ARG	3	3	1	1	2	2	1	3	11	1	0.0909
ARG	3	3	1	1	2	2	3	1	11	1	0.0909
ARG	3	3	1	1	3	1	2	3	11	1	0.0909
ARG	3	3	1	1	3	3	1	2	11	1	0.0909
ARG	3	3	1	1	3	3	2	2	11	1	0.0909
ARG	3	3	1	1	3	3	3	3	11	1	0.0909
ARG	3	3	1	2	1	2	3	2	44	1	0.0227
ARG	3	3	1	2	2	1	1	1	44	1	0.0227
ARG	3	3	1	2	2	2	3	1	44	3	0.0682
ARG	3	3	1	2	2	2	3	2	44	6	0.1364
ARG	3	3	1	2	2	2	3	3	44	3	0.0682
ARG	3	3	1	2	2	3	2	2	44	1	0.0227
ARG	3	3	1	2	3	1	2	2	44	2	0.0455
ARG	3	3	1	2	3	2	1	1	44	1	0.0227
ARG	3	3	1	2	3	2	2	2	44	1	0.0227
ARG	3	3	1	2	3	3	2	2	44	9	0.2045
ARG	3	3	1	2	3	3	2	3	44	11	0.2500
ARG	3	3	1	2	3	3	3	2	44	1	0.0227
ARG	3	3	1	2	3	3	3	3	44	3	0.0682
ARG	3	3	2	1	2	2	1	2	21	1	0.0476
ARG	3	3	2	1	2	3	2	2	21	1	0.0476
ARG	3	3	2	1	3	1	2	3	21	1	0.0476
ARG	3	3	2	1	3	2	1	2	21	9	0.4286
ARG	3	3	2	1	3	2	2	2	21	1	0.0476
ARG	3	3	2	1	3	2	2	3	21	1	0.0476
ARG	3	3	2	1	3	3	2	2	21	1	0.0476
ARG	3	3	2	1	3	3	3	1	21	1	0.0476
ARG	3	3	2	1	3	3	3	3	21	1	0.0476
ARG	3	3	2	2	1	2	2	1	91	1	0.0110
ARG	3	3	2	2	1	2	2	2	91	1	0.0110
ARG	3	3	2	2	2	3	2	2	91	1	0.0110
ARG	3	3	2	2	2	3	3	2	91	1	0.0110
ARG	3	3	2	2	3	1	1	2	91	1	0.0110
ARG	3	3	2	2	3	3	1	2	91	6	0.0659
ARG	3	3	2	2	3	3	2	3	91	12	0.1319
ARG	3	3	2	2	3	3	3	3	91	12	0.1319
ARG	3	3	2	3	1	2	1	1	49	1	0.0204
ARG	3	3	2	3	2	1	1	1	49	1	0.0204
ARG	3	3	2	3	2	2	1	1	49	1	0.0204
ARG	3	3	2	3	2	2	2	3	49	1	0.0204
ARG	3	3	2	3	2	2	3	2	49	1	0.0204
ARG	3	3	2	3	2	2	3	3	49	1	0.0204
ARG	3	3	2	3	2	3	3	2	49	1	0.0204
ARG	3	3	2	3	3	2	1	2	49	1	0.0204
ARG	3	3	2	3	3	2	1	3	49	1	0.0204
ARG	3	3	2	3	3	2	2	1	49	2	0.0408
ARG	3	3	2	3	3	2	2	2	49	1	0.0204
ARG	3	3	2	3	3	3	1	2	49	2	0.0408
ARG	3	3	2	3	3	3	2	2	49	20	0.4082
ARG	3	3	2	3	3	3	3	1	49	3	0.0612
ARG	3	3	2	3	3	3	3	2	49	2	0.0408
ARG	3	3	2	3	3	3	3	3	49	4	0.0816
ARG	3	3	3	2	3	1	2	2	44	2	0.0455
ARG	3	3	3	2	3	2	1	1	44	1	0.0227
ARG	3	3	3	2	3	2	2	2	44	15	0.3409
ARG	3	3	3	2	3	2	2	3	44	2	0.0455
ARG	3	3	3	2	3	3	1	2	44	1	0.0227
ARG	3	3	3	2	3	3	2	2	44	9	0.2045

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
ARG	3	3	3	2	3	3	2	3	44	11	0.2500
ARG	3	3	3	2	3	3	3	3	44	3	0.0682
ARG	3	3	3	3	2	1	1	1	83	1	0.0120
ARG	3	3	3	3	2	1	2	3	83	1	0.0120
ARG	3	3	3	3	2	2	1	1	83	1	0.0120
ARG	3	3	3	3	2	2	2	3	83	1	0.0120
ARG	3	3	3	3	2	2	3	2	83	1	0.0120
ARG	3	3	3	3	2	2	3	3	83	1	0.0120
ARG	3	3	3	3	2	3	3	2	83	1	0.0120
ARG	3	3	3	3	3	1	2	2	83	8	0.0964
ARG	3	3	3	3	3	2	1	1	83	4	0.0482
ARG	3	3	3	3	3	2	1	3	83	1	0.0120
ARG	3	3	3	3	3	2	2	1	83	1	0.0120
ARG	3	3	3	3	3	2	2	2	83	5	0.0602
ARG	3	3	3	3	3	3	1	2	83	1	0.0120
ARG	3	3	3	3	3	3	2	2	83	9	0.1084
ARG	3	3	3	3	3	3	2	3	83	38	0.4578
ARG	3	3	3	3	3	3	3	1	83	3	0.0361
ASN	1	1	0	0	1	2	0	0	385	18	0.0468
ASN	1	1	0	0	1	3	0	0	385	30	0.0779
ASN	1	1	0	0	2	1	0	0	385	55	0.1429
ASN	1	1	0	0	2	3	0	0	385	37	0.0961
ASN	1	1	0	0	3	1	0	0	385	1	0.0026
ASN	1	1	0	0	3	2	0	0	385	1	0.0026
ASN	1	2	0	0	1	1	0	0	196	89	0.4541
ASN	1	2	0	0	1	3	0	0	196	13	0.0663
ASN	1	2	0	0	2	1	0	0	196	34	0.1735
ASN	1	2	0	0	2	2	0	0	196	1	0.0051
ASN	1	2	0	0	2	3	0	0	196	22	0.1122
ASN	1	2	0	0	3	1	0	0	196	1	0.0051
ASN	1	2	0	0	3	3	0	0	196	5	0.0255
ASN	1	3	0	0	1	1	0	0	41	3	0.0732
ASN	1	3	0	0	2	1	0	0	41	3	0.0732
ASN	1	3	0	0	3	1	0	0	41	8	0.1951
ASN	1	3	0	0	3	3	0	0	41	16	0.3902
ASN	2	1	0	0	1	3	0	0	689	14	0.0203
ASN	2	1	0	0	2	2	0	0	689	149	0.2163
ASN	2	1	0	0	2	3	0	0	689	75	0.1089
ASN	2	1	0	0	3	1	0	0	689	1	0.0015
ASN	2	1	0	0	3	2	0	0	689	5	0.0073
ASN	2	1	0	0	3	3	0	0	689	27	0.0392
ASN	2	2	0	0	1	1	0	0	1165	1	0.0009
ASN	2	2	0	0	1	3	0	0	1165	2	0.0017
ASN	2	2	0	0	2	1	0	0	1165	255	0.2189
ASN	2	2	0	0	2	3	0	0	1165	51	0.0438
ASN	2	2	0	0	3	1	0	0	1165	4	0.0034
ASN	2	2	0	0	3	2	0	0	1165	3	0.0026
ASN	2	2	0	0	3	3	0	0	1165	25	0.0215
ASN	2	3	0	0	2	1	0	0	1261	4	0.0032
ASN	2	3	0	0	2	2	0	0	1261	163	0.1293
ASN	3	1	0	0	1	1	0	0	22	3	0.1364
ASN	3	1	0	0	1	3	0	0	22	2	0.0909
ASN	3	1	0	0	2	1	0	0	22	1	0.0455
ASN	3	1	0	0	2	3	0	0	22	3	0.1364
ASN	3	1	0	0	3	2	0	0	22	2	0.0909
ASN	3	1	0	0	3	3	0	0	22	10	0.4545
ASN	3	2	0	0	1	1	0	0	829	10	0.0121
ASN	3	2	0	0	1	3	0	0	829	2	0.0024
ASN	3	2	0	0	2	1	0	0	829	5	0.0060
ASN	3	2	0	0	2	2	0	0	829	1	0.0012
ASN	3	2	0	0	2	3	0	0	829	8	0.0097
ASN	3	2	0	0	3	1	0	0	829	53	0.0639
ASN	3	2	0	0	3	3	0	0	829	190	0.2292
ASN	3	3	0	0	1	1	0	0	3309	26	0.0079
ASN	3	3	0	0	1	3	0	0	3309	1	0.0003
ASN	3	3	0	0	2	1	0	0	3309	276	0.0834
ASN	3	3	0	0	2	2	0	0	3309	39	0.0118
ASN	3	3	0	0	2	3	0	0	3309	35	0.0106
ASN	3	3	0	0	3	1	0	0	3309	101	0.0305
ASN	3	3	0	0	3	2	0	0	3309	203	0.0613
ASP	1	1	0	0	1	2	0	0	10	3	0.3000
ASP	1	1	0	0	1	3	0	0	10	2	0.2000
ASP	1	1	0	0	3	1	0	0	10	1	0.1000
ASP	1	1	0	0	3	3	0	0	10	2	0.2000
ASP	1	2	0	0	1	1	0	0	379	66	0.1741
ASP	1	2	0	0	1	3	0	0	379	77	0.2032
ASP	1	2	0	0	2	1	0	0	379	2	0.0053
ASP	1	2	0	0	2	3	0	0	379	2	0.0053

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
ASP	1	2	0	0	3	1	0	0	379	1	0.0026
ASP	1	2	0	0	3	3	0	0	379	2	0.0053
ASP	1	3	0	0	1	1	0	0	51	4	0.0784
ASP	1	3	0	0	1	2	0	0	51	16	0.3137
ASP	1	3	0	0	2	1	0	0	51	7	0.1373
ASP	1	3	0	0	2	2	0	0	51	13	0.2549
ASP	1	3	0	0	2	3	0	0	51	5	0.0980
ASP	1	3	0	0	3	1	0	0	51	2	0.0392
ASP	1	3	0	0	3	2	0	0	51	1	0.0196
ASP	1	3	0	0	3	3	0	0	51	2	0.0392
ASP	2	1	0	0	1	1	0	0	464	16	0.0345
ASP	2	1	0	0	1	2	0	0	464	56	0.1207
ASP	2	1	0	0	2	2	0	0	464	22	0.0474
ASP	2	1	0	0	2	3	0	0	464	19	0.0409
ASP	2	1	0	0	3	1	0	0	464	10	0.0216
ASP	2	2	0	0	2	1	0	0	1401	164	0.1171
ASP	2	2	0	0	2	3	0	0	1401	3	0.0021
ASP	2	2	0	0	3	2	0	0	1401	1	0.0007
ASP	2	3	0	0	1	2	0	0	6	1	0.1667
ASP	2	3	0	0	2	1	0	0	6	1	0.1667
ASP	2	3	0	0	2	2	0	0	6	1	0.1667
ASP	2	3	0	0	3	1	0	0	6	1	0.1667
ASP	2	3	0	0	3	2	0	0	6	1	0.1667
ASP	2	3	0	0	3	3	0	0	6	1	0.1667
ASP	3	1	0	0	2	1	0	0	17	8	0.4706
ASP	3	1	0	0	2	3	0	0	17	6	0.3529
ASP	3	1	0	0	3	2	0	0	17	1	0.0588
ASP	3	2	0	0	1	1	0	0	176	3	0.0170
ASP	3	2	0	0	1	2	0	0	176	9	0.0511
ASP	3	2	0	0	2	1	0	0	176	17	0.0966
ASP	3	2	0	0	2	2	0	0	176	1	0.0057
ASP	3	2	0	0	2	3	0	0	176	13	0.0739
ASP	3	2	0	0	3	1	0	0	176	6	0.0341
ASP	3	2	0	0	3	3	0	0	176	76	0.4318
ASP	3	3	0	0	2	1	0	0	1112	107	0.0962
ASP	3	3	0	0	2	2	0	0	1112	4	0.0036
ASP	3	3	0	0	2	3	0	0	1112	70	0.0629
ASP	3	3	0	0	3	1	0	0	1112	52	0.0468
ASP	3	3	0	0	3	2	0	0	1112	225	0.2023
CYS	2	0	0	0	1	0	0	0	1172	4	0.0034
CYS	2	0	0	0	3	0	0	0	1172	2	0.0017
CYS	3	0	0	0	1	0	0	0	4569	1	0.0002
GLN	1	1	2	0	2	1	1	0	14	1	0.0714
GLN	1	1	2	0	2	1	2	0	14	12	0.8571
GLN	1	1	2	0	3	1	1	0	14	1	0.0714
GLN	1	2	1	0	1	2	3	0	162	2	0.0123
GLN	1	2	1	0	3	1	1	0	162	77	0.4753
GLN	1	2	1	0	3	1	2	0	162	44	0.2716
GLN	1	2	1	0	3	1	3	0	162	11	0.0679
GLN	1	2	1	0	3	2	3	0	162	1	0.0062
GLN	1	2	1	0	3	3	2	0	162	11	0.0679
GLN	1	2	1	0	3	3	3	0	162	11	0.0679
GLN	1	2	2	0	2	1	2	0	56	1	0.0179
GLN	1	2	2	0	2	1	3	0	56	1	0.0179
GLN	1	2	2	0	2	2	1	0	56	8	0.1429
GLN	1	2	2	0	2	2	2	0	56	8	0.1429
GLN	1	2	2	0	2	2	3	0	56	2	0.0357
GLN	1	2	2	0	3	1	1	0	56	14	0.2500
GLN	1	2	2	0	3	1	2	0	56	8	0.1429
GLN	1	2	2	0	3	1	3	0	56	2	0.0357
GLN	1	2	2	0	3	2	1	0	56	2	0.0357
GLN	1	2	2	0	3	2	3	0	56	5	0.0893
GLN	1	2	2	0	3	3	1	0	56	1	0.0179
GLN	1	2	2	0	3	3	2	0	56	2	0.0357
GLN	1	2	2	0	3	3	3	0	56	2	0.0357
GLN	1	2	3	0	1	2	1	0	7	1	0.1429
GLN	1	2	3	0	2	1	1	0	7	1	0.1429
GLN	1	2	3	0	3	1	1	0	7	1	0.1429
GLN	1	2	3	0	3	2	3	0	7	3	0.4286
GLN	1	3	2	0	3	1	1	0	14	7	0.5000
GLN	1	3	2	0	3	1	2	0	14	4	0.2857
GLN	1	3	2	0	3	1	3	0	14	1	0.0714
GLN	1	3	2	0	3	3	2	0	14	1	0.0714
GLN	1	3	2	0	3	3	3	0	14	1	0.0714
GLN	1	3	3	0	2	1	1	0	29	1	0.0345
GLN	1	3	3	0	3	1	1	0	29	14	0.4828
GLN	1	3	3	0	3	1	2	0	29	8	0.2759
GLN	1	3	3	0	3	1	3	0	29	2	0.0690

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
GLN	1	3	3	0	3	3	2	0	29	2	0.0690
GLN	1	3	3	0	3	3	3	0	29	2	0.0690
GLN	2	1	1	0	1	3	2	0	426	1	0.0023
GLN	2	1	1	0	2	1	2	0	426	29	0.0681
GLN	2	1	1	0	2	2	2	0	426	2	0.0047
GLN	2	1	1	0	2	3	3	0	426	1	0.0023
GLN	2	1	1	0	3	2	2	0	426	2	0.0047
GLN	2	1	1	0	3	3	2	0	426	1	0.0023
GLN	2	1	1	0	3	3	3	0	426	1	0.0023
GLN	2	1	2	0	2	2	2	0	22	10	0.4545
GLN	2	1	2	0	3	2	2	0	22	5	0.2273
GLN	2	1	2	0	3	3	3	0	22	6	0.2727
GLN	2	1	3	0	2	1	1	0	7	2	0.2857
GLN	2	1	3	0	2	2	1	0	7	1	0.1429
GLN	2	1	3	0	2	2	3	0	7	3	0.4286
GLN	2	1	3	0	3	3	3	0	7	1	0.1429
GLN	2	2	1	0	2	1	1	0	578	10	0.0173
GLN	2	2	1	0	2	1	2	0	578	121	0.2093
GLN	2	2	1	0	2	1	3	0	578	1	0.0017
GLN	2	2	1	0	2	2	2	0	578	11	0.0190
GLN	2	2	1	0	2	2	3	0	578	35	0.0606
GLN	2	2	1	0	2	3	2	0	578	3	0.0052
GLN	2	2	1	0	3	1	1	0	578	10	0.0173
GLN	2	2	1	0	3	2	1	0	578	3	0.0052
GLN	2	2	1	0	3	2	2	0	578	2	0.0035
GLN	2	2	1	0	3	2	3	0	578	7	0.0121
GLN	2	2	1	0	3	3	1	0	578	1	0.0017
GLN	2	2	1	0	3	3	3	0	578	1	0.0017
GLN	2	2	2	0	2	1	2	0	99	3	0.0303
GLN	2	2	2	0	2	1	3	0	99	2	0.0202
GLN	2	2	2	0	2	2	1	0	99	17	0.1717
GLN	2	2	2	0	2	2	3	0	99	5	0.0505
GLN	2	2	2	0	3	1	1	0	99	15	0.1515
GLN	2	2	2	0	3	1	2	0	99	8	0.0808
GLN	2	2	2	0	3	1	3	0	99	2	0.0202
GLN	2	2	2	0	3	2	1	0	99	5	0.0505
GLN	2	2	2	0	3	2	2	0	99	2	0.0202
GLN	2	2	2	0	3	2	3	0	99	11	0.1111
GLN	2	2	2	0	3	3	1	0	99	2	0.0202
GLN	2	2	2	0	3	3	2	0	99	2	0.0202
GLN	2	2	2	0	3	3	3	0	99	3	0.0303
GLN	2	2	3	0	1	2	2	0	45	2	0.0444
GLN	2	2	3	0	2	1	1	0	45	1	0.0222
GLN	2	2	3	0	2	1	2	0	45	12	0.2667
GLN	2	2	3	0	2	2	1	0	45	20	0.4444
GLN	2	2	3	0	2	2	2	0	45	1	0.0222
GLN	2	2	3	0	3	1	1	0	45	1	0.0222
GLN	2	2	3	0	3	2	3	0	45	2	0.0444
GLN	2	3	1	0	3	1	1	0	15	7	0.4667
GLN	2	3	1	0	3	1	2	0	15	5	0.3333
GLN	2	3	1	0	3	1	3	0	15	1	0.0667
GLN	2	3	1	0	3	3	2	0	15	1	0.0667
GLN	2	3	1	0	3	3	3	0	15	1	0.0667
GLN	2	3	2	0	2	1	2	0	28	2	0.0714
GLN	2	3	2	0	2	1	3	0	28	2	0.0714
GLN	2	3	2	0	2	2	1	0	28	16	0.5714
GLN	2	3	2	0	2	2	2	0	28	2	0.0714
GLN	2	3	2	0	2	2	3	0	28	4	0.1429
GLN	2	3	2	0	3	3	1	0	28	2	0.0714
GLN	2	3	3	0	2	1	2	0	57	4	0.0702
GLN	2	3	3	0	2	1	3	0	57	4	0.0702
GLN	2	3	3	0	2	2	1	0	57	32	0.5614
GLN	2	3	3	0	2	2	2	0	57	5	0.0877
GLN	2	3	3	0	2	2	3	0	57	8	0.1404
GLN	2	3	3	0	3	3	1	0	57	4	0.0702
GLN	3	1	1	0	2	1	1	0	74	1	0.0135
GLN	3	1	1	0	2	1	2	0	74	14	0.1892
GLN	3	1	1	0	2	1	3	0	74	1	0.0135
GLN	3	1	1	0	2	2	1	0	74	8	0.1081
GLN	3	1	1	0	2	2	2	0	74	1	0.0135
GLN	3	1	1	0	2	2	3	0	74	3	0.0405
GLN	3	1	1	0	3	1	2	0	74	12	0.1622
GLN	3	1	1	0	3	1	3	0	74	3	0.0405
GLN	3	1	1	0	3	2	2	0	74	1	0.0135
GLN	3	1	1	0	3	3	1	0	74	2	0.0270
GLN	3	1	1	0	3	3	2	0	74	3	0.0405
GLN	3	1	1	0	3	3	3	0	74	3	0.0405
GLN	3	1	2	0	2	1	2	0	39	1	0.0256

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
GLN	3	1	2	0	2	1	3	0	39	1	0.0256
GLN	3	1	2	0	2	2	1	0	39	8	0.2051
GLN	3	1	2	0	2	2	2	0	39	3	0.0769
GLN	3	1	2	0	2	2	3	0	39	2	0.0513
GLN	3	1	2	0	3	1	1	0	39	7	0.1795
GLN	3	1	2	0	3	1	3	0	39	1	0.0256
GLN	3	1	2	0	3	2	2	0	39	1	0.0256
GLN	3	1	2	0	3	3	1	0	39	1	0.0256
GLN	3	1	2	0	3	3	2	0	39	1	0.0256
GLN	3	1	2	0	3	3	3	0	39	2	0.0513
GLN	3	1	3	0	1	2	3	0	4	1	0.2500
GLN	3	1	3	0	1	3	1	0	4	1	0.2500
GLN	3	1	3	0	3	1	1	0	4	1	0.2500
GLN	3	1	3	0	3	1	2	0	4	1	0.2500
GLN	3	2	1	0	1	2	1	0	802	9	0.0112
GLN	3	2	1	0	1	2	3	0	802	1	0.0012
GLN	3	2	1	0	1	3	1	0	802	1	0.0012
GLN	3	2	1	0	2	1	1	0	802	5	0.0062
GLN	3	2	1	0	2	1	2	0	802	60	0.0748
GLN	3	2	1	0	2	1	3	0	802	9	0.0112
GLN	3	2	1	0	2	2	1	0	802	1	0.0012
GLN	3	2	1	0	2	2	2	0	802	53	0.0661
GLN	3	2	1	0	2	2	3	0	802	2	0.0025
GLN	3	2	1	0	3	1	1	0	802	5	0.0062
GLN	3	2	1	0	3	1	3	0	802	9	0.0112
GLN	3	2	1	0	3	2	2	0	802	13	0.0162
GLN	3	2	1	0	3	2	3	0	802	55	0.0686
GLN	3	2	1	0	3	3	2	0	802	18	0.0224
GLN	3	2	1	0	3	3	3	0	802	37	0.0461
GLN	3	2	2	0	1	2	1	0	146	1	0.0068
GLN	3	2	2	0	1	2	2	0	146	2	0.0137
GLN	3	2	2	0	2	1	1	0	146	6	0.0411
GLN	3	2	2	0	2	1	2	0	146	48	0.3288
GLN	3	2	2	0	2	1	3	0	146	5	0.0342
GLN	3	2	2	0	2	2	2	0	146	1	0.0068
GLN	3	2	2	0	2	2	3	0	146	4	0.0274
GLN	3	2	2	0	3	1	1	0	146	18	0.1233
GLN	3	2	2	0	3	1	2	0	146	8	0.0548
GLN	3	2	2	0	3	1	3	0	146	3	0.0205
GLN	3	2	2	0	3	2	1	0	146	7	0.0479
GLN	3	2	2	0	3	2	3	0	146	6	0.0411
GLN	3	2	2	0	3	3	2	0	146	8	0.0548
GLN	3	2	2	0	3	3	3	0	146	20	0.1370
GLN	3	2	3	0	1	2	1	0	748	1	0.0013
GLN	3	2	3	0	1	2	2	0	748	1	0.0013
GLN	3	2	3	0	2	1	1	0	748	7	0.0094
GLN	3	2	3	0	2	1	2	0	748	84	0.1123
GLN	3	2	3	0	2	1	3	0	748	4	0.0053
GLN	3	2	3	0	2	2	2	0	748	126	0.1684
GLN	3	2	3	0	2	2	3	0	748	1	0.0013
GLN	3	2	3	0	3	1	1	0	748	14	0.0187
GLN	3	2	3	0	3	1	2	0	748	4	0.0053
GLN	3	2	3	0	3	1	3	0	748	3	0.0040
GLN	3	2	3	0	3	2	1	0	748	382	0.5107
GLN	3	2	3	0	3	2	2	0	748	6	0.0080
GLN	3	2	3	0	3	3	2	0	748	5	0.0067
GLN	3	2	3	0	3	3	3	0	748	13	0.0174
GLN	3	3	1	0	2	1	3	0	63	4	0.0635
GLN	3	3	1	0	3	2	1	0	63	8	0.1270
GLN	3	3	1	0	3	3	2	0	63	8	0.1270
GLN	3	3	1	0	3	3	3	0	63	39	0.6190
GLN	3	3	2	0	2	2	2	0	449	10	0.0223
GLN	3	3	2	0	3	1	1	0	449	7	0.0156
GLN	3	3	2	0	3	1	2	0	449	4	0.0089
GLN	3	3	2	0	3	1	3	0	449	1	0.0022
GLN	3	3	2	0	3	2	2	0	449	5	0.0111
GLN	3	3	2	0	3	3	3	0	449	7	0.0156
GLN	3	3	3	0	1	2	1	0	566	3	0.0053
GLN	3	3	3	0	1	3	3	0	566	1	0.0018
GLN	3	3	3	0	2	1	1	0	566	1	0.0018
GLN	3	3	3	0	2	1	3	0	566	23	0.0406
GLN	3	3	3	0	2	2	2	0	566	32	0.0565
GLN	3	3	3	0	3	1	1	0	566	14	0.0247
GLN	3	3	3	0	3	1	2	0	566	8	0.0141
GLN	3	3	3	0	3	1	3	0	566	4	0.0071
GLN	3	3	3	0	3	2	1	0	566	49	0.0866
GLN	3	3	3	0	3	2	2	0	566	19	0.0336
GLN	3	3	3	0	3	2	3	0	566	6	0.0106

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
GLN	3	3	3	0	3	3	1	0	566	47	0.0830
GLN	3	3	3	0	3	3	2	0	566	48	0.0848
GLU	1	2	1	0	1	2	2	0	25	2	0.0800
GLU	1	2	1	0	1	3	2	0	25	1	0.0400
GLU	1	2	1	0	2	1	1	0	25	2	0.0800
GLU	1	2	1	0	2	2	2	0	25	1	0.0400
GLU	1	2	1	0	3	1	1	0	25	4	0.1600
GLU	1	2	1	0	3	2	1	0	25	4	0.1600
GLU	1	2	2	0	2	2	1	0	2	2	1.0000
GLU	1	2	3	0	1	2	2	0	9	2	0.2222
GLU	1	2	3	0	2	2	1	0	9	2	0.2222
GLU	1	2	3	0	3	2	2	0	9	1	0.1111
GLU	1	3	1	0	1	2	1	0	32	13	0.4063
GLU	1	3	1	0	1	3	2	0	32	1	0.0313
GLU	1	3	1	0	2	2	2	0	32	1	0.0313
GLU	1	3	1	0	3	1	1	0	32	11	0.3438
GLU	1	3	1	0	3	2	1	0	32	6	0.1875
GLU	1	3	2	0	1	1	1	0	28	1	0.0357
GLU	1	3	2	0	1	1	3	0	28	1	0.0357
GLU	1	3	2	0	1	2	1	0	28	11	0.3929
GLU	1	3	2	0	2	1	1	0	28	1	0.0357
GLU	1	3	2	0	2	2	2	0	28	1	0.0357
GLU	1	3	2	0	3	1	1	0	28	5	0.1786
GLU	1	3	2	0	3	1	2	0	28	1	0.0357
GLU	1	3	2	0	3	2	1	0	28	5	0.1786
GLU	1	3	3	0	1	1	3	0	66	1	0.0152
GLU	1	3	3	0	1	2	1	0	66	24	0.3636
GLU	1	3	3	0	1	3	2	0	66	1	0.0152
GLU	1	3	3	0	2	1	1	0	66	1	0.0152
GLU	1	3	3	0	3	1	1	0	66	12	0.1818
GLU	1	3	3	0	3	1	2	0	66	14	0.2121
GLU	1	3	3	0	3	1	3	0	66	1	0.0152
GLU	1	3	3	0	3	2	1	0	66	12	0.1818
GLU	2	1	1	0	1	2	2	0	7	1	0.1429
GLU	2	1	1	0	2	1	2	0	7	2	0.2857
GLU	2	1	1	0	2	2	1	0	7	1	0.1429
GLU	2	1	1	0	2	2	2	0	7	1	0.1429
GLU	2	1	1	0	3	2	3	0	7	1	0.1429
GLU	2	1	1	0	3	3	3	0	7	1	0.1429
GLU	2	1	2	0	2	2	1	0	1	1	1.0000
GLU	2	1	3	0	2	2	1	0	14	1	0.0714
GLU	2	1	3	0	2	2	2	0	14	1	0.0714
GLU	2	1	3	0	2	2	3	0	14	1	0.0714
GLU	2	2	2	0	2	2	1	0	155	52	0.3355
GLU	2	2	2	0	2	2	3	0	155	3	0.0194
GLU	2	2	2	0	3	2	3	0	155	1	0.0065
GLU	2	2	3	0	2	2	1	0	25	5	0.2000
GLU	2	2	3	0	2	2	2	0	25	15	0.6000
GLU	2	3	1	0	2	1	2	0	3	1	0.3333
GLU	2	3	1	0	2	2	1	0	3	1	0.3333
GLU	2	3	1	0	3	3	3	0	3	1	0.3333
GLU	2	3	3	0	1	2	3	0	1	1	1.0000
GLU	3	1	1	0	1	2	1	0	6	2	0.3333
GLU	3	1	1	0	1	3	2	0	6	1	0.1667
GLU	3	1	1	0	2	2	3	0	6	1	0.1667
GLU	3	1	1	0	3	2	1	0	6	1	0.1667
GLU	3	1	2	0	1	2	3	0	360	1	0.0028
GLU	3	1	2	0	2	2	1	0	360	5	0.0139
GLU	3	1	2	0	2	2	2	0	360	9	0.0250
GLU	3	1	2	0	3	1	3	0	360	26	0.0722
GLU	3	1	2	0	3	3	2	0	360	1	0.0028
GLU	3	1	2	0	3	3	3	0	360	1	0.0028
GLU	3	1	3	0	3	1	2	0	61	56	0.9180
GLU	3	2	1	0	2	2	2	0	40	1	0.0250
GLU	3	2	1	0	2	2	3	0	40	1	0.0250
GLU	3	2	1	0	3	2	2	0	40	4	0.1000
GLU	3	2	1	0	3	2	3	0	40	25	0.6250
GLU	3	2	2	0	1	3	2	0	12	1	0.0833
GLU	3	2	2	0	2	2	2	0	12	1	0.0833
GLU	3	2	2	0	2	2	3	0	12	1	0.0833
GLU	3	2	2	0	3	1	3	0	12	1	0.0833
GLU	3	2	2	0	3	2	3	0	12	1	0.0833
GLU	3	2	2	0	3	3	3	0	12	1	0.0833
GLU	3	2	3	0	2	1	3	0	385	1	0.0026
GLU	3	2	3	0	2	2	3	0	385	1	0.0026
GLU	3	2	3	0	2	3	3	0	385	1	0.0026
GLU	3	2	3	0	3	2	1	0	385	81	0.2104
GLU	3	2	3	0	3	2	2	0	385	3	0.0078

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
GLU	3	3	1	0	3	3	2	0	15	14	0.9333
GLU	3	3	1	0	3	3	3	0	15	1	0.0667
GLU	3	3	2	0	1	3	2	0	571	1	0.0018
GLU	3	3	2	0	1	3	3	0	571	1	0.0018
GLU	3	3	2	0	2	2	3	0	571	1	0.0018
GLU	3	3	2	0	3	2	1	0	571	1	0.0018
GLU	3	3	2	0	3	2	2	0	571	1	0.0018
GLU	3	3	2	0	3	2	3	0	571	2	0.0035
GLU	3	3	2	0	3	3	3	0	571	3	0.0053
GLU	3	3	3	0	1	2	1	0	20	2	0.1000
GLU	3	3	3	0	2	1	1	0	20	1	0.0500
GLU	3	3	3	0	2	1	3	0	20	4	0.2000
GLU	3	3	3	0	2	2	2	0	20	3	0.1500
GLU	3	3	3	0	2	3	1	0	20	1	0.0500
GLU	3	3	3	0	3	1	1	0	20	3	0.1500
GLU	3	3	3	0	3	2	1	0	20	1	0.0500
GLU	3	3	3	0	3	2	2	0	20	2	0.1000
GLU	3	3	3	0	3	3	2	0	20	2	0.1000
HIS	2	1	0	0	1	1	0	0	859	1	0.0012
HIS	2	1	0	0	2	2	0	0	859	1	0.0012
HIS	2	1	0	0	3	1	0	0	859	1	0.0012
HIS	2	1	0	0	3	2	0	0	859	1	0.0012
HIS	2	2	0	0	1	1	0	0	9	1	0.1111
HIS	2	2	0	0	2	1	0	0	9	4	0.4444
HIS	3	1	0	0	1	1	0	0	205	1	0.0049
HIS	3	1	0	0	2	1	0	0	205	2	0.0098
HIS	3	1	0	0	3	2	0	0	205	1	0.0049
HIS	3	2	0	0	2	1	0	0	8	2	0.2500
HIS	3	2	0	0	3	1	0	0	8	2	0.2500
ILE	1	1	0	0	1	2	0	0	22	10	0.4545
ILE	1	1	0	0	2	1	0	0	22	1	0.0455
ILE	1	2	0	0	1	1	0	0	1497	45	0.0301
ILE	1	2	0	0	2	1	0	0	1497	71	0.0474
ILE	1	2	0	0	2	2	0	0	1497	27	0.0180
ILE	1	2	0	0	3	1	0	0	1497	153	0.1022
ILE	1	2	0	0	3	2	0	0	1497	56	0.0374
ILE	1	2	0	0	3	3	0	0	1497	17	0.0114
ILE	2	1	0	0	1	2	0	0	511	4	0.0078
ILE	2	1	0	0	2	2	0	0	511	36	0.0705
ILE	2	2	0	0	1	2	0	0	568	6	0.0106
ILE	2	2	0	0	2	1	0	0	568	77	0.1356
ILE	2	2	0	0	2	3	0	0	568	1	0.0018
ILE	2	2	0	0	3	2	0	0	568	1	0.0018
ILE	2	3	0	0	3	2	0	0	2	1	0.5000
ILE	2	3	0	0	3	3	0	0	2	1	0.5000
ILE	3	1	0	0	1	1	0	0	51	1	0.0196
ILE	3	1	0	0	1	2	0	0	51	1	0.0196
ILE	3	1	0	0	2	1	0	0	51	1	0.0196
ILE	3	1	0	0	2	2	0	0	51	2	0.0392
ILE	3	1	0	0	3	2	0	0	51	13	0.2549
ILE	3	1	0	0	3	3	0	0	51	4	0.0784
ILE	3	2	0	0	1	1	0	0	3813	1	0.0003
ILE	3	2	0	0	1	2	0	0	3813	1	0.0003
ILE	3	2	0	0	2	2	0	0	3813	1	0.0003
ILE	3	2	0	0	3	1	0	0	3813	98	0.0257
ILE	3	2	0	0	3	3	0	0	3813	61	0.0160
ILE	3	3	0	0	1	2	0	0	446	2	0.0045
ILE	3	3	0	0	2	3	0	0	446	1	0.0022
ILE	3	3	0	0	3	1	0	0	446	3	0.0067
ILE	3	3	0	0	3	2	0	0	446	11	0.0247
LEU	1	1	0	0	3	2	0	0	1	1	1.0000
LEU	1	2	0	0	1	1	0	0	1	1	1.0000
LEU	2	1	0	0	2	2	0	0	2448	43	0.0176
LEU	2	1	0	0	2	3	0	0	2448	56	0.0229
LEU	2	1	0	0	3	1	0	0	2448	10	0.0041
LEU	2	1	0	0	3	2	0	0	2448	27	0.0110
LEU	2	1	0	0	3	3	0	0	2448	4	0.0016
LEU	2	2	0	0	2	1	0	0	60	55	0.9167
LEU	2	2	0	0	2	3	0	0	60	2	0.0333
LEU	2	2	0	0	3	2	0	0	60	3	0.0500
LEU	2	3	0	0	2	2	0	0	1	1	1.0000
LEU	3	1	0	0	2	1	0	0	246	2	0.0081
LEU	3	1	0	0	2	2	0	0	246	1	0.0041
LEU	3	1	0	0	3	2	0	0	246	140	0.5691
LEU	3	1	0	0	3	3	0	0	246	3	0.0122
LEU	3	2	0	0	2	1	0	0	4443	44	0.0099
LEU	3	2	0	0	2	2	0	0	4443	1	0.0002
LEU	3	2	0	0	3	1	0	0	4443	492	0.1107

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
LEU	3	2	0	0	3	3	0	0	4443	278	0.0626
LEU	3	3	0	0	3	1	0	0	49	24	0.4898
LEU	3	3	0	0	3	2	0	0	49	22	0.4490
LYS	1	2	1	1	2	2	2	3	1	1	1.0000
LYS	1	2	1	2	1	2	3	2	51	1	0.0196
LYS	1	2	1	2	2	2	2	1	51	6	0.1176
LYS	1	2	1	2	2	2	2	2	51	4	0.0784
LYS	1	2	1	2	2	2	2	3	51	12	0.2353
LYS	1	2	1	2	2	2	3	2	51	4	0.0784
LYS	1	2	1	2	3	2	1	2	51	7	0.1373
LYS	1	2	1	2	3	2	2	1	51	3	0.0588
LYS	1	2	1	2	3	2	2	2	51	12	0.2353
LYS	1	2	1	2	3	3	3	1	51	1	0.0196
LYS	1	2	1	2	3	3	3	2	51	1	0.0196
LYS	1	2	2	2	2	2	2	1	6	1	0.1667
LYS	1	2	2	2	2	2	2	2	6	1	0.1667
LYS	1	2	2	2	2	2	2	3	6	1	0.1667
LYS	1	2	2	3	1	2	2	2	2	2	1.0000
LYS	1	2	3	1	1	2	3	2	14	1	0.0714
LYS	1	2	3	1	2	2	3	2	14	1	0.0714
LYS	1	2	3	1	3	2	2	2	14	12	0.8571
LYS	1	2	3	2	2	2	2	1	274	4	0.0146
LYS	1	2	3	2	2	2	2	2	274	2	0.0073
LYS	1	2	3	2	2	2	2	3	274	8	0.0292
LYS	1	2	3	2	2	2	3	2	274	20	0.0730
LYS	1	2	3	2	3	2	1	2	274	4	0.0146
LYS	1	2	3	2	3	2	2	1	274	2	0.0073
LYS	1	2	3	2	3	2	2	2	274	216	0.7883
LYS	1	2	3	3	2	1	1	2	10	1	0.1000
LYS	1	2	3	3	2	2	2	2	10	3	0.3000
LYS	1	2	3	3	2	2	3	2	10	1	0.1000
LYS	1	2	3	3	3	2	1	2	10	1	0.1000
LYS	1	2	3	3	3	2	2	2	10	3	0.3000
LYS	1	2	3	3	3	2	3	2	10	1	0.1000
LYS	2	1	1	2	2	1	2	2	14	1	0.0714
LYS	2	1	1	2	2	2	2	2	14	4	0.2857
LYS	2	1	1	2	2	2	2	3	14	1	0.0714
LYS	2	1	1	2	2	2	3	1	14	1	0.0714
LYS	2	1	1	2	2	2	3	2	14	1	0.0714
LYS	2	1	1	2	3	2	1	2	14	1	0.0714
LYS	2	1	1	2	3	2	2	2	14	4	0.2857
LYS	2	1	2	1	2	1	2	2	3	1	0.3333
LYS	2	1	2	1	2	2	2	2	3	1	0.3333
LYS	2	1	2	1	3	2	1	2	3	1	0.3333
LYS	2	1	2	2	2	1	1	2	173	1	0.0058
LYS	2	1	2	2	2	2	2	2	173	1	0.0058
LYS	2	1	2	2	2	2	3	2	173	1	0.0058
LYS	2	1	2	2	2	3	2	3	173	1	0.0058
LYS	2	1	2	2	3	2	1	2	173	1	0.0058
LYS	2	1	2	3	1	2	2	2	4	2	0.5000
LYS	2	1	2	3	2	2	3	1	4	2	0.5000
LYS	2	2	1	1	3	1	2	1	3	1	0.3333
LYS	2	2	1	1	3	2	2	2	3	1	0.3333
LYS	2	2	1	1	3	2	3	1	3	1	0.3333
LYS	2	2	1	2	1	2	3	2	16	1	0.0625
LYS	2	2	1	2	2	2	3	2	16	1	0.0625
LYS	2	2	1	2	3	1	2	1	16	1	0.0625
LYS	2	2	1	2	3	2	2	2	16	12	0.7500
LYS	2	2	1	2	3	2	3	1	16	1	0.0625
LYS	2	2	1	3	1	2	3	2	31	1	0.0323
LYS	2	2	1	3	2	1	1	2	31	4	0.1290
LYS	2	2	1	3	2	2	2	2	31	5	0.1613
LYS	2	2	1	3	2	2	3	2	31	5	0.1613
LYS	2	2	1	3	3	2	1	2	31	4	0.1290
LYS	2	2	1	3	3	2	2	2	31	12	0.3871
LYS	2	2	2	1	2	1	1	2	99	1	0.0101
LYS	2	2	2	1	2	1	2	2	99	1	0.0101
LYS	2	2	2	1	2	2	1	2	99	1	0.0101
LYS	2	2	2	1	2	2	1	3	99	4	0.0404
LYS	2	2	2	1	2	2	2	2	99	62	0.6263
LYS	2	2	2	1	2	2	2	3	99	13	0.1313
LYS	2	2	2	1	2	2	3	1	99	1	0.0101
LYS	2	2	2	1	2	2	3	2	99	2	0.0202
LYS	2	2	2	1	3	2	1	2	99	3	0.0303
LYS	2	2	2	1	3	2	2	2	99	5	0.0505
LYS	2	2	2	2	1	2	3	2	1139	5	0.0044
LYS	2	2	2	2	2	1	1	2	1139	24	0.0211
LYS	2	2	2	2	2	1	2	3	1139	5	0.0044

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
LYS	2	2	2	2	2	2	1	2	1139	4	0.0035
LYS	2	2	2	2	2	2	1	3	1139	6	0.0053
LYS	2	2	2	2	2	2	2	1	1139	12	0.0105
LYS	2	2	2	2	2	2	2	3	1139	59	0.0518
LYS	2	2	2	2	2	2	3	2	1139	29	0.0255
LYS	2	2	2	2	2	2	3	3	1139	11	0.0097
LYS	2	2	2	2	2	3	2	2	1139	2	0.0018
LYS	2	2	2	2	3	1	2	2	1139	2	0.0018
LYS	2	2	2	2	3	2	1	2	1139	23	0.0202
LYS	2	2	2	2	3	2	2	2	1139	72	0.0632
LYS	2	2	2	2	3	2	3	2	1139	29	0.0255
LYS	2	2	2	2	3	3	2	1	1139	5	0.0044
LYS	2	2	2	2	3	3	2	3	1139	2	0.0018
LYS	2	2	2	3	2	1	1	2	49	1	0.0204
LYS	2	2	2	3	2	2	1	2	49	1	0.0204
LYS	2	2	2	3	2	2	1	3	49	2	0.0408
LYS	2	2	2	3	2	2	2	1	49	2	0.0408
LYS	2	2	2	3	2	2	2	2	49	34	0.6939
LYS	2	2	2	3	2	2	3	3	49	1	0.0204
LYS	2	2	2	3	3	2	2	2	49	2	0.0408
LYS	2	2	3	1	2	1	1	2	15	1	0.0667
LYS	2	2	3	1	2	1	2	2	15	1	0.0667
LYS	2	2	3	1	2	2	2	2	15	4	0.2667
LYS	2	2	3	1	2	2	2	3	15	1	0.0667
LYS	2	2	3	1	2	2	3	2	15	1	0.0667
LYS	2	2	3	1	3	2	1	2	15	1	0.0667
LYS	2	2	3	1	3	2	2	2	15	5	0.3333
LYS	2	2	3	2	1	2	3	2	138	2	0.0145
LYS	2	2	3	2	2	1	1	2	138	2	0.0145
LYS	2	2	3	2	2	1	2	2	138	3	0.0217
LYS	2	2	3	2	2	2	2	1	138	12	0.0870
LYS	2	2	3	2	2	2	2	2	138	22	0.1594
LYS	2	2	3	2	2	2	2	3	138	26	0.1884
LYS	2	2	3	2	2	2	3	1	138	2	0.0145
LYS	2	2	3	2	3	2	1	2	138	14	0.1014
LYS	2	2	3	2	3	2	2	1	138	6	0.0435
LYS	2	2	3	2	3	2	2	2	138	32	0.2319
LYS	2	2	3	2	3	2	3	2	138	1	0.0072
LYS	2	2	3	2	3	3	3	2	138	1	0.0072
LYS	2	2	3	3	2	1	2	3	13	1	0.0769
LYS	2	2	3	3	2	2	2	3	13	9	0.6923
LYS	2	2	3	3	3	2	2	2	13	2	0.1538
LYS	2	2	3	3	3	3	2	1	13	1	0.0769
LYS	2	3	2	2	3	2	2	1	1	1	1.0000
LYS	2	3	3	1	2	2	2	2	2	2	1.0000
LYS	3	1	1	2	2	2	2	2	2	1	0.5000
LYS	3	1	1	2	3	2	1	2	2	1	0.5000
LYS	3	1	2	2	3	2	2	1	36	4	0.1111
LYS	3	1	2	2	3	2	2	2	36	8	0.2222
LYS	3	1	2	2	3	2	3	2	36	1	0.0278
LYS	3	1	2	2	3	3	2	2	36	17	0.4722
LYS	3	1	2	2	3	3	2	3	36	5	0.1389
LYS	3	1	2	3	3	1	1	1	1	1	1.0000
LYS	3	2	1	1	3	2	2	1	4	1	0.2500
LYS	3	2	1	1	3	3	2	2	4	3	0.7500
LYS	3	2	1	2	2	1	1	2	259	1	0.0039
LYS	3	2	1	2	2	1	2	2	259	1	0.0039
LYS	3	2	1	2	2	1	2	3	259	16	0.0618
LYS	3	2	1	2	2	2	1	1	259	1	0.0039
LYS	3	2	1	2	2	2	2	2	259	4	0.0154
LYS	3	2	1	2	2	2	2	3	259	145	0.5598
LYS	3	2	1	2	2	2	3	1	259	1	0.0039
LYS	3	2	1	2	2	2	3	2	259	1	0.0039
LYS	3	2	1	2	3	2	1	1	259	1	0.0039
LYS	3	2	1	2	3	2	2	1	259	1	0.0039
LYS	3	2	1	2	3	2	2	2	259	39	0.1506
LYS	3	2	1	2	3	2	2	3	259	15	0.0579
LYS	3	2	1	2	3	2	3	2	259	5	0.0193
LYS	3	2	1	2	3	3	2	1	259	16	0.0618
LYS	3	2	1	2	3	3	2	3	259	1	0.0039
LYS	3	2	1	3	3	2	2	1	16	4	0.2500
LYS	3	2	1	3	3	2	2	2	16	12	0.7500
LYS	3	2	2	1	2	1	1	2	126	2	0.0159
LYS	3	2	2	1	2	1	2	2	126	2	0.0159
LYS	3	2	2	1	2	2	2	2	126	8	0.0635
LYS	3	2	2	1	2	2	2	3	126	2	0.0159
LYS	3	2	2	1	2	2	3	1	126	2	0.0159
LYS	3	2	2	1	2	2	3	2	126	2	0.0159

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
LYS	3	2	2	1	3	1	2	3	126	1	0.0079
LYS	3	2	2	1	3	2	1	2	126	5	0.0397
LYS	3	2	2	1	3	2	2	2	126	71	0.5635
LYS	3	2	2	1	3	2	2	3	126	3	0.0238
LYS	3	2	2	1	3	2	3	2	126	4	0.0317
LYS	3	2	2	1	3	3	2	2	126	3	0.0238
LYS	3	2	2	1	3	3	2	3	126	3	0.0238
LYS	3	2	2	2	1	1	2	3	811	2	0.0025
LYS	3	2	2	2	2	1	1	2	811	1	0.0012
LYS	3	2	2	2	2	1	2	2	811	1	0.0012
LYS	3	2	2	2	2	1	2	3	811	1	0.0012
LYS	3	2	2	2	2	2	1	2	811	1	0.0012
LYS	3	2	2	2	2	2	2	1	811	34	0.0419
LYS	3	2	2	2	2	2	2	2	811	21	0.0259
LYS	3	2	2	2	2	2	2	3	811	78	0.0962
LYS	3	2	2	2	2	2	3	1	811	1	0.0012
LYS	3	2	2	2	2	2	3	2	811	18	0.0222
LYS	3	2	2	2	2	2	3	3	811	2	0.0025
LYS	3	2	2	2	3	2	1	2	811	102	0.1258
LYS	3	2	2	2	3	2	2	1	811	76	0.0937
LYS	3	2	2	2	3	2	2	3	811	91	0.1122
LYS	3	2	2	2	3	2	3	2	811	35	0.0432
LYS	3	2	2	2	3	3	2	1	811	1	0.0012
LYS	3	2	2	2	3	3	2	2	811	3	0.0037
LYS	3	2	2	2	3	3	3	2	811	1	0.0012
LYS	3	2	2	3	2	1	1	2	87	3	0.0345
LYS	3	2	2	3	2	1	2	2	87	3	0.0345
LYS	3	2	2	3	2	1	2	3	87	1	0.0115
LYS	3	2	2	3	2	2	2	1	87	2	0.0230
LYS	3	2	2	3	2	2	2	2	87	13	0.1494
LYS	3	2	2	3	2	2	2	3	87	16	0.1839
LYS	3	2	2	3	2	2	3	1	87	3	0.0345
LYS	3	2	2	3	2	2	3	2	87	4	0.0460
LYS	3	2	2	3	3	2	1	2	87	6	0.0690
LYS	3	2	2	3	3	2	2	1	87	5	0.0575
LYS	3	2	2	3	3	2	2	2	87	25	0.2874
LYS	3	2	2	3	3	2	3	2	87	1	0.0115
LYS	3	2	2	3	3	3	2	1	87	1	0.0115
LYS	3	2	3	1	2	2	2	2	50	1	0.0200
LYS	3	2	3	1	3	1	2	2	50	1	0.0200
LYS	3	2	3	1	3	2	2	2	50	36	0.7200
LYS	3	2	3	1	3	2	3	2	50	12	0.2400
LYS	3	2	3	2	2	2	2	2	28	2	0.0714
LYS	3	2	3	2	2	2	3	1	28	1	0.0357
LYS	3	2	3	2	3	2	1	2	28	4	0.1429
LYS	3	2	3	2	3	2	2	2	28	9	0.3214
LYS	3	2	3	2	3	2	2	3	28	6	0.2143
LYS	3	2	3	2	3	2	3	1	28	1	0.0357
LYS	3	2	3	3	1	2	3	2	4	1	0.2500
LYS	3	2	3	3	1	2	3	3	4	1	0.2500
LYS	3	3	1	2	3	1	3	2	13	1	0.0769
LYS	3	3	1	2	3	2	2	2	13	2	0.1538
LYS	3	3	1	2	3	2	3	3	13	1	0.0769
LYS	3	3	1	2	3	3	2	2	13	7	0.5385
LYS	3	3	1	2	3	3	2	3	13	1	0.0769
LYS	3	3	1	2	3	3	3	2	13	1	0.0769
LYS	3	3	2	1	2	2	2	1	33	1	0.0303
LYS	3	3	2	1	3	2	2	2	33	8	0.2424
LYS	3	3	2	1	3	3	2	2	33	12	0.3636
LYS	3	3	2	1	3	3	2	3	33	12	0.3636
LYS	3	3	2	2	1	2	2	2	877	1	0.0011
LYS	3	3	2	2	1	3	3	3	877	1	0.0011
LYS	3	3	2	2	2	2	3	1	877	1	0.0011
LYS	3	3	2	2	3	1	2	2	877	29	0.0331
LYS	3	3	2	2	3	1	3	2	877	25	0.0285
LYS	3	3	2	2	3	2	2	1	877	65	0.0741
LYS	3	3	2	2	3	2	2	2	877	95	0.1083
LYS	3	3	2	2	3	2	2	3	877	1	0.0011
LYS	3	3	2	2	3	2	3	2	877	1	0.0011
LYS	3	3	2	2	3	3	2	3	877	88	0.1003
LYS	3	3	2	3	1	2	2	2	140	1	0.0071
LYS	3	3	2	3	1	3	3	3	140	1	0.0071
LYS	3	3	2	3	2	2	3	1	140	1	0.0071
LYS	3	3	2	3	3	1	2	2	140	4	0.0286
LYS	3	3	2	3	3	1	3	2	140	3	0.0214
LYS	3	3	2	3	3	2	2	1	140	18	0.1286
LYS	3	3	2	3	3	2	2	2	140	13	0.0929
LYS	3	3	2	3	3	2	3	3	140	1	0.0071

AA	unbound				complex				#	# change	P change
	r1	r2	r3	r4	r1	r2	r3	r4			
LYS	3	3	2	3	3	3	2	2	140	78	0.5571
LYS	3	3	2	3	3	3	3	2	140	1	0.0071
LYS	3	3	3	3	1	2	3	2	6	1	0.1667
LYS	3	3	3	3	1	2	3	3	6	1	0.1667
LYS	3	3	3	3	3	2	2	1	6	1	0.1667
LYS	3	3	3	3	3	3	2	2	6	3	0.5000
MET	1	2	1	0	1	2	2	0	4	1	0.2500
MET	1	2	1	0	1	2	3	0	4	3	0.7500
MET	1	2	3	0	2	2	1	0	14	13	0.9286
MET	1	2	3	0	2	2	3	0	14	1	0.0714
MET	2	2	1	0	2	2	3	0	400	28	0.0700
MET	2	2	1	0	3	3	3	0	400	1	0.0025
MET	2	2	3	0	2	1	1	0	2	1	0.5000
MET	3	2	1	0	3	2	3	0	175	1	0.0057
MET	3	2	3	0	3	2	2	0	9	2	0.2222
MET	3	3	1	0	3	2	1	0	6	4	0.6667
PHE	2	1	0	0	3	1	0	0	47	4	0.0851
PHE	2	1	0	0	3	2	0	0	47	2	0.0426
PHE	3	1	0	0	1	1	0	0	551	4	0.0073
PHE	3	1	0	0	2	1	0	0	551	1	0.0018
PHE	3	1	0	0	3	2	0	0	551	40	0.0726
PHE	3	2	0	0	3	1	0	0	449	23	0.0512
SER	1	0	0	0	0	0	0	0	6454	1	0.0002
SER	1	0	0	0	2	0	0	0	6454	455	0.0705
SER	1	0	0	0	3	0	0	0	6454	624	0.0967
SER	2	0	0	0	1	0	0	0	2604	291	0.1118
SER	2	0	0	0	3	0	0	0	2604	168	0.0645
SER	3	0	0	0	1	0	0	0	4853	1597	0.3291
SER	3	0	0	0	2	0	0	0	4853	131	0.0270
THR	1	0	0	0	2	0	0	0	2399	28	0.0117
THR	1	0	0	0	3	0	0	0	2399	247	0.1030
THR	2	0	0	0	1	0	0	0	71	30	0.4225
THR	2	0	0	0	3	0	0	0	71	21	0.2958
THR	3	0	0	0	1	0	0	0	2735	93	0.0340
THR	3	0	0	0	2	0	0	0	2735	34	0.0124
TRP	2	1	0	0	3	3	0	0	13	1	0.0769
TRP	2	3	0	0	1	3	0	0	584	1	0.0017
TRP	2	3	0	0	2	2	0	0	584	1	0.0017
TRP	3	1	0	0	3	3	0	0	614	28	0.0456
TRP	3	2	0	0	1	1	0	0	20	1	0.0500
TRP	3	2	0	0	3	3	0	0	20	14	0.7000
TYR	1	1	0	0	3	1	0	0	1240	4	0.0032
TYR	2	1	0	0	2	2	0	0	1027	1	0.0010
TYR	3	1	0	0	1	1	0	0	2475	2	0.0008
TYR	3	1	0	0	1	2	0	0	2475	2	0.0008
TYR	3	1	0	0	3	2	0	0	2475	1	0.0004
TYR	3	2	0	0	3	1	0	0	12	2	0.1667
VAL	1	0	0	0	2	0	0	0	446	151	0.3386
VAL	1	0	0	0	3	0	0	0	446	11	0.0247
VAL	2	0	0	0	1	0	0	0	7094	152	0.0214
VAL	2	0	0	0	3	0	0	0	7094	646	0.0911
VAL	3	0	0	0	1	0	0	0	785	8	0.0102
VAL	3	0	0	0	2	0	0	0	785	34	0.0433

Table D.4.: Probability for a rotamer change of unbound rotamer sets to different complex rotamer sets

AName	avg Chi1	$P(r1 \mid \phi\psi))$	#	Phirot	Psirot	$P(\chi_1 \mid \phi)$	$P(chi_1 \mid \psi)$
ARG	58.864	0.002360	10	1	18	0.008914	0.008911
ARG	66.438	0.001866	8	2	17	0.012707	0.009670
ARG	60.218	0.001615	7	2	18	0.012707	0.008911
ARG	54.560	0.002590	11	3	10	0.017639	0.006636
ARG	65.195	0.002330	10	3	11	0.017639	0.005498
ARG	63.800	0.003359	14	3	17	0.017639	0.009670
ARG	75.402	0.001615	7	3	18	0.017639	0.008911
ARG	66.620	0.003350	14	6	8	0.011949	0.008532
ARG	68.391	0.003330	14	6	9	0.011949	0.006636
ARG	181.536	0.002156	9	3	16	0.021099	0.042002
ARG	184.805	0.002410	10	4	16	0.007413	0.042002
ARG	189.965	0.002156	9	5	14	0.031743	0.005132
ARG	176.880	0.003425	14	5	15	0.031743	0.014634
ARG	186.333	0.009006	36	5	16	0.031743	0.042002
ARG	185.371	0.002664	11	5	17	0.031743	0.012734
ARG	184.992	0.015348	61	6	7	0.072040	0.050745
ARG	183.448	0.014840	59	6	8	0.072040	0.034020
ARG	187.395	0.003425	14	6	15	0.072040	0.014634
ARG	194.232	0.009513	38	6	16	0.072040	0.042002
ARG	176.930	0.002664	11	6	17	0.072040	0.012734
ARG	186.729	0.002156	9	7	6	0.043908	0.005512
ARG	187.001	0.016109	64	7	7	0.043908	0.050745
ARG	187.266	0.006723	27	7	8	0.043908	0.034020
ARG	197.485	0.003425	14	7	16	0.043908	0.042002
ARG	280.871	0.001385	6	2	16	0.021107	0.024526
ARG	278.666	0.003659	15	2	17	0.021107	0.035934
ARG	286.227	0.002628	11	2	18	0.021107	0.013879
ARG	309.880	0.005686	23	3	9	0.067503	0.048862
ARG	294.795	0.002391	10	3	15	0.067503	0.023385
ARG	295.248	0.002140	9	3	16	0.067503	0.024526
ARG	302.379	0.006435	26	3	17	0.067503	0.035934
ARG	295.126	0.002127	9	3	18	0.067503	0.013879
ARG	299.393	0.002401	10	4	8	0.056094	0.050763
ARG	300.829	0.007455	30	4	9	0.056094	0.048862
ARG	303.521	0.002373	10	4	10	0.056094	0.012358
ARG	298.700	0.008933	36	4	15	0.056094	0.023385
ARG	294.774	0.003650	15	4	16	0.056094	0.024526
ARG	296.370	0.005173	21	4	17	0.056094	0.035934
ARG	298.368	0.002878	12	4	18	0.056094	0.013879
ARG	291.559	0.005181	21	5	8	0.049249	0.050763
ARG	295.907	0.012256	49	5	9	0.049249	0.048862
ARG	290.261	0.001374	6	5	10	0.049249	0.012358
ARG	278.455	0.002391	10	5	15	0.049249	0.023385
ARG	288.888	0.002895	12	5	16	0.049249	0.024526
ARG	291.375	0.004164	17	5	17	0.049249	0.035934
ARG	286.858	0.006667	27	6	7	0.062179	0.023005
ARG	291.552	0.018829	75	6	8	0.062179	0.050763
ARG	292.838	0.005938	24	6	9	0.062179	0.048862
ARG	275.005	0.003398	14	6	16	0.062179	0.024526
ARG	289.388	0.003659	15	6	17	0.062179	0.035934
ARG	285.475	0.007171	29	7	7	0.024910	0.023005
ARG	293.099	0.005939	24	7	8	0.024910	0.050763
ARG	284.749	0.001888	8	7	16	0.024910	0.024526
ARG	291.024	0.007668	31	12	11	0.017304	0.021104
ARG	296.131	0.002569	11	12	12	0.017304	0.005133
ARG	291.233	0.002873	12	13	10	0.010839	0.012358
ARG	298.885	0.003646	15	13	11	0.010839	0.021104
ASN	60.368	0.008954	17	1	1	0.033981	0.029715
ASN	71.157	0.006794	39	1	18	0.033981	0.015590
ASN	55.838	0.005940	32	2	1	0.015858	0.029715
ASN	50.452	0.002380	14	2	2	0.015858	0.004397
ASN	68.057	0.001147	7	2	18	0.015858	0.015590
ASN	63.321	0.001139	7	3	10	0.008395	0.008395
ASN	61.425	0.002889	17	3	11	0.008395	0.008128
ASN	60.204	0.001664	10	4	10	0.006796	0.008395
ASN	46.405	0.001313	8	4	11	0.006796	0.008128
ASN	65.379	0.001489	9	5	10	0.013193	0.008395
ASN	61.533	0.000963	6	5	11	0.013193	0.008128
ASN	57.146	0.001614	10	5	14	0.013193	0.002798
ASN	58.830	0.000941	6	5	17	0.013193	0.003331
ASN	174.962	0.001331	8	2	15	0.012672	0.047349
ASN	180.736	0.000976	6	2	16	0.012672	0.053751
ASN	183.387	0.001646	10	3	13	0.029746	0.004935
ASN	181.591	0.004171	24	3	15	0.029746	0.047349
ASN	182.874	0.007190	41	3	16	0.029746	0.053751
ASN	185.493	0.012334	70	4	15	0.033214	0.047349
ASN	181.195	0.007545	43	4	16	0.033214	0.053751
ASN	185.294	0.008962	51	5	15	0.028145	0.047349

AName	avg Chi1	$P(r1 \mid \phi\psi))$	#	Phirot	Psirot	$P(\chi_1 \mid \phi)$	$P(chi_1 \mid \psi)$
ASN	184.978	0.005237	30	5	16	0.028145	0.053751
ASN	186.807	0.002033	12	5	17	0.028145	0.019606
ASN	196.755	0.000930	6	5	18	0.028145	0.002534
ASN	180.434	0.001459	9	6	7	0.038549	0.003601
ASN	190.020	0.001806	11	6	8	0.038549	0.003868
ASN	196.187	0.002850	17	6	9	0.038549	0.004401
ASN	185.329	0.004171	24	6	15	0.038549	0.047349
ASN	187.709	0.011273	64	6	16	0.038549	0.053751
ASN	186.382	0.002563	15	6	17	0.038549	0.019606
ASN	190.633	0.002929	17	7	16	0.015873	0.053751
ASN	191.246	0.006805	39	7	17	0.015873	0.019606
ASN	204.988	0.005206	30	12	11	0.017207	0.015872
ASN	197.561	0.005721	33	12	12	0.017207	0.012137
ASN	208.227	0.004853	28	13	11	0.009471	0.015872
ASN	285.091	0.004312	25	2	16	0.029752	0.011873
ASN	287.056	0.001144	7	3	16	0.013742	0.011873
ASN	288.877	0.000967	6	3	17	0.013742	0.010539
ASN	281.958	0.005588	32	4	9	0.049231	0.042557
ASN	295.440	0.014293	81	4	10	0.049231	0.057765
ASN	298.284	0.003279	19	4	11	0.049231	0.032418
ASN	297.596	0.000968	6	4	16	0.049231	0.011873
ASN	299.058	0.003427	20	4	17	0.049231	0.010539
ASN	298.822	0.002360	14	4	18	0.049231	0.007337
ASN	285.458	0.002392	14	5	7	0.055635	0.030817
ASN	289.071	0.008066	46	5	8	0.055635	0.034019
ASN	287.743	0.009846	56	5	9	0.055635	0.042557
ASN	294.439	0.010920	62	5	10	0.055635	0.057765
ASN	299.331	0.000975	6	5	11	0.055635	0.032418
ASN	289.537	0.001142	7	5	17	0.055635	0.010539
ASN	304.354	0.001835	11	5	18	0.055635	0.007337
ASN	285.746	0.011075	63	6	7	0.055101	0.030817
ASN	288.919	0.011966	68	6	8	0.055101	0.034019
ASN	290.055	0.010024	57	6	9	0.055101	0.042557
ASN	288.094	0.000977	6	6	10	0.055101	0.057765
ASN	287.912	0.001142	7	6	17	0.055101	0.010539
ASN	286.509	0.006468	37	7	7	0.013742	0.030817
ASN	288.911	0.001507	9	7	8	0.013742	0.034019
ASN	295.951	0.001687	10	12	10	0.030819	0.057765
ASN	301.983	0.010368	59	12	11	0.030819	0.032418
ASN	290.732	0.007525	43	12	12	0.030819	0.025748
ASN	293.631	0.009499	54	13	10	0.036690	0.057765
ASN	293.509	0.005937	34	13	11	0.036690	0.032418
ASN	301.715	0.008056	46	13	12	0.036690	0.025748
ASN	293.963	0.001153	7	14	9	0.002535	0.042557
ASP	63.245	0.001932	8	3	10	0.005232	0.042508
ASP	61.639	0.001666	7	4	9	0.023835	0.019992
ASP	65.030	0.009403	37	4	10	0.023835	0.042508
ASP	63.070	0.001369	6	4	11	0.023835	0.004853
ASP	67.991	0.006023	24	5	9	0.032362	0.019992
ASP	67.326	0.010949	43	5	10	0.032362	0.042508
ASP	59.562	0.002613	11	6	8	0.029261	0.004853
ASP	66.830	0.004998	20	6	9	0.029261	0.019992
ASP	67.397	0.005281	21	6	10	0.029261	0.042508
ASP	67.877	0.003424	14	6	18	0.029261	0.009511
ASP	188.653	0.002377	10	1	17	0.023077	0.005629
ASP	195.343	0.009599	38	1	18	0.023077	0.017663
ASP	192.261	0.002701	11	2	13	0.026956	0.031250
ASP	183.219	0.006250	25	3	6	0.022689	0.013393
ASP	189.949	0.001675	7	3	15	0.022689	0.047554
ASP	187.442	0.001667	7	3	16	0.022689	0.020769
ASP	181.043	0.003995	16	4	15	0.013381	0.047554
ASP	189.220	0.009902	39	5	13	0.059148	0.031250
ASP	180.705	0.008336	33	5	14	0.059148	0.021545
ASP	181.128	0.014563	57	5	15	0.059148	0.047554
ASP	180.699	0.004231	17	5	16	0.059148	0.020769
ASP	183.564	0.002679	11	6	7	0.054493	0.013393
ASP	182.226	0.004679	19	6	8	0.054493	0.008346
ASP	199.341	0.002364	10	6	9	0.054493	0.004852
ASP	187.851	0.006301	25	6	13	0.054493	0.031250
ASP	181.223	0.003719	15	6	14	0.054493	0.021545
ASP	188.743	0.008892	35	6	15	0.054493	0.047554
ASP	188.591	0.004487	18	6	16	0.054493	0.020769
ASP	194.200	0.004719	19	7	7	0.012605	0.013393
ASP	292.391	0.003484	14	4	8	0.026182	0.068517
ASP	292.825	0.002702	11	4	9	0.026182	0.033579
ASP	295.084	0.005259	21	4	16	0.026182	0.021933
ASP	299.136	0.001654	7	4	17	0.026182	0.011452
ASP	283.519	0.001415	6	5	7	0.070787	0.032803

AName	avg Chi1	$P(r1 \mid \phi\psi))$	#	Phirot	Psirot	$P(\chi_1 \mid \phi)$	$P(chi_1 \mid \psi)$
ASP	292.379	0.012774	50	5	8	0.070787	0.068517
ASP	290.277	0.016082	63	5	9	0.070787	0.033579
ASP	297.856	0.002914	12	5	10	0.070787	0.009123
ASP	296.305	0.007055	28	5	16	0.070787	0.021933
ASP	298.749	0.002672	11	5	17	0.070787	0.011452
ASP	296.152	0.002139	9	5	18	0.070787	0.006793
ASP	285.964	0.011449	45	6	7	0.063418	0.032803
ASP	287.812	0.023613	92	6	8	0.063418	0.068517
ASP	287.569	0.002702	11	6	9	0.063418	0.033579
ASP	285.960	0.008104	32	7	7	0.022303	0.032803
ASP	282.908	0.004516	18	7	8	0.022303	0.068517
ASP	290.859	0.002646	11	13	11	0.007951	0.007182
CYS	55.687	0.002146	8	1	17	0.009354	0.013266
CYS	65.240	0.003291	12	2	17	0.016315	0.013266
CYS	59.683	0.003585	13	2	18	0.016315	0.015441
CYS	68.263	0.002432	9	3	17	0.023276	0.013266
CYS	60.879	0.003585	13	3	18	0.023276	0.015441
CYS	59.977	0.001578	6	5	18	0.003698	0.015441
CYS	62.471	0.006717	24	7	8	0.016315	0.012396
CYS	61.947	0.003823	14	7	9	0.016315	0.008047
CYS	197.744	0.001874	7	1	1	0.016338	0.014149
CYS	197.915	0.001569	6	1	18	0.016338	0.011101
CYS	198.892	0.007353	26	2	1	0.040737	0.014149
CYS	199.219	0.005563	20	2	18	0.040737	0.011101
CYS	178.652	0.001466	6	3	17	0.029845	0.002394
CYS	187.106	0.025629	89	6	7	0.055987	0.049412
CYS	190.407	0.008733	31	6	8	0.055987	0.013714
CYS	185.332	0.005647	20	7	7	0.009367	0.049412
CYS	275.099	0.002467	9	1	17	0.023331	0.093092
CYS	274.014	0.006516	23	1	18	0.023331	0.050797
CYS	304.869	0.005940	21	2	16	0.029872	0.056466
CYS	288.988	0.002467	9	2	17	0.029872	0.093092
CYS	276.984	0.007965	28	2	18	0.029872	0.050797
CYS	316.871	0.018399	64	3	16	0.167239	0.056466
CYS	287.813	0.019007	66	3	17	0.167239	0.093092
CYS	305.640	0.014626	51	3	18	0.167239	0.050797
CYS	301.748	0.007957	28	4	7	0.049060	0.041205
CYS	298.722	0.003042	11	4	16	0.049060	0.056466
CYS	293.311	0.016686	58	4	17	0.049060	0.093092
CYS	294.333	0.002462	9	5	8	0.073481	0.049489
CYS	289.767	0.014857	52	5	9	0.073481	0.026380
CYS	289.213	0.008752	31	5	10	0.073481	0.015915
CYS	302.257	0.005940	21	5	16	0.073481	0.056466
CYS	299.547	0.012623	44	5	17	0.073481	0.093092
CYS	297.970	0.002462	9	5	18	0.073481	0.050797
CYS	295.958	0.009403	33	6	7	0.072609	0.041205
CYS	292.523	0.021575	75	6	8	0.072609	0.049489
CYS	291.989	0.001587	6	6	9	0.072609	0.026380
CYS	299.324	0.003622	13	6	16	0.072609	0.056466
CYS	303.781	0.007980	28	6	17	0.072609	0.093092
CYS	300.473	0.008825	31	7	7	0.025511	0.041205
CYS	297.647	0.007674	27	7	8	0.025511	0.049489
GLN	58.066	0.002304	9	2	17	0.006782	0.012527
GLN	66.523	0.003117	12	3	17	0.010893	0.012527
GLN	59.906	0.002530	10	3	18	0.010893	0.005955
GLN	36.617	0.001762	7	6	17	0.006371	0.012527
GLN	30.639	0.001709	7	7	16	0.004316	0.004312
GLN	210.624	0.002225	9	3	10	0.023731	0.003920
GLN	179.399	0.006180	23	3	16	0.023731	0.067884
GLN	178.330	0.001766	7	3	17	0.023731	0.015062
GLN	180.406	0.011947	44	4	16	0.024969	0.067884
GLN	184.907	0.002324	9	5	8	0.017540	0.027030
GLN	189.143	0.005356	20	5	16	0.017540	0.067884
GLN	198.842	0.002038	8	5	17	0.017540	0.015062
GLN	182.262	0.010254	38	6	7	0.067065	0.027442
GLN	183.743	0.011073	41	6	8	0.067065	0.027030
GLN	174.144	0.017715	65	6	16	0.067065	0.067884
GLN	185.764	0.003125	12	6	17	0.067065	0.015062
GLN	183.080	0.006152	23	7	7	0.020016	0.027442
GLN	194.462	0.003144	12	7	8	0.020016	0.027030
GLN	183.735	0.002609	10	7	16	0.020016	0.067884
GLN	286.584	0.002893	11	2	17	0.023748	0.053891
GLN	297.293	0.002617	10	3	9	0.094784	0.046457
GLN	259.821	0.001791	7	3	10	0.094784	0.022093
GLN	295.057	0.005372	20	3	15	0.094784	0.014247
GLN	294.474	0.009504	35	3	16	0.094784	0.046044
GLN	302.120	0.007300	27	3	17	0.094784	0.053891
GLN	295.080	0.003444	13	3	18	0.094784	0.011356

AName	avg Chi1	$P(r1 \mid \phi\psi))$	#	Phirot	Psirot	$P(\chi_1 \mid \phi)$	$P(chi_1 \mid \psi)$
GLN	293.003	0.008402	31	4	8	0.062570	0.058846
GLN	301.445	0.016942	62	4	9	0.062570	0.046457
GLN	293.252	0.002893	11	4	10	0.062570	0.022093
GLN	294.462	0.002066	8	4	15	0.062570	0.014247
GLN	300.928	0.002617	10	4	16	0.062570	0.046044
GLN	300.416	0.006198	23	4	17	0.062570	0.053891
GLN	297.225	0.004821	18	5	8	0.066287	0.058846
GLN	298.091	0.005096	19	5	9	0.066287	0.046457
GLN	297.462	0.009780	36	5	10	0.066287	0.022093
GLN	288.688	0.006749	25	5	16	0.066287	0.046044
GLN	293.867	0.011983	44	5	17	0.066287	0.053891
GLN	287.148	0.016116	59	6	7	0.081155	0.034069
GLN	285.401	0.021901	80	6	8	0.081155	0.058846
GLN	287.698	0.005372	20	6	9	0.081155	0.046457
GLN	287.600	0.001791	7	6	16	0.081155	0.046044
GLN	287.398	0.005923	22	6	17	0.081155	0.053891
GLN	295.035	0.002342	9	6	18	0.081155	0.011356
GLN	285.187	0.005096	19	7	7	0.029117	0.034069
GLN	290.732	0.003168	12	7	8	0.029117	0.058846
GLN	294.383	0.008953	33	7	16	0.029117	0.046044
GLN	291.091	0.008402	31	12	12	0.015488	0.013834
GLU	71.454	0.002150	7	5	9	0.007767	0.018289
GLU	52.896	0.008770	27	6	8	0.036832	0.018790
GLU	62.838	0.009097	28	6	9	0.036832	0.018289
GLU	64.383	0.003776	12	6	17	0.036832	0.011775
GLU	61.759	0.002813	9	7	8	0.005763	0.018790
GLU	165.080	0.001831	6	2	16	0.012806	0.033892
GLU	182.966	0.004494	14	3	16	0.017326	0.033892
GLU	176.373	0.001831	6	4	16	0.007784	0.033892
GLU	184.088	0.001793	6	5	15	0.018832	0.008787
GLU	180.631	0.005493	17	5	16	0.018832	0.033892
GLU	183.474	0.002788	9	5	17	0.018832	0.011297
GLU	183.531	0.018514	56	6	7	0.059510	0.047449
GLU	187.478	0.010456	32	6	8	0.059510	0.024352
GLU	188.948	0.005825	18	6	16	0.059510	0.033892
GLU	187.991	0.010841	33	7	7	0.028876	0.047449
GLU	192.737	0.005145	16	7	8	0.028876	0.024352
GLU	190.803	0.002164	7	7	16	0.028876	0.033892
GLU	291.147	0.002850	9	3	16	0.023875	0.018847
GLU	303.423	0.002515	8	3	17	0.023875	0.058049
GLU	299.085	0.007210	22	4	9	0.030409	0.046992
GLU	304.166	0.002515	8	4	10	0.030409	0.015832
GLU	297.236	0.002850	9	4	11	0.030409	0.011308
GLU	287.679	0.006875	21	5	7	0.102285	0.054029
GLU	292.833	0.004527	14	5	8	0.102285	0.059055
GLU	296.805	0.014252	43	5	9	0.102285	0.046992
GLU	294.653	0.006204	19	5	10	0.102285	0.015832
GLU	303.983	0.002515	8	5	13	0.102285	0.006282
GLU	288.431	0.003856	12	5	16	0.102285	0.018847
GLU	294.347	0.023307	70	5	17	0.102285	0.058049
GLU	301.247	0.002180	7	5	18	0.102285	0.010303
GLU	291.475	0.017270	52	6	7	0.105300	0.054029
GLU	290.475	0.027331	82	6	8	0.105300	0.059055
GLU	290.936	0.008216	25	6	9	0.105300	0.046992
GLU	290.574	0.002180	7	6	16	0.105300	0.018847
GLU	301.509	0.010228	31	6	17	0.105300	0.058049
GLU	284.866	0.010563	32	7	7	0.029404	0.054029
GLU	291.809	0.006875	21	7	8	0.029404	0.059055
GLU	296.157	0.001844	6	12	11	0.005780	0.011308
HIS	80.249	0.031911	50	6	9	0.055783	0.048988
HIS	178.729	0.004091	7	2	17	0.032649	0.014129
HIS	184.517	0.006097	10	3	7	0.022903	0.033616
HIS	188.188	0.003530	6	5	7	0.016081	0.033616
HIS	191.616	0.006738	11	6	7	0.083328	0.033616
HIS	171.584	0.043140	67	6	16	0.083328	0.135927
HIS	190.346	0.003530	6	7	7	0.062862	0.033616
HIS	191.102	0.034706	54	7	16	0.062862	0.135927
HIS	294.090	0.004120	7	3	17	0.050198	0.018024
HIS	296.251	0.003545	6	4	10	0.027779	0.048226
HIS	270.779	0.004754	8	4	17	0.027779	0.018024
HIS	294.691	0.003528	6	5	8	0.089185	0.032638
HIS	288.068	0.012509	20	5	9	0.089185	0.032638
HIS	286.356	0.024814	39	5	10	0.089185	0.048226
HIS	314.257	0.008647	14	5	16	0.089185	0.029715
HIS	290.267	0.003357	6	6	7	0.051172	0.007307
HIS	289.305	0.016358	26	6	8	0.051172	0.032638
HIS	296.318	0.006094	10	6	9	0.051172	0.032638
HIS	283.519	0.004164	7	6	16	0.051172	0.029715

AName	avg Chi1	$P(r1 \mid \phi\psi))$	#	Phirot	Psirot	$P(\chi_1 \mid \phi)$	$P(chi_1 \mid \psi)$
ILE	46.042	0.003616	18	2	17	0.007033	0.015472
ILE	59.526	0.003616	18	3	17	0.011409	0.015472
ILE	68.055	0.002325	12	3	18	0.011409	0.004845
ILE	56.543	0.012577	61	4	9	0.032665	0.038915
ILE	59.229	0.006320	31	4	10	0.032665	0.021411
ILE	59.962	0.003638	18	5	9	0.014535	0.038915
ILE	64.031	0.005284	26	5	10	0.014535	0.021411
ILE	67.536	0.005242	26	6	8	0.027351	0.010159
ILE	55.667	0.008835	43	6	9	0.027351	0.038915
ILE	67.679	0.001761	9	6	10	0.027351	0.021411
ILE	56.178	0.001137	6	6	17	0.027351	0.015472
ILE	190.214	0.003818	19	2	16	0.056749	0.013905
ILE	188.222	0.014869	72	2	17	0.056749	0.042341
ILE	182.977	0.002374	12	3	16	0.006097	0.013905
ILE	185.837	0.001342	7	4	16	0.016415	0.013905
ILE	183.450	0.008838	43	4	17	0.016415	0.042341
ILE	188.313	0.001352	7	5	17	0.003283	0.042341
ILE	198.568	0.006732	33	6	8	0.014226	0.020467
ILE	197.539	0.005696	28	7	8	0.011725	0.020467
ILE	190.068	0.001144	6	7	17	0.011725	0.042341
ILE	301.443	0.010916	53	3	15	0.155702	0.040525
ILE	300.131	0.036800	177	3	16	0.155702	0.144107
ILE	305.062	0.002374	12	3	17	0.155702	0.014239
ILE	299.255	0.001076	6	4	14	0.059308	0.002347
ILE	302.807	0.008837	43	4	15	0.059308	0.040525
ILE	300.970	0.023456	113	4	16	0.059308	0.144107
ILE	301.774	0.003819	19	4	17	0.059308	0.014239
ILE	294.160	0.002188	11	5	7	0.045224	0.088091
ILE	302.278	0.005926	29	5	15	0.045224	0.040525
ILE	299.785	0.019078	92	5	16	0.045224	0.144107
ILE	295.302	0.001961	10	5	17	0.045224	0.014239
ILE	294.337	0.036566	176	6	7	0.105314	0.088091
ILE	291.867	0.015047	73	6	8	0.105314	0.027382
ILE	294.715	0.015533	75	6	16	0.105314	0.144107
ILE	294.154	0.017814	86	7	7	0.031766	0.088091
ILE	288.465	0.002802	14	7	8	0.031766	0.027382
LEU	62.729	0.000929	7	2	18	0.002118	0.001929
LEU	173.926	0.007326	49	2	16	0.022115	0.064117
LEU	230.307	0.001248	9	3	14	0.009640	0.003745
LEU	176.966	0.001432	10	3	15	0.009640	0.030300
LEU	183.869	0.002492	17	3	16	0.009640	0.064117
LEU	180.362	0.004296	29	4	15	0.019393	0.030300
LEU	185.017	0.008534	57	4	16	0.019393	0.064117
LEU	184.774	0.000955	7	5	14	0.025744	0.003745
LEU	184.066	0.006859	46	5	15	0.025744	0.030300
LEU	186.397	0.007779	52	5	16	0.025744	0.064117
LEU	186.142	0.014560	97	6	7	0.062035	0.040059
LEU	186.867	0.006092	41	6	8	0.062035	0.018952
LEU	183.146	0.007311	49	6	15	0.062035	0.030300
LEU	191.062	0.011253	75	6	16	0.062035	0.064117
LEU	194.782	0.011542	77	7	7	0.033909	0.040059
LEU	192.457	0.005941	40	7	8	0.033909	0.018952
LEU	187.992	0.004456	30	7	16	0.033909	0.064117
LEU	292.046	0.003858	26	3	16	0.028021	0.055152
LEU	292.140	0.002497	17	3	17	0.028021	0.081253
LEU	289.698	0.000984	7	3	18	0.028021	0.019746
LEU	300.217	0.000984	7	4	8	0.078618	0.060145
LEU	299.424	0.004161	28	4	9	0.078618	0.036314
LEU	288.435	0.009759	65	4	10	0.078618	0.023831
LEU	298.028	0.003707	25	4	15	0.078618	0.010213
LEU	294.524	0.012634	84	4	16	0.078618	0.055152
LEU	299.343	0.012634	84	4	17	0.078618	0.081253
LEU	301.377	0.005825	39	4	18	0.078618	0.019746
LEU	284.186	0.005825	39	5	8	0.101987	0.060145
LEU	292.571	0.008851	59	5	9	0.101987	0.036314
LEU	298.713	0.003253	22	5	10	0.101987	0.023831
LEU	306.931	0.000832	6	5	11	0.101987	0.009532
LEU	281.683	0.001437	10	5	14	0.101987	0.003858
LEU	293.996	0.002043	14	5	15	0.101987	0.010213
LEU	298.592	0.012937	86	5	16	0.101987	0.055152
LEU	298.534	0.025798	171	5	17	0.101987	0.081253
LEU	297.382	0.004766	32	5	18	0.101987	0.019746
LEU	285.273	0.012180	81	6	7	0.101533	0.027009
LEU	287.408	0.026857	178	6	8	0.101533	0.060145
LEU	290.070	0.010516	70	6	9	0.101533	0.036314
LEU	292.658	0.004766	32	6	16	0.101533	0.055152
LEU	296.253	0.010213	68	6	17	0.101533	0.081253
LEU	289.663	0.001135	8	6	18	0.101533	0.019746

AName	avg Chi1	$P(r1 \mid \phi\psi))$	#	Phirot	Psirot	$P(\chi_1 \mid \phi)$	$P(chi_1 \mid \psi)$
LEU	276.097	0.004918	33	7	7	0.023256	0.027009
LEU	288.471	0.005674	38	7	8	0.023256	0.060145
LEU	303.100	0.001891	13	7	16	0.023256	0.055152
LEU	297.254	0.002345	16	7	17	0.023256	0.081253
LEU	284.057	0.002194	15	13	10	0.009643	0.023831
LEU	285.798	0.004161	28	13	11	0.009643	0.009532
LYS	65.388	0.002079	10	1	17	0.015737	0.012736
LYS	65.002	0.002482	12	1	18	0.015737	0.006120
LYS	64.623	0.005360	25	2	17	0.017062	0.012736
LYS	59.027	0.001187	6	2	18	0.017062	0.006120
LYS	68.636	0.002909	14	6	8	0.008448	0.005789
LYS	180.437	0.001437	7	1	16	0.017109	0.059290
LYS	177.845	0.001649	8	1	17	0.017109	0.020096
LYS	180.167	0.001203	6	2	15	0.032390	0.012788
LYS	193.126	0.002763	13	2	16	0.032390	0.059290
LYS	191.045	0.005385	25	2	17	0.032390	0.020096
LYS	190.374	0.001658	8	3	16	0.017109	0.059290
LYS	178.898	0.001860	9	4	15	0.012126	0.012788
LYS	193.310	0.002763	13	4	16	0.012126	0.059290
LYS	186.682	0.002088	10	4	17	0.012126	0.020096
LYS	184.410	0.001436	7	5	7	0.039367	0.050654
LYS	188.579	0.001210	6	5	8	0.039367	0.021424
LYS	192.734	0.008228	38	5	14	0.039367	0.016442
LYS	184.385	0.002516	12	5	15	0.039367	0.012788
LYS	187.780	0.009835	45	5	16	0.039367	0.059290
LYS	186.958	0.017783	81	6	7	0.072588	0.050654
LYS	184.973	0.010666	49	6	8	0.072588	0.021424
LYS	207.510	0.001164	6	6	9	0.072588	0.003488
LYS	197.553	0.001426	7	6	14	0.072588	0.016442
LYS	190.978	0.002079	10	6	15	0.072588	0.012788
LYS	183.797	0.011382	52	6	16	0.072588	0.059290
LYS	176.149	0.002528	12	6	17	0.072588	0.020096
LYS	182.142	0.013586	62	7	7	0.038370	0.050654
LYS	182.434	0.002089	10	7	8	0.038370	0.021424
LYS	185.492	0.008951	41	7	16	0.038370	0.059290
LYS	282.750	0.003203	15	2	17	0.017113	0.048342
LYS	273.347	0.002302	11	2	18	0.017113	0.015117
LYS	290.303	0.011314	52	3	13	0.122114	0.018772
LYS	297.912	0.001196	6	3	15	0.122114	0.008805
LYS	294.118	0.010714	49	3	16	0.122114	0.050668
LYS	303.655	0.012258	56	3	17	0.122114	0.048342
LYS	297.560	0.002302	11	4	9	0.053331	0.015117
LYS	289.055	0.001413	7	4	10	0.053331	0.008472
LYS	296.260	0.001414	7	4	15	0.053331	0.008805
LYS	289.544	0.011377	52	4	16	0.053331	0.050668
LYS	296.064	0.012038	55	4	17	0.053331	0.048342
LYS	311.864	0.005371	25	4	18	0.053331	0.015117
LYS	294.735	0.001216	6	5	8	0.032065	0.061299
LYS	290.453	0.003617	17	5	9	0.032065	0.015117
LYS	291.290	0.002500	12	5	10	0.032065	0.008472
LYS	288.986	0.007400	34	5	16	0.032065	0.050668
LYS	294.445	0.002540	12	5	17	0.032065	0.048342
LYS	288.301	0.013345	61	6	7	0.075262	0.035052
LYS	287.107	0.028845	131	6	8	0.075262	0.061299
LYS	288.149	0.002302	11	6	9	0.075262	0.015117
LYS	290.562	0.002320	11	6	16	0.075262	0.050668
LYS	298.496	0.001215	6	6	17	0.075262	0.048342
LYS	289.659	0.009375	43	7	7	0.031069	0.035052
LYS	286.647	0.009836	45	7	8	0.031069	0.061299
LYS	286.527	0.001603	8	13	11	0.003489	0.004485
MET	73.376	0.007862	10	2	17	0.013223	0.014504
MET	160.119	0.027687	33	2	17	0.094912	0.047280
MET	175.514	0.014869	18	4	16	0.026472	0.039508
MET	180.805	0.006372	8	5	16	0.009685	0.039508
MET	184.181	0.009639	12	6	7	0.023889	0.021373
MET	199.258	0.004410	6	6	8	0.023889	0.008420
MET	296.557	0.009018	11	2	17	0.023929	0.117008
MET	289.634	0.006932	9	3	14	0.188200	0.010990
MET	301.440	0.046805	55	3	17	0.188200	0.117008
MET	305.825	0.005301	7	3	18	0.188200	0.010990
MET	297.423	0.010735	13	4	17	0.029103	0.117008
MET	299.991	0.004707	6	5	8	0.031690	0.071756
MET	287.813	0.007930	10	5	10	0.031690	0.018747
MET	287.059	0.024288	29	6	7	0.140342	0.048484
MET	289.037	0.036373	43	6	8	0.140342	0.071756
MET	286.509	0.018246	22	6	9	0.140342	0.036848
MET	292.612	0.005582	7	6	17	0.140342	0.117008
MET	285.804	0.007244	9	7	7	0.025223	0.048484

AName	avg Chi1	$P(r1 \mid \phi\psi))$	#	Phirot	Psirot	$P(\chi_1 \mid \phi)$	$P(chi_1 \mid \psi)$
MET	288.073	0.004707	6	7	8	0.025223	0.071756
PHE	69.509	0.018074	41	2	17	0.044796	0.036031
PHE	67.485	0.009542	22	2	18	0.044796	0.022561
PHE	67.477	0.003258	8	3	8	0.035365	0.009092
PHE	68.324	0.015824	36	3	9	0.035365	0.031990
PHE	66.925	0.002901	7	3	17	0.035365	0.036031
PHE	57.011	0.002897	7	4	9	0.006399	0.031990
PHE	181.237	0.004221	10	2	16	0.019894	0.024589
PHE	172.479	0.002888	7	4	16	0.007755	0.024589
PHE	190.340	0.003279	8	5	15	0.015173	0.011116
PHE	176.565	0.002888	7	5	16	0.015173	0.024589
PHE	181.421	0.019442	44	6	7	0.054286	0.042779
PHE	188.549	0.009518	22	6	8	0.054286	0.019200
PHE	194.881	0.003777	9	6	16	0.054286	0.024589
PHE	173.038	0.007821	18	7	7	0.017196	0.042779
PHE	185.917	0.002877	7	7	8	0.017196	0.019200
PHE	297.137	0.003791	9	3	16	0.061766	0.034081
PHE	301.040	0.011431	26	3	17	0.061766	0.069849
PHE	302.077	0.002876	7	4	11	0.082691	0.018559
PHE	296.108	0.021205	48	4	15	0.082691	0.037455
PHE	293.091	0.007806	18	4	16	0.082691	0.034081
PHE	294.571	0.012776	29	4	17	0.082691	0.069849
PHE	293.724	0.003266	8	4	18	0.082691	0.009786
PHE	295.471	0.013141	30	5	9	0.052990	0.030032
PHE	300.640	0.005099	12	5	10	0.052990	0.021259
PHE	295.318	0.002902	7	5	15	0.052990	0.037455
PHE	286.707	0.003345	8	5	16	0.052990	0.034081
PHE	285.445	0.005155	12	5	17	0.052990	0.069849
PHE	284.669	0.009075	21	6	7	0.070541	0.019234
PHE	288.482	0.014007	32	6	8	0.070541	0.025983
PHE	283.928	0.002450	6	6	9	0.070541	0.030032
PHE	279.387	0.005575	13	6	16	0.070541	0.034081
PHE	278.599	0.014121	32	6	17	0.070541	0.069849
PHE	290.956	0.002877	7	7	7	0.007763	0.019234
PHE	299.808	0.006873	16	13	10	0.021939	0.021259
PHE	303.248	0.006857	16	13	11	0.021939	0.018559
SER	76.572	0.000963	9	1	17	0.013502	0.043578
SER	68.438	0.002775	25	1	18	0.013502	0.044258
SER	54.773	0.006286	56	2	17	0.021994	0.043578
SER	56.656	0.005153	46	2	18	0.021994	0.044258
SER	67.334	0.002775	25	3	9	0.024372	0.056490
SER	65.705	0.001869	17	3	10	0.024372	0.014186
SER	53.325	0.004700	42	3	12	0.024372	0.008240
SER	60.067	0.002888	26	3	17	0.024372	0.043578
SER	60.791	0.001416	13	3	18	0.024372	0.044258
SER	55.171	0.000736	7	4	9	0.014182	0.056490
SER	61.768	0.002095	19	4	10	0.014182	0.014186
SER	60.851	0.000849	8	4	11	0.014182	0.002124
SER	59.279	0.001416	13	4	17	0.014182	0.043578
SER	52.991	0.001982	18	4	18	0.014182	0.044258
SER	48.451	0.004927	44	5	8	0.054942	0.050544
SER	62.733	0.007192	64	5	9	0.054942	0.056490
SER	61.182	0.004474	40	5	10	0.054942	0.014186
SER	54.230	0.005380	48	5	17	0.054942	0.043578
SER	57.307	0.012742	113	5	18	0.054942	0.044258
SER	38.588	0.004247	38	6	7	0.099100	0.008240
SER	56.302	0.020784	184	6	8	0.099100	0.050544
SER	59.358	0.024861	220	6	9	0.099100	0.056490
SER	55.659	0.000849	8	6	10	0.099100	0.014186
SER	63.378	0.010024	89	6	17	0.099100	0.043578
SER	66.115	0.004700	42	6	18	0.099100	0.044258
SER	54.312	0.000849	8	7	7	0.018427	0.008240
SER	62.040	0.007192	64	7	8	0.018427	0.050544
SER	64.434	0.001529	14	7	9	0.018427	0.056490
SER	69.771	0.001756	16	7	17	0.018427	0.043578
SER	184.865	0.005477	49	2	1	0.036590	0.009259
SER	190.962	0.003775	34	2	15	0.036590	0.016735
SER	182.375	0.001749	16	2	16	0.036590	0.023191
SER	180.399	0.000617	6	3	6	0.011800	0.008919
SER	187.136	0.000605	6	3	14	0.011800	0.002803
SER	189.495	0.001521	14	3	15	0.011800	0.016735
SER	169.662	0.001975	18	3	16	0.011800	0.023191
SER	185.421	0.004656	42	4	6	0.025214	0.008919
SER	182.058	0.000843	8	4	7	0.025214	0.010279
SER	182.185	0.002423	22	4	15	0.025214	0.016735
SER	169.786	0.005925	53	4	16	0.025214	0.023191
SER	187.867	0.001184	11	4	17	0.025214	0.018094
SER	184.255	0.002648	24	5	15	0.008914	0.016735

AName	avg Chi1	$P(r1 \mid \phi\psi))$	#	Phirot	Psirot	$P(\chi_1 \mid \phi)$	$P(chi_1 \mid \psi)$
SER	186.658	0.000734	7	5	16	0.008914	0.023191
SER	195.475	0.001409	13	5	17	0.008914	0.018094
SER	176.260	0.003089	28	6	7	0.020799	0.010279
SER	182.581	0.001156	11	6	8	0.020799	0.002973
SER	177.619	0.002088	19	6	16	0.020799	0.023191
SER	174.689	0.005919	53	6	17	0.020799	0.018094
SER	184.006	0.002191	20	7	7	0.010272	0.010279
SER	178.832	0.002201	20	7	16	0.010272	0.023191
SER	174.969	0.001635	15	7	17	0.010272	0.018094
SER	319.134	0.000623	6	3	14	0.047467	0.002803
SER	297.102	0.008551	76	3	16	0.047467	0.027778
SER	300.918	0.004814	43	3	17	0.047467	0.034744
SER	291.799	0.000623	6	4	9	0.023352	0.011468
SER	304.637	0.006173	55	4	10	0.023352	0.012657
SER	311.196	0.002888	26	4	11	0.023352	0.006881
SER	312.599	0.001189	11	4	14	0.023352	0.002803
SER	296.794	0.000963	9	4	15	0.023352	0.007730
SER	298.072	0.001416	13	4	16	0.023352	0.027778
SER	308.698	0.000849	8	4	17	0.023352	0.034744
SER	313.630	0.001076	10	5	8	0.033371	0.024040
SER	298.822	0.003908	35	5	9	0.033371	0.011468
SER	306.152	0.001756	16	5	10	0.033371	0.012657
SER	299.397	0.003908	35	5	15	0.033371	0.007730
SER	290.078	0.001303	12	5	16	0.033371	0.027778
SER	311.815	0.007305	65	5	17	0.033371	0.034744
SER	308.936	0.001416	13	5	18	0.033371	0.005182
SER	299.000	0.010137	90	6	7	0.054090	0.021152
SER	301.260	0.012969	115	6	8	0.054090	0.024040
SER	297.782	0.002548	23	6	9	0.054090	0.011468
SER	292.528	0.002095	19	6	16	0.054090	0.027778
SER	294.177	0.007532	67	6	17	0.054090	0.034744
SER	299.670	0.003228	29	7	7	0.016728	0.021152
SER	300.257	0.001642	15	7	8	0.016728	0.024040
SER	295.774	0.003455	31	7	16	0.016728	0.027778
SER	292.443	0.002095	19	7	17	0.016728	0.034744
SER	303.888	0.001303	12	8	16	0.002123	0.027778
SER	302.908	0.001189	11	12	11	0.007048	0.006881
SER	304.769	0.003341	30	12	12	0.007048	0.005861
THR	59.222	0.002821	15	2	17	0.007465	0.038495
THR	55.211	0.002435	13	3	9	0.040250	0.065133
THR	56.853	0.001649	9	3	10	0.040250	0.022101
THR	56.940	0.001801	10	3	16	0.040250	0.004830
THR	57.506	0.011769	61	3	17	0.040250	0.038495
THR	60.739	0.006911	36	3	18	0.040250	0.048740
THR	51.125	0.002432	13	4	8	0.064253	0.037909
THR	58.941	0.012177	63	4	9	0.064253	0.065133
THR	58.784	0.009990	52	4	10	0.064253	0.022101
THR	56.698	0.001601	9	4	11	0.064253	0.003952
THR	57.341	0.003599	19	4	17	0.064253	0.038495
THR	65.228	0.011000	57	4	18	0.064253	0.048740
THR	45.093	0.004571	24	5	8	0.060155	0.037909
THR	57.282	0.018997	98	5	9	0.060155	0.065133
THR	54.883	0.002231	12	5	10	0.060155	0.022101
THR	62.002	0.001070	6	5	17	0.060155	0.038495
THR	65.758	0.011000	57	5	18	0.060155	0.048740
THR	59.805	0.014297	74	6	8	0.052544	0.037909
THR	61.079	0.009450	49	6	9	0.052544	0.065133
THR	59.618	0.005933	31	6	17	0.052544	0.038495
THR	63.203	0.002239	12	6	18	0.052544	0.048740
THR	52.081	0.003599	19	7	8	0.006001	0.037909
THR	189.620	0.003378	18	2	1	0.022050	0.006557
THR	186.081	0.001419	8	2	17	0.022050	0.004517
THR	190.927	0.001591	9	2	18	0.022050	0.003351
THR	188.548	0.002206	12	3	8	0.010076	0.008014
THR	183.520	0.001438	8	6	8	0.003359	0.008014
THR	285.720	0.001658	9	2	16	0.010101	0.113726
THR	310.310	0.002394	13	3	14	0.079490	0.007465
THR	300.703	0.002999	16	3	15	0.079490	0.015954
THR	301.376	0.011214	58	3	16	0.079490	0.113726
THR	299.937	0.009021	47	3	17	0.079490	0.022394
THR	312.989	0.002011	11	4	14	0.022691	0.007465
THR	305.709	0.003386	18	4	15	0.022691	0.015954
THR	302.812	0.007898	41	4	16	0.022691	0.113726
THR	302.669	0.001850	10	5	7	0.030596	0.049911
THR	300.210	0.001838	10	5	15	0.030596	0.015954
THR	300.301	0.012969	67	5	16	0.030596	0.113726
THR	295.950	0.001261	7	5	17	0.030596	0.022394
THR	299.423	0.019373	100	6	7	0.104962	0.049911

AName	avg Chi1	$P(r1 \mid \phi\psi))$	#	Phirot	Psirot	$P(\chi_1 \mid \phi)$	$P(chi_1 \mid \psi)$
THR	299.592	0.010386	54	6	8	0.104962	0.024736
THR	301.157	0.001645	9	6	15	0.104962	0.015954
THR	302.431	0.034617	178	6	16	0.104962	0.113726
THR	302.030	0.002619	14	6	17	0.104962	0.022394
THR	300.156	0.011195	58	7	7	0.035866	0.049911
THR	300.547	0.004562	24	7	8	0.035866	0.024736
THR	304.239	0.006923	36	7	16	0.035866	0.113726
TRP	71.077	0.002676	6	1	18	0.011585	0.014593
TRP	59.902	0.014050	29	2	1	0.033259	0.046025
TRP	62.841	0.005596	12	2	18	0.033259	0.014593
TRP	60.758	0.002592	6	3	17	0.004111	0.006361
TRP	173.505	0.047054	95	6	7	0.087389	0.086575
TRP	168.459	0.003142	7	6	8	0.087389	0.010869
TRP	167.219	0.005076	11	6	16	0.087389	0.010869
TRP	177.219	0.010208	21	7	7	0.021379	0.086575
TRP	176.383	0.003626	8	7	8	0.021379	0.010869
TRP	307.929	0.005234	11	3	8	0.154369	0.080732
TRP	308.656	0.007048	15	3	13	0.154369	0.012391
TRP	289.063	0.015617	32	3	16	0.154369	0.037174
TRP	292.197	0.019605	40	3	17	0.154369	0.041680
TRP	308.539	0.002719	6	4	7	0.119815	0.028913
TRP	306.689	0.021683	44	4	8	0.119815	0.080732
TRP	282.294	0.003218	7	4	9	0.119815	0.032668
TRP	291.324	0.008056	17	4	10	0.119815	0.014645
TRP	289.768	0.021553	44	4	11	0.119815	0.034921
TRP	288.683	0.009146	19	4	15	0.119815	0.028913
TRP	292.773	0.004710	10	4	16	0.119815	0.037174
TRP	290.268	0.003722	8	4	17	0.119815	0.041680
TRP	307.559	0.003739	8	5	8	0.040940	0.080732
TRP	280.338	0.005199	11	5	9	0.040940	0.032668
TRP	286.346	0.009146	19	5	15	0.040940	0.028913
TRP	290.752	0.002730	6	5	17	0.040940	0.041680
TRP	288.317	0.012606	26	6	7	0.066480	0.028913
TRP	295.880	0.015702	32	6	8	0.066480	0.080732
TRP	290.623	0.012625	26	6	9	0.066480	0.032668
TRP	288.851	0.006231	13	7	8	0.012395	0.080732
TYR	52.340	0.006785	24	1	17	0.045075	0.029450
TYR	52.924	0.007369	26	1	18	0.045075	0.033377
TYR	66.255	0.003032	11	2	17	0.013283	0.029450
TYR	57.222	0.004479	16	2	18	0.013283	0.033377
TYR	61.881	0.004764	17	3	17	0.021993	0.029450
TYR	58.560	0.009681	34	3	18	0.021993	0.033377
TYR	63.402	0.002454	9	4	17	0.003702	0.029450
TYR	50.104	0.010784	38	5	9	0.022429	0.018979
TYR	56.193	0.002638	10	5	10	0.022429	0.004581
TYR	74.469	0.001832	7	6	8	0.007186	0.006763
TYR	83.659	0.002113	8	7	8	0.005008	0.006763
TYR	185.718	0.002145	8	5	15	0.007188	0.012871
TYR	183.359	0.002174	8	5	16	0.007188	0.067408
TYR	180.720	0.003888	14	6	7	0.059897	0.022033
TYR	176.890	0.001520	6	6	8	0.059897	0.004145
TYR	184.880	0.004433	16	6	15	0.059897	0.012871
TYR	176.116	0.027398	95	6	16	0.059897	0.067408
TYR	167.226	0.001553	6	6	17	0.059897	0.007199
TYR	178.514	0.009937	35	7	7	0.038552	0.022033
TYR	172.318	0.012322	43	7	16	0.038552	0.067408
TYR	301.008	0.003927	14	3	10	0.213457	0.040794
TYR	299.300	0.013235	46	3	13	0.213457	0.022033
TYR	298.249	0.001891	7	3	14	0.213457	0.005017
TYR	298.388	0.004508	16	3	15	0.213457	0.028142
TYR	296.352	0.023415	81	3	16	0.213457	0.071336
TYR	300.233	0.020506	71	3	17	0.213457	0.047339
TYR	296.257	0.001600	6	3	18	0.213457	0.009381
TYR	284.393	0.001891	7	4	4	0.085251	0.005017
TYR	302.913	0.004218	15	4	9	0.085251	0.017234
TYR	300.434	0.017888	62	4	10	0.085251	0.040794
TYR	296.224	0.006835	24	4	15	0.085251	0.028142
TYR	287.864	0.012653	44	4	16	0.085251	0.071336
TYR	292.973	0.005090	18	4	17	0.085251	0.047339
TYR	293.617	0.002472	9	4	18	0.085251	0.009381
TYR	293.049	0.005381	19	5	9	0.047750	0.017234
TYR	282.793	0.004799	17	5	10	0.047750	0.040794
TYR	292.501	0.006254	22	5	15	0.047750	0.028142
TYR	284.735	0.008581	30	5	16	0.047750	0.071336
TYR	285.841	0.005381	19	6	7	0.025074	0.013744
TYR	283.418	0.003054	11	6	8	0.025074	0.009817
TYR	283.206	0.001891	7	6	16	0.025074	0.071336
TYR	287.397	0.003345	12	6	17	0.025074	0.047339

AName	avg Chi1	$P(r1 \mid \phi\psi))$	#	Phirot	Psirot	$P(\chi_1 \mid \phi)$	$P(chi_1 \mid \psi)$
TYR	281.382	0.002472	9	7	7	0.005887	0.013744
TYR	307.107	0.003345	12	13	11	0.005451	0.007635
VAL	56.701	0.001389	10	2	16	0.005633	0.010716
VAL	67.334	0.001974	14	2	17	0.005633	0.011158
VAL	44.777	0.001243	9	3	16	0.004750	0.010716
VAL	62.900	0.001389	10	3	17	0.004750	0.011158
VAL	53.625	0.003289	23	4	16	0.008726	0.010716
VAL	60.718	0.001974	14	4	17	0.008726	0.011158
VAL	57.004	0.000804	6	5	17	0.003645	0.011158
VAL	74.384	0.003864	27	6	8	0.012703	0.008507
VAL	79.010	0.002811	20	6	9	0.012703	0.004530
VAL	81.629	0.000804	6	6	17	0.012703	0.011158
VAL	58.273	0.000920	7	7	7	0.003424	0.002541
VAL	66.409	0.000948	7	7	8	0.003424	0.008507
VAL	179.450	0.000811	6	2	15	0.008305	0.080062
VAL	182.668	0.000810	6	2	17	0.008305	0.036875
VAL	175.323	0.000811	6	3	7	0.099995	0.066773
VAL	187.611	0.005621	39	3	13	0.099995	0.009634
VAL	182.825	0.002544	18	3	14	0.099995	0.006976
VAL	178.190	0.020132	137	3	15	0.099995	0.080062
VAL	177.116	0.023538	160	3	16	0.099995	0.143401
VAL	169.455	0.010675	73	3	17	0.099995	0.036875
VAL	176.171	0.001843	13	4	7	0.092465	0.066773
VAL	179.568	0.000800	6	4	14	0.092465	0.006976
VAL	180.223	0.014380	98	4	15	0.092465	0.080062
VAL	179.091	0.040214	273	4	16	0.092465	0.143401
VAL	172.414	0.002871	20	4	17	0.092465	0.036875
VAL	177.608	0.002433	17	5	7	0.059465	0.066773
VAL	176.912	0.000810	6	5	8	0.059465	0.033553
VAL	176.294	0.009955	68	5	15	0.059465	0.080062
VAL	178.066	0.018521	126	5	16	0.059465	0.143401
VAL	169.940	0.007289	50	5	17	0.059465	0.036875
VAL	172.429	0.025581	174	6	7	0.082055	0.066773
VAL	174.929	0.012733	87	6	8	0.082055	0.033553
VAL	177.920	0.004646	32	6	15	0.082055	0.080062
VAL	175.880	0.008043	55	6	16	0.082055	0.143401
VAL	173.669	0.002282	16	6	17	0.082055	0.036875
VAL	173.341	0.013344	91	7	7	0.043076	0.066773
VAL	180.512	0.008170	56	7	8	0.043076	0.033553
VAL	181.859	0.002728	19	7	15	0.043076	0.080062
VAL	178.499	0.003911	27	7	16	0.043076	0.143401
VAL	296.821	0.001395	10	2	18	0.099969	0.019579
VAL	295.605	0.000946	7	3	1	0.049043	0.002544
VAL	290.990	0.000745	6	3	12	0.049043	0.001217
VAL	304.003	0.005662	39	3	17	0.049043	0.026437
VAL	298.317	0.006242	43	3	18	0.049043	0.019579
VAL	293.899	0.001642	12	4	11	0.022474	0.003208
VAL	299.635	0.008457	58	4	17	0.022474	0.026437
VAL	297.555	0.002423	17	4	18	0.022474	0.019579
VAL	332.445	0.001517	11	5	16	0.010517	0.004978
VAL	302.678	0.001397	10	5	17	0.010517	0.026437
VAL	297.967	0.002423	17	5	18	0.010517	0.019579
VAL	301.016	0.004012	28	6	8	0.016053	0.008960
VAL	294.823	0.004740	33	6	9	0.016053	0.008738
VAL	306.278	0.001397	10	6	17	0.016053	0.026437
VAL	271.911	0.000802	6	7	8	0.002325	0.008960

Table D.5.: Backbone dependent probabilities for χ_1 rotamer combinations in unbound residues

Bibliography

[1] E. Althaus, O. Kohlbacher, H.-P. Lenhof, and P. Müller. A combinatorial approach to protein docking with flexible side-chains. In R. Shamir, S. Miyano, S. Istrail, P. Pevzner, and M. Waterman, editors, *Recomb 2000 – proceedings of the fourth annual international conference on computational molecular biology*, pages 15–24. ACM press, 2000.

[2] E. Althaus, O. Kohlbacher, H.-P. Lenhof, and P. Müller. A combinatorial approach to protein docking with flexible side-chains. *Journal of Computational Biology*, 9:597 – 612, August 2002.

[3] E. Althaus, O. Kohlbacher, H.-P. Lenhof, and P. Müller. A branch and cut algorithm for the optimal solution of the side-chain placement problem. Research Report MPI-I-2000-1-001, Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany, January 2000.

[4] H. M. Bermann, J. Westbrook, F. Zukang, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindylalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28:235–242, 2000.

[5] K. D. Berndt. Protein secondary structure. `http://www.cryst.bbk.ac.uk/PPS2/course/section8/ss-960531_5.html`. Link vom 21.05.2003.

[6] M. J. Betts and M. J. Sternberg. An analysis of conformational changes on protein-protein association: implications for predictive docking. *Protein Engineering*, 12(4):271–283, 1999.

[7] M. Bower, F. E. Cohen, and R. L. Dunbrack. Prediction of proteins side-chain rotamers from a backbone-dependend rotamer library: A new homology modelling tool. *Journal of Molecular Biology*, 267(5):1268–1282, 1997.

[8] H. Claussen, C. Buning, M. Rarey, and T. Lengauer. Flexe: Efficient molecular docking considering protein structure variations. *JOURNAL OF MOLECULAR BIOLOGY*, 2001, Vol. 308, T. 2, S. 377-395, 2001.

[9] M. L. Connolly. Solvent–accessible surfaces of proteins and nucleic acids. *Science*, 221:709, 1983.

[10] L. L. Conte, C. Chothia, and J. Janin. The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology*, 285:2177–2198, 1999.

[11] R. L. Dunbrack and M. J. Karplus. Backbone-dependent rotamer library for proteins. application to side-chain prediction. *Journal of Molecular Biology*, 230:543–574, 1993.

[12] F. Eisenhaber, P. Argos, and M. Scharf. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*, 16(3):273–284, 1995.

[13] G. A. Fink. Mustererkennung mit markov-modellen. State doctorate thesis, Bielefeld University, Technical Faculty, 2002.

[14] H. A. Gabb, R. M. Jackson, and M. J. E. Sternberg. Modelling protein docking using shape complementarity, electrostatics and biochemical information. 272:106–120, 1997.

[15] M. Gerstein, A. Lesk, and C. Chothia. Structural mechanisms for domain movements in proteins. *Biochemistry*, 33:6739–7649, 1994.

[16] I. Halperin, B. Ma, H. Wolfson, and R. Nussinov. Principles of docking: An overview of search algorithms and a guide to scoring functions. *PROTEINS -NEW YORK-*, 2002, Vol. 47, T. 4, S. 409-443, 2002.

[17] A. Heifetz, E. Katchalski-Katzir, and M. Eisenstein. Electrostatics in protein-protein docking. *PROTEIN SCIENCE*, 2002, Vol. 11, T. 3, S. 571-587, 2002.

[18] S. J. Hubbard, S. Campell, and J. M. Thornton. Molecular recognition conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *Journal of Molecular Biology*, 220:507–530, 1991.

[19] W. Humphrey, A. Dalke, and K. Schulten. Vmd - visual molecular dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[20] IUPAC-IUB Commission on Biochemical Nomenclature (CBN). Abbreviations and symbols for the description of the conformation of polypeptide chains. `http://www.chem.qmw.ac.uk/iupac/misc/noGreek/ppep1.html`, 1967. Link vom 11.12.2001.

[21] R. M. Jackson, H. A. Gabb, and M. J. E. Sternberg. Rapid refinement of protein interfaces incorporating solvation: Application to the docking problem. *Journal of Molecular Biology*, 276:265–285, 1998.

[22] J. Janin and S. Wodak. Conformation of amino acid side-chains in proteins. *Journal of Molecular Biology*, 125:357–386, 1978.

[23] W. Kabsch and C. Sander. *Biopolymers*, 22:2577–2637, 1983.

[24] T. Klein, F. Ackermann, and S. Posch. viwish: A visualisation server for protein modelling and docking. *Gene–COMBIS*, Gene 183:GC51–GC58, 1996.

[25] K. Koch, F. Zöllner, S. Neumann, F. Kummert, and G. Sagerer. Comparing bound and unbound protein structures using energy calculation and rotamer statistics. *In Silico Biology*, 2:32, 2002.

[26] O. Kohlbacher. *New approaches to protein docking.* PhD thesis, University Saarbrücken, 2000.

[27] Koshland. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA. 44:98 ?104.*, 44:98–104, 1958.

[28] A. R. Leach. Ligand docking to proteins with discrete side-chain flexibility. *Journal of Molecular Biology*, 235:345–356, 1994.

[29] B. Lee and F. Richards. The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology*, 55:379–400, 1971.

[30] S. C. Lovell, M. Word, J. S. Richardson, and D. C. Richardson. The penultimate rotamer library. *PROTEINS: Structure, Function, and Genetics*, 40:389–408, 2000.

[31] M. Mc Gregor, S. Islam, and M. J. E. Sternberg. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *Journal of Molecular Biology*, 198:295–310, 1987.

[32] J. Mendes, A. Baptista, M. A. Carronda, and C. Soares. Improved modeling of side-chains in proteins with rotamer-based methods: A flexible rotamer model. *PROTEINS: Structure, Function, and Genetics*, pages 530–543, 1999.

[33] R. Najmanovich, J. Kuttner, V. Sobolev, and M. Edelmann. Side-chain flexibility in proteins upon ligand binding. *PROTEINS: Structure, Function, and Genetics*, 39:261–268, 2000.

[34] S. Neumann, F. Zöllner, K. Koch, F. Kummert, and G. Sagerer. Elmar: A protein docking system using flexibility information. 2002. poster 113, ECCB 2002.

[35] R. Norrel, D. Petrey, H. J. Wolfson, and R. Nussinov. Examination of shape complementarity in docking of unbound proteins. *PROTEINS: Structure, Function, and Genetics*, 36(3):307–317, 1999.

[36] N. C. of the International Union of Biochemistry and M. B. (NC-IUBMB). Enzyme nomenclature. `http://www.chem.qmul.ac.uk/iubmb/enzyme/EC1/cont1aa.html`, 1992. Link vom 21.05.2003.

[37] C. Orengo, A. E. Todd, and J. M. Thornton. From protein structure to function. *Current Opinion in Structural Biology*, (9):374–382, 1999.

[38] C. A. Orengo, F. M. G. Pearl, and J. E. Bray. The cath database provides insights into protein structure/function relationships. *Nucleic Acids Research*, 27(1):275–, 1999.

[39] J. W. Ponder and F. M. Richards. Tertiary templates for proteins: Use of packing criteria in the enumeration of allowed sequences for different structural classes. *JMB*, 193:775–791, 1987.

[40] P. R. Pearlstein, C. FitzGerald, and NIH. Pdb–at–a–glance. `http://cmm.info.nih.gov/modeling/pdb_at_a_glance.html`, 1996. Link 11.12.2001.

[41] G. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, pages 95–99, 1963. Klassisches Paper, in dem Ramachandran Plots beschrieben werden.

[42] B. Sandak, H. J. Wolfson, and R. Nussinov. Flexible docking allowing induced fit in proteins: Insights from an open to closed conformational isomers. *PROTEINS: Structure, Function, and Genetics*, 32(2):159–, 1998.

[43] D. Schomburg. Brenda, the comprehensive enzyme information system. `http://www.brenda.uni-koeln.de`, 2003. Link 16.04.2003.

[44] H. Schrauber, F. Eisenhaber, and P. Argos. Rotamers: To be or not to be? an analysis of amino acid side-chain conformations in globular proteins. *Journal of Molecular Biology*, 230:592–612, 1993.

[45] L. Stryer. *Biochemistry*. W. H. Freeman, 1994.

[46] P. Tuffery, C. Etchebest, and S. Hazout. Prediction of protein side chain conformations: a study on the influence of backbone accuracy on conformation stability in the rotamer space. *Protein Engineering*, 10(4):361–372.

[47] P. v. Sengbusch. Aminosäuren. http://www.rrz.uni-hamburg.de/biologie/b_online/d16/16j.htm, 1999. Link vom 31.5.2001.

[48] S. Zhao, D. S. Goodsell, and A. J. Olson. Analysis of a data set of paired uncomplexed protein structures: New metrices for side-chain flexibility and model evaluation. *PROTEINS: Structure, Function, and Genetics*, pages 271–279, 2001.

[49] G. Zubay. *Biochemistry*. WCB, 1993.