

**Secbase:
A Novel Tool to Correlate
Secondary Structure Elements with Ligand Binding**

Dissertation
zur
Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem

Fachbereich
der Philipps-Universität Marburg
vorgelegt von

**Oliver Koch
aus Herford**

Marburg/Lahn 2008

Vom Fachbereich Pharmazie der Philipps-Universität Marburg
als Dissertation angenommen am:

Erstgutachter:

Prof. Dr. G. Klebe

Zweitgutachter:

Prof. Dr. G. Schneider

Tag der mündlichen Prüfung:

10. September 2008

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr. G. Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von April 2003 bis September 2007 durchgeführt.

We cannot learn from the hidden.

Tudor I. Oprea

Table of Contents

1	Introduction	1
1.1	Secondary structure elements and ligand binding.....	1
1.1.1	Secondary Structure Elements	1
1.1.2	Influence of secondary structure elements	3
1.1.3	Turns as functional elements.....	5
1.1.4	Functional similarity within related folding patterns	5
1.1.5	Secbase	6
1.2	Turn Classification	8
1.3	Publications arising from this work	11
2	Secbase: Methods and Materials	13
2.1	PDB, Relibase and Reliscript	13
2.1.1	The Protein Data Bank	13
2.1.2	PDB select list	14
2.1.3	Relibase +	14
2.1.4	Reliscript	15
2.2	Hydrogen bonding.....	17
2.3	Conformational Parameter	18
2.4	Relibase Substructure Search	20
2.5	Geometric Description	21
2.5.1	α -Helices and kinks	21
2.5.2	β -Strands and kinks	22
3	Secbase: Applications	24
3.1	Conformational parameter.....	24
3.2	Ligand to backbone amide interactions.....	29
3.2.1	Methods and Materials	29
3.2.2	Water – Backbone Amide	31
3.2.3	Water mediated ligand binding to a backbone amide	38
3.2.4	Ligand – Backbone Amide.....	40
3.2.5	Discussion & Conclusion	44
4	Secbase: Summary	46
5	Cooperative Effects in Hydrogen Bonding	47

5.1	Introduction	47
5.2	Methods	50
5.3	Results	51
5.3.1	Helical structures	51
5.3.2	β sheet structures	54
5.4	Discussion and Conclusions	56
6	Turn Classification: Existing turn families and types	60
6.1	δ -turn	60
6.2	ϵ -turn	60
6.3	γ -turn	60
6.4	β -turn	62
6.5	α -turn	67
6.6	π -turns	70
6.7	Additional information	71
6.7.1	Deviations from standard angles	71
6.7.2	Intermolecular hydrogen bonding	71
6.7.3	Nomenclature based on position in Ramachandran plot	71
7	Turn Classification: Methods and Materials	73
7.1	Emergent Self-Organizing Maps (ESOM) and U-matrix	73
7.2	Turn Data Collection	77
7.2.1	$C\alpha$ - $C\alpha$ distance cut-off	77
7.2.2	Final data sets	78
7.3	Data Preparation	79
7.3.1	ω transformation	79
7.3.2	Z transformation	80
7.4	Cluster assignment and evaluation	81
7.5	Conformation of amino acids in protein	84
7.5.1	Ramachandran plots	86
8	Turn Classification: Results and discussion	90
8.1	Introduction	90
8.1.1	Definitions	90
8.1.2	Description of an example ESOM map	91
8.1.3	Torsion angle analysis	92

8.2	Two residue turns	93
8.2.1	δ -turn	93
8.3	Three residues turns	96
8.3.1	γ -turns.....	96
8.3.2	ε -turns.....	99
8.4	Four residue turns.....	102
8.4.1	Normal β -turn.....	102
8.4.2	Open β -turns.....	106
8.4.3	Reverse β -turns.....	111
8.5	Five residue turns	115
8.5.1	Normal α -turns	115
8.5.2	Open α -turns.....	119
8.5.3	Reverse α -turns.....	125
8.6	Six residue turns	129
8.6.1	Normal π -turns	129
8.6.2	Open π -turns.....	133
8.6.3	Reverse π -turns.....	138
8.7	Discussion	142
8.8	Differences to known classification.....	145
8.8.1	Benefit of clustering using ESOMs.....	145
8.8.2	Increase of $C\alpha$ - $C\alpha$ distance for open turns.....	147
8.8.3	Differentiation between open and normal turns.....	148
8.8.4	Introduce of reverse turns.....	149
8.8.5	The peptide bond torsion angle	150
8.8.6	Summary of turn-types.....	151
8.9	Further results and discussion	152
8.9.1	Principal Component Analysis.....	152
8.9.2	Prediction of turns in protein structure.....	156
9	Turn Classification: Summary and Outlook.....	157
10	Literature	159
11	Summary, Zusammenfassung	159
11.1	Summary	159
11.2	Zusammenfassung.....	161

12	Acknowledgements	169
13	Appendix	170
13.1	two residue turns	170
13.1.1	δ -turns.....	170
13.2	three residue turns	176
13.2.1	γ -turns.....	176
13.2.2	ε -turns.....	178
13.3	four residue turns.....	181
13.3.1	normal β -turns	181
13.3.2	open β -turns.....	185
13.3.3	reverse β -turns	195
13.4	five residue turns	205
13.4.1	normal α -turns	205
13.4.2	open α -turns.....	210
13.4.3	reverse α -turns	222
13.5	six residue turns.....	234
13.5.1	normal π -turns	234
13.5.2	open π -turns.....	239
13.5.3	reverse π -turns	254

1 Introduction

1.1 Secondary structure elements and ligand binding

1.1.1 Secondary Structure Elements

There are three types of secondary structure elements that build up the rigid framework of a protein (Branden 1999). Helices and β -sheets are responsible for neutralizing the highly polar main chain within the hydrophobic interior of the protein. They exhibit clearly defined torsion angles and for this reason they are called regular secondary structure elements. Based on this regularity, they were already proposed by Pauling and Corey in the year 1951 (Pauling 1951; Pauling 1951; Eisenberg 2003). In contrast, the irregular turns cover a wide range of possible torsion angle and give the protein chain the opportunity to fold back upon itself (see section 1.2).

β -sheets are built up from adjacent β -strands that are connected through hydrogen bonds between the main chain functional groups. β -strands are extended parts of the polypeptide chain. They normally have a length between five and ten amino acids. There are different type of β -sheets whose classification is based on the spatial direction of the adjacent β -strands and the hydrogen bonding pattern. The so-called parallel β -sheets (Figure 1.1 a)) contains β -strands with the same spatial direction and an evenly spaced hydrogen bonding pattern between one residue on one β -strand and two residues on the adjacent β -strand. Within an antiparallel β -sheet (Figure 1.1 b)) β -strands show opposite orientation and parallel hydrogen bonding pattern between one residue on neighbouring strands. This leads to alternating pairs of parallel narrowly spaced and widely spaced hydrogen bonds.

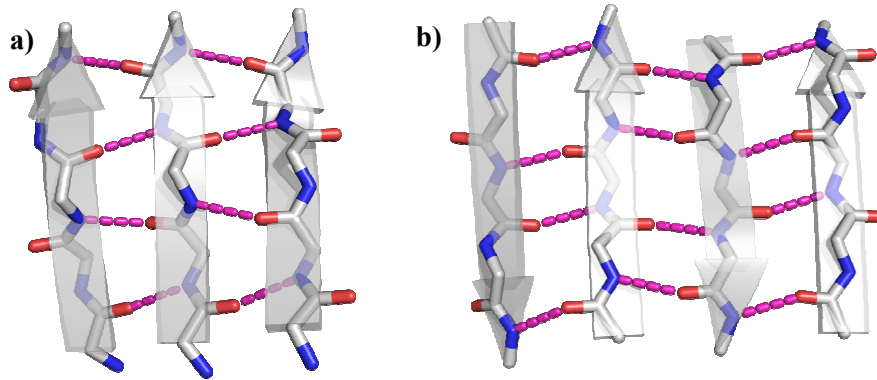


Figure 1.1: parallel (a) and antiparallel (b) β -sheet with main chain hydrogen bonds (magenta)

The helices are mainly stabilized by intrahelical hydrogen bonds between the main chain CO-group of residue i and the main-chain NH-group of residue $i+n$. This leads to three type of helices: The α -helix ($n = 4$, Figure 1.1 a)), the π -helix ($n = 5$) and the 3_{10} -helix ($n=3$, Figure 1.1 b)) and each of them is described by specific properties (the repeat unit and the backbone torsion angles, see Table 1.1). The side chains, which are attached to the C_{α}

helix	φ	ψ	repeat unit
3_{10}	-49°	-26°	3.010
α	-57°	-47°	3.613
π	-57°	-70°	4.416

Table 1.1: properties of helices

atom, are pointing away from the helix axis. The helix termini exhibit amide groups at the helix N-terminus and carbonyl groups at the C-terminus respectively that can be involved in hydrogen bonding. The helix boundary residues (Figure 1.2) are called N_{cap} and C_{cap} . In more detail, N_{cap} position is the last residue with a non-helical backbone conformation. It could be defined as to be the first hydrogen bonded position (Penel 1999) or without having a hydrogen bond (Engel 2004), but the last non-helical conformation is essential. The other way round, C_{cap} is the first non-helical position at the helix C-terminus, having a hydrogen-bond or not. The following and preceding residues are called: N_{one} , N_{two} and N_{three} at N-terminus of a α -helix and C_{three} , C_{two} and C_{one} at C-terminus of a α -helix (Figure 1.2 c)). This leads to four amide groups at the N-terminus and four carbonyl groups at the C-terminus not involved in intrahelical hydrogen-bonding.

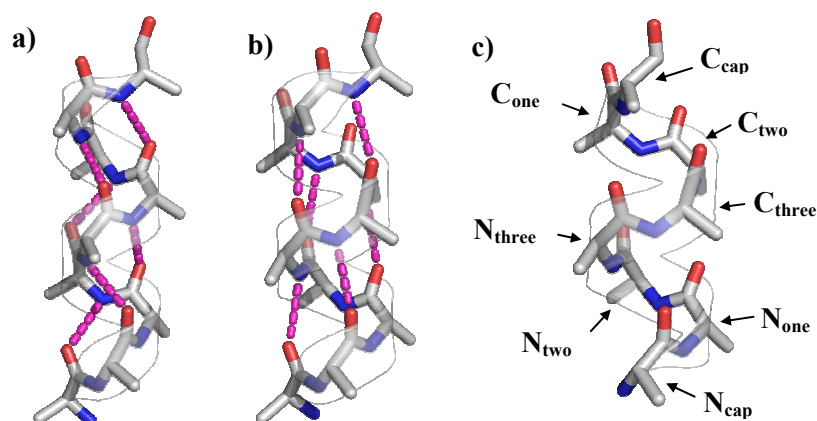


Figure 1.2: a) 3_{10} -helix and b) α -helix with main chain hydrogen bonding (magenta), c) different positions at helix terminus

Turns are irregular secondary structures with a hydrogen bond or a specific $C\alpha$ -distance between the first and the last residue that is involved in building a turn. For further details about turns and the new uniform classification see section 1.2.

1.1.2 Influence of secondary structure elements

In general, the function of secondary structure elements is described in building the rigid framework of a protein or presenting sidechains in the right conformation for ligand binding. However, they could also take influence on the physicochemical properties of functional groups neighbored to the secondary structure elements or within themselves. One example are cooperative effects in hydrogen bonding that lead to an decrease of the mean hydrogen bond length in α -helices and parallel β -sheets with an increasing length (see chapter 5). Another example is the effect of α -helices on the pK_a values of an amino acid at the N-terminus. Presumably, this could also have dramatic effects on the protein-ligand interactions and therefore secondary structure elements should clearly kept in mind during development of new ligands.

pK_a shift of Cys at the α -helix N-terminus

A mutagenesis study (Miranda 2003) of cysteine at different position at the N-terminus of an α -helix in combination with a experimental determination of the thiol pK_a values clearly showed a pK_a shift at the N_{cap} position (Table 1.2) in contrast to the pK_a -values at the other positions or a control cysteine not involved in helices.

Cys position	thiol- pK_a
control peptide	8.62 ± 0.04
N_{two}	8.12 ± 0.05
N_{one}	8.43 ± 0.03
N_{cap}	6.53 ± 0.05

Table 1.2: Thiol- pK_a at different positions at α -helix N-terminus

In contrast to cysteine at the N_{one} and N_{two} position, the cysteine at the N_{cap} position has the conformational space to place the sulphur close to the helix axis (Figure 1.3). Hydrogen bonds between the backbone amides and the cysteine thiolate are possible. Additionally, favourable electrostatic interaction to a charged thiolate is conceivable. Miranda (Miranda 2003) suggests that both, the possible hydrogen bond formation and the overall helix dipole, will contribute to this large pK_a shift at the N_{cap} position. Therefore, the complete secondary structure element has to be considered and not only local interactions.

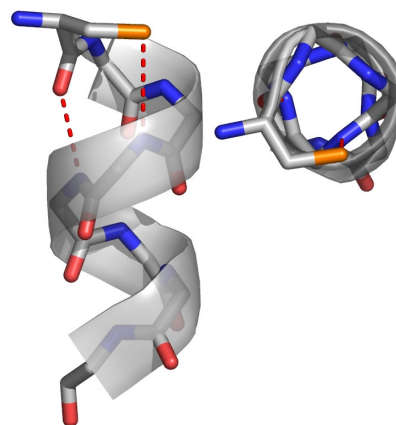


Figure 1.3 Cys at N_{Cap} position of a α -helix with hydrogen bonds between the carbonyl group and the thiol to the helix backbone

In nature, a charged thiolate is essential in human thioredoxin since a nucleophilic attack by this charged thiolate is necessary. The catalytic cysteine involved in this mechanism is located at the N_{cap} position of a α -helix and shows a pK_a of 6.3 (Miranda 2003).

Carboxyl pK_a values at the α -helix N-terminus

An analysis of over 200 carboxyl pK_a values of aspartates and glutamates in proteins (Forsyth 2002) revealed lower mean pK_a values at the N-termini of helices compared to the overall mean pK_a values within proteins and model peptides.

	mean pK_a values	
	aspartate	glutamate
model peptide	3.8 – 4.1	4.1 – 4.6
protein	$\leq 3.4 (\pm 1.0)$	4.1 (± 0.8)
helix N-terminus	2.8 (± 0.5)	3.4 (± 0.6)

Table 1.3: Mean pK_a values of aspartate and glutamate

Additionally, they described a relation between the numbers of hydrogen bonds to the carboxylate and the mean pK_a values for aspartate; the mean pK_a decreases with an increase in the number of hydrogen bonds to the carboxyl group.

Subsequently Porter et. al (Porter 2006) explained the lower pK_a values at the helical N-terminus with a greater number of hydrogen bond interactions to backbone amides and sidechains compared to the rest of the protein. Statistically, aspartate shows 1.1 backbone hydrogen bonds among the overall distribution and 1.6 at the N-terminus. Glutamate shows 0.4 and 0.6 respectively, although glutamate is more likely stabilized by sidechain hydrogen bond interactions (0.3 more sidechain hydrogen bonds and 0.2 more backbone hydrogen bonds). Furthermore, they demonstrated the lack of a correlation between the different pK_a

values at the helical N-terminus and an increasing magnitude of the helix macrodipole, although a correlation could be expected. They finally concluded that a backbone and a sidechain hydrogen bond can contribute up to 1.2 and 0.8 pH units to the pK_a , but could not completely eliminate any influence of the helix macrodipole.

1.1.3 Turns as functional elements

Turns are also important from a functional point of view (see chapter 1.2 and 6). They can be involved in ligand binding accounting for presenting sidechains in the correct conformation. For example, the aspartate of the catalytic triad in serine proteases like trypsinogen and chymotrypsin is part of a strong γ -turn. They also participate in molecular recognition processes between proteins or between a peptide substrate that mimics turn conformation of a protein. Therefore, the analysis of turns could provide information for designing new ligands as turn mimetics.

1.1.4 Functional similarity within related folding patterns

The number of predicted possible protein folds varied over the last years between ~650 and ~10000 (Koonin 2002). However, the majority of protein folds presumably belong to ~1000 common folds, because an increasing number of proteins consist only of a small number of protein folds (Koonin 2002).

Since protein families with similar folding pattern and completely dissimilar sequences could have evolved from the same ancestor and although they are involved in different biochemical reactions they could still bind similar ligands (Breinbauer 2002) or show similar biochemical activities involved in different biological functions (Koonin 2002). Thus, identifying proteins with similar folding pattern could be used for recognizing unexpected cross-reactivity and discovering new lead structures.

Koch et al. have suggested the idea of ligand-sensing cores (Koch 2005), which comprise the catalytic site of a protein from the structural point of view. They propagate the concept that some structural fold characteristics could presumably be identified in various proteins, which could also bind similar ligands. Therefore, ligand-sensing cores are grouped into so-called protein structure similarity clusters (PSSC). A ligand of one member of these clusters could be used as a lead or starting point for the development of new ligands for the remaining part of this cluster. Among other examples, which were analyzed afterwards, they describe an example that was successful in the face of grouping ligand sensing cores into protein structure

similarity clusters. Cdc25a phosphatase, acetylcholinesterase (AChE) and 11 β -hydroxysteroid dehydrogenase type 1 and 2 (11 β HSD1 and 11 β HSD2) show structural similarities within their catalytic cores. Therefore, a compound collection was synthesized based on an inhibitor skeleton known to bind at Cdc25A. Interestingly enough testing the same compound library against the other two proteins revealed a hit rate of 2-3% (Koch 2005).

1.1.5 Secbase

Relibase is an object-oriented data management system (Hendlich 2003) that stores structures of three dimensional protein-ligand complexes deposited in the PDB. Secbase is a modular extension of Relibase that integrates information about secondary structural elements assigned to each individual protein structure in Relibase. In addition to information about helices and β -sheets retrieved from the PDB, a new uniform classification of all turn families is integrated into Secbase (see section 1.2). This new classification is based on recent clustering methods and highly accurate sequence-based protein turn prediction has confirmed this new categorization as consistent and well-defined.

The function of a protein is often based on the structure and the spatial structure is more conserved in evolution than amino acid sequence. Furthermore, local conformational similarity in conserved motifs can occur in globally different structures (Grishin 2001). Therefore, the motivation for Secbase is guided through two main ideas: Firstly, Secbase should provide means to analyse protein-ligand interactions with respect to secondary structure elements and, secondly, should allow analysis and discovery of functional similarity within related folding patterns. In general, Secbase, in combination with Relibase, can be used for knowledge discovery about the influence of secondary structural elements on protein-ligand interactions and should be valuable for structure based drug design and molecular modelling.

At first, the python-based interface to Relibase (Reliscript) was used to add the information about helices and β -sheets from the PDB in terms of python objects (see chapter 2.1.3 and 2.1.4). The created turn classification (see section 1.2) was used to classify the turn conformations within all protein structures. This information was also added. Furthermore, the C++ core code of Relibase was extended to get access to this data and to add Secbase constraints to the substructure search. This leads to the opportunity to specify Secbase constraints within the substructure search accessible through the Relibase webinterface (see chapter 2.4). Three analyses were performed using Reliscript and the webinterface substructure search. These are the analysis of the amino acid distribution within helices (see

chapter 3.1), an analysis of hydrogen bonding within helices and β -sheets (see chapter 5) and the ligand to backbone amide interactions with respect to secondary structure elements (see chapter 3.2)

1.2 Turn Classification

In contrast to helices and β -sheets, turns are irregular secondary structure elements. Initial definitions describe non-helical turns as sites where the polypeptide chain folds back on itself (Venkatachalam 1968). Turns are thus deemed responsible for the globularity of proteins. A more obvious and widely accepted definition (Chou 2000) is based on an intra-turn hydrogen bond (Table 1.4) between the main-chain carbonyl group from the first residue of the turn and the main-chain amide group from the last residue of the turn. This leads to four different turn families (γ , β , α and π -turns) which are differentiated by the number of residues separating the pair connected via the hydrogen bond. Two further families exist, δ and ε -turns. They are characterized by a hydrogen bond between the main-chain amide group of the first residue and the main-chain carbonyl group of the last residue (Toniolo 1980). In these conformations a ‘reverse’ hydrogen bond is observed. In addition, there are also ‘open’ conformations described for β - and α -turns, where the distance between the first and last $C\alpha$ must fall into a specific range (Lewis 1973; Dasgupta 2004).

turn family	nr of residues	H bond type
δ	2	$NH_i - CO_{i+1}$
ε	3	$NH_i - CO_{i+2}$
γ	3	$CO_i - NH_{i+2}$
β	4	$CO_i - NH_{i+3}$
α	5	$CO_i - NH_{i+4}$
π	6	$CO_i - NH_{i+5}$

Table 1.4: Number of residues and hydrogen bond pattern of different turn types

Finally, the backbone dihedral angles of the inner residues (the torsion angles between the hydrogen bond) define different subfamilies or turn-types within a turn family. Figure 1.4 shows a normal β -turn and the torsion angles that are used to describe the specific conformation.

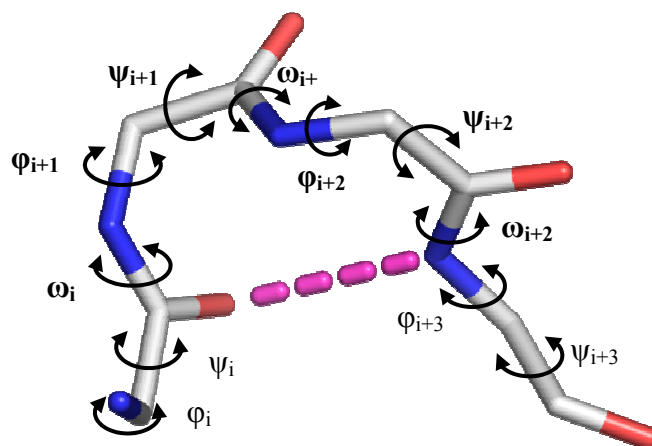


Figure 1.4: Normal β -turn and torsion angles that describe the individual turn conformation

Turn structures are not only important structural elements responsible for folding the protein chain back on itself. They also have an influence on function, when they occur at the surface of a protein. They can be involved in molecular recognition processes between different proteins, or between a protein and a peptide substrate that imitates turn conformation (a so-called turn-mimetic). For example, the large group of GPCRs presumably recognize peptidic ligands that adopt turn-type conformations within the bioactive form (Tyndall 2005). The turn conformation could be considered as a kind of scaffold for presenting a sidechain in the correct orientation. Therefore, turns provide information required for the design of new ligands as turn mimetics (Chou 2000; Tyndall 2005). Additionally, turns are also involved in ligand binding; they could be responsible for presenting the correct sidechain orientation within a binding site. For example, an Asp of the binding site in serine proteases such as trypsinogen and chymotrypsin corresponds to residue $i+1$ in a strong γ -turn (Milner-White 1990), at the same time this Asp is a central residue in the catalytic triad in these proteins.

The classification of turns based on main-chain conformations has been modified and extended over the last decades for different turn families due to an increasing amount of data from protein structures, preventing uniform classification. Additionally, turns are not yet completely integrated into current databases; examining the Protein Data Bank (www.pdb.org) reveals that although there is an entry for turns, on average only one turn per protein chain is described, which appears as not sufficient. In the present analysis there are more than 140000 hydrogen bonded turns identified within a dataset of 1903 protein chains (on average: ~ 73 turns per chain). Also, when a turn is described, the information about the assigned classification is usually not given. There are methods available to detect turns in proteins, e.g. the program PROMOTIF (Hutchinson 1996) or the PROSS-Server (<http://roselab.jhu.edu>) (Srinivasan 1999), but they do not cover all possible turn families.

Furthermore, some turn types are only theoretically described (δ , ϵ -turns) and the question about the correct distance cut-off between the $C\alpha_i$ and $C\alpha_{i+n}$ for open/distorted turns is not yet satisfactorily answered (Chou 2000).

Correct turn classification is important for identifying structural similarity among different proteins and for structural prediction based on amino-acid sequence. Over the last decade it turned out (see section 6) that due to the irregularity of turn structures, classification into turn types is a fundamental problem. To obtain a uniform classification for all turn families that is based on the current data, the analysis of all turn families was started from scratch, including ω torsion angles. So far, the ω torsion angle was not considered generally in any previous analysis.

At first, data for all turn structures were collected within a non-redundant dataset based on the PDB-select list (Hobohm 1994). Reliscript (a Python based interface of Relibase+ (Hendlich 2003)) was used to identify all possible turn structures, based on the hydrogen bonding between the main chain amide and carbonyl groups and the distance between the $C\alpha$ atoms of the first and the last residue, respectively. To identify all similar turn types within all families, the backbone torsion angles ϕ , ψ and ω were used to cluster the turn structures with Emergent Self-Organising Maps (ESOMs) (Ultsch 2003). This leads to a completely new, uniform turn classification for all known turn families based on machine learning methods, considering current data and including the ω torsion angle (see chapter 6, 7, 1 and 6).

1.3 Publications arising from this work

This manuscript is mainly based on published or accepted articles or articles in preparation, whereas some articles are contained in an extended form with additional information. The content of the original publication are integrated and slightly adapted to fit into this thesis and the formatting.

The following article is based on Section 1.1, 2, 1 and 4:

Koch, O., Cole, J., Block, P., Klebe, G.

Secbase – Secondary Structure Elements and Ligand Binding

manuscript ready for submission

Section 5 represents the following article with only slightly adaptations:

Koch, O., Bocola, M., Klebe, G.

Cooperative effects in hydrogen-bonding of protein secondary structure elements: a systematic analysis of crystal data using Secbase

Proteins, 2005, 61, 310-317

The following article represents an extraction of section 1.2, 7, 1 and 6. Section 6 contains additional background information:

Koch, O., Klebe, G.

Turns revisited: A uniform and comprehensive classification of normal, open and reverse turn families minimizing unassigned random chain portions.

Proteins, accepted

As described in section 8.9, the data compiled in this study were already confirmed as consistent and well-defined:

Meissner, M. Koch, O. Klebe, G., Schneider, G.,

Prediction of Turns Types in Protein Structure by Machine-Learning Classifiers

Proteins, accepted

Furthermore, the following poster contributions and talks are based on this work:

Talks

Secbase - Ein Werkzeug zur Korrelation von sekundären Strukturelementen mit der Ligandenbindung, DPhG-Doktorandentagung, Freudenstadt-Lauterbad, Germany, (03/2004)

Secbase - Information about Secondary Structure Elements for Molecular Modeling, Molecular Modelling Workshop, Erlangen, Germany, (05/2006)

Secbase and DrugscoreMaps - useful tools for structure-based drug design, Center for Bioinformatics, University of Hamburg, Hamburg, Germany (04/2008)

Turns revisited: Clustering turn structures using ESOMs leads to a uniform classification for open, normal and reverse turn families, The 8th International Conference on Chemical Structures, Noordwijkerhout, Netherlands, (06/2008)

Poster

Koch, O.; Block, P.; Bocola, M.; Klebe, G.: *Secbase - A tool for correlating secondary structure elements with ligand binding*. 18. CIC-Workshop der GDCh-Fachgruppe Chemie-Information-Computer, Boppard (Rhein), Deutschland, (11/2004)

Koch, O.; Klebe, G.: *Secbase - Secondary structure elements and ligand binding*. Molecular Modelling Workshop, Erlangen, Deutschland, (05/2005)

Koch, O., Klebe, G.: *Using Secbase to analyse secondary structure elements: Involvement of turns in ligand binding*. German Conference on Bioinformatics, Hamburg, (10/2005)

Koch, O., Klebe, G.: *Are turn clusters of functional relevance? Using Secbase to analyse secondary structure elements*. 1. German Conference on Chemoinformatics, Goslar, (11/2005)

Meissner, M.; Koch, O.; Klebe, G.; Schneider, G.: *Prediction of novel turn categories in proteins using adaptive kernel methods*, 2. German Conference on Chemoinformatics, Goslar, (11/2006)

Koch, O.; Klebe, G.: *Secbase - Finding related proteins using a 2D topology search*. 2. German Conference on Chemoinformatics, Goslar, (11/2006)

Koch, O.; Cole, J.; Klebe, G.: *Secbase - Secondary Structure Elements and Ligand Binding: An Analysis using Secbase*, 3. German Conference on Chemoinformatics, Goslar, (11/2007)

Abstract published in meeting abstracts for 3rd German Conference on Chemoinformatics:

Koch, O., Cole, J., Klebe, G.: *Secbase - Secondary Structure Elements and Ligand Binding*, Poster presentation - Chemistry Central Journal 2008, 2(Suppl 1): P21

2 Secbase: Methods and Materials

The information about the helices and β -sheets are retrieved from the corresponding pdb-files and assigned to the existing Relibase data. All possible turn conformations were detected using Python programs and classified according to the procedure described in section 7. The hydrogen bonds of the normal and reverse turn families were identified using an implementation of the DSSP energy-function from Kabsch and Sander (Kabsch 1983) which calculates the electrostatic energy between the two main chain hydrogen bonding groups. This method is similar to the automatic assignment of secondary structure elements of the PDB. Furthermore, open turns are recognized based on a $C\alpha_i-C\alpha_{i+n}$ distance cutoff of 10\AA .

2.1 PDB, Relibase and Reliscript

2.1.1 The Protein Data Bank

The Protein Data Base is a project of the Research Collaboratory for Structural Bioinformatics (RCSB¹) and contains the structural information of biological macromolecules including proteins and nucleic acids. It was founded by the “Brookhaven National Laboratory” in 1971 and later (1998) taken over by the RCSB, whereas three institutions² participate in developing the PDB. As of July 2007 the PDB contains 44476 structure of these, 37853 are X-ray and 6386 NMR structures. Since the PDB is the single worldwide depository, it is the only publicly available source for structural information.

In fact, the PDB is not a real databank; it is more a package of text files with structural information about the macromolecules. These files contain lists of atomic coordinates with some additional information like the secondary structure elements or the biological source and it will be often compared with a loosely sheet collection. Therefore, it is difficult and inefficient to search for similar properties and mainly, structural similarities are really cumbersome to identify within the whole PDB.

¹ <http://www.rcsb.org/pdb/>

² Rutgers (University of New Jersey); San Diego Supercomputer Center (SDSC, University of California); Center for Advanced Research in Biotechnology (CARB, National Institute of Standards and Technology)

Additionally, the databank grows over the years with changes of the file format and other arrangement which leads to an inconsistent data collection.

2.1.2 PDB select list

The PDB select list is a project from Uwe Hobohm und Chris Sander (Hobohm 1994) with the aim to make a list of PDB entries available that is representative for the whole PDB.

Similar protein chains with a high sequence similarity can occur within the PDB. For example, the same protein can be solved with different ligands or with a higher resolution or as a mutant. Additionally, protein families have members with a high sequence similarity. This leads to a redundancy of the whole PDB with different weighting of the different proteins and protein families. Therefore, it is problematic to use the whole PDB for an analysis because some of the protein structures or protein families are over represented and others are under represented. Thus, using the entire pdb would obtain biased results. The PDB select list (Hobohm 1994) contains only protein chains with a sequence homology smaller than 25% and one can assume that the list consists of unique protein chains.

Before creating the PDB select list, pdb files were excluded that meet the following requirements:

- smaller than 30 residues or more than 5% of non-standard amino acid residues
- resolution bigger than 3.5 Å or a R-factor bigger than 30%

In case of similar protein chains, the chain was chosen based on the following criteria:

- at least 90% of residues must have sidechain info
- at least 90% of main chain coordinates must be present
- newer structures were preferred

2.1.3 Relibase +

The basic principles of knowledge-based approaches are to search for and gain information and correlations in order to find new insights and principles that could not be elucidated with other methods. Using the PDB to analyse the structural information about protein-ligand complexes is inefficient and very difficult (see 2.1.1 for reasons) and to solve this problem and to make the information easily accessible, Relibase (Hendlich 2003) (*Receptor-Ligand-Database*) was developed in 1994 in consideration of its use for structure-based drug design and understanding protein-ligand interactions.

Relibase is an object-orientated data management system that stores the three-dimensional structural information of protein-ligand complexes deposited in the PDB. The core contains the information as pre-processed data that is accessible via a web-based user interface which allows an access of the data in a fast and efficient way.

2.1.4 Reliscript

Reliscript is a Python based command-line interface to Relibase that can be used to access the whole functionality of Relibase. It allows more complex queries to be constructed using Python and the Relibase search functions. Python³ is an interpretable and object-oriented programming language which was developed by Guido van Rossum. It is easy to learn and to use and enjoys great popularity among scientists. Since Python is an object-oriented language, Reliscript itself is also object-oriented and the available data are expressed in terms of objects. There are data objects, container objects and operation objects available.

In more detail, the data object gives access to the data stored in Relibase. The PDB object is the root object, it contains all the other data objects (e.g. the CHAIN object that includes the information about the individual protein chains) and the information about a complete PDB entry, e.g. author name or resolution. The container objects stores collections of data objects, for example the results of text searches that were performed using an operation object. Additionally, Reliscript searches can be stored as Relibase+ hitlists for further analysis using the Relibase+ web interface and vice versa. Figure 2.1 shows an example script that performs a text search and creates a Relibase+ hitlist. For further information see the Reliscript User Guide⁴.

```
[1] >>> import reliscript
[2] >>> pdbset = reliscript.set('pdb')
[3] >>> sersearch = reliscript.text_search('SERINE', field='header')
[4] >>> sersearch(pdbset)
[5] >>> pdbset.save_to_hitlist('serine search')
```

Figure 2.1: Simple Reliscript example that search for all pdb entries containing ,SERINE' in the header and stores the results as a Relibase+ hitlist

- 1) imports the Reliscript module that gives access to Relibase*
- 2) creates a container object with all available pdb objects*
- 3) creates a operation object that search for 'SERINE' in the pdb header*
- 4) performs a search; the results are stored in the 'pdbset' container*
- 5) stores the results as a Relibase+ hitlist*

³ www.python.org

⁴ <http://www.ccdc.cam.ac.uk/>

The toolkit permits a lower access to the Relibase+ data and gives the opportunity to extend the Relibase data. This toolkit was used to add the information about secondary structure elements and makes the additional data accessible in Reliscript.

The information about the secondary structure elements are processed and stored in terms of Python objects. The main object type is the SSE object that stores the basic information and the other secondary structure element objects (see Figure 2.2). The base objects are STRAND, HELIX and TURN and each object contains the information about an individual element. Additionally, STRAND and HELIX objects also have the described geometric description (see section 2.5) together with information about kinks. Furthermore, STRAND objects that belong to one β -sheet are encapsulated in SHEET objects (“.strands”) and finally, HELIX, TURN and SHEET objects that belong to one protein chain are encapsulated in a SSE object (“.helices”, “.sheets”, “.turns”).

To get access to the secondary structure elements of a protein the corresponding PDB object contained a new attribute (‘.sse’) that contains a list of SSE objects.

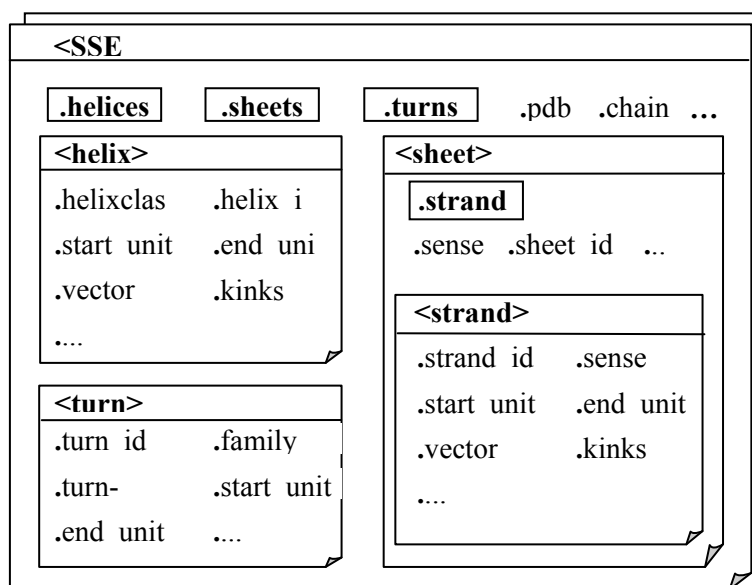


Figure 2.2: Schematic representation of the data about secondary structure elements

2.2 Hydrogen bonding

To identify hydrogen bonds the electrostatic interaction energy between two hydrogen bonded groups was calculated based on the equation implemented into the DSSP algorithm of Kabsch and Sander (Kabsch 1983).

$$E = q_1 * q_2 \left(\frac{1}{r(ON)} + \frac{1}{r(CH)} - \frac{1}{r(OH)} - \frac{1}{r(CN)} \right) * f$$

The same energy function is used for assigning secondary structure elements in the PDB files, where the electron unit charges are $q_1 = 0.42e$ and $q_2 = 0.20e$ for C,O (+ q_1 , - q_1) and N,H (+ q_2 , - q_2). $r(AB)$ is the interatomic distance from A to B in Ångstroms (Figure 2.3) and f the dimensional factor with $f = 332$. The resulting energy E is in kcal/mol with an ordinary H bond having a binding energy of -3 kcal/mol. A cut-off of $E < -0.5$ kcal/mol is used to identify hydrogen bonds as described by Kabsch and Sander.

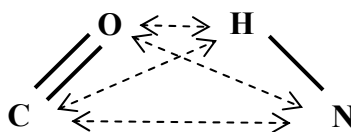


Figure 2.3: Atomic distances that are taken into account in DSSP energy function

Since hydrogen atoms are lacking in most of the pdb files the main-chain amide hydrogen has to be added manually. In reference to McDonald (McDonald 1994) the hydrogen atom was placed in the plane with the $C\alpha$ atom and the carbonyl carbon atom, but 4° closer to the $C\alpha$ atom with a 1.0 \AA distance to the nitrogen atom.

2.3 Conformational Parameter

The knowledge about different amino acid propensities throughout the helices, β -sheet and random coiled is available since decades (Chou 1974) and was adjusted several times to the growing number of structurally characterized proteins.

The conformational parameter P (Chou 1977; Wilmot 1988) describes the amino acid propensity within the different secondary structure elements and is built up from the frequency (f) and the average frequency ($\langle f \rangle$):

$$P_i = \frac{f_i}{\langle f_i \rangle}$$

The frequency of a specific residue is defined as

$$f_{sse} = \frac{n_{i,sse}}{n_{i,p}} = \frac{\text{number of residue } i \text{ in sse}}{\text{total number of residue } i \text{ in proteins}},$$

and the average frequency of finding all amino acid residues is defined as

$$\langle f_{sse} \rangle = \frac{n_{tot,sse}}{n_{tot,p}} = \frac{\text{total number of all residues in sse}}{\text{total number of all residues in proteins}},$$

With the help of Reliscript and a small Python script it is straight-forward to calculate this confirmative parameter P. The following example script counts the amino acid types within β -strands, which is needed for the calculation:

```
[1] import reliscript
[2] pdb = reliscript.create('1ets')
[3] strand = {}
[4] for sse in pdb.sse:
[5]   for sht in sse.sheets:
[6]     for str in sht.strands:
[7]       for res in str.residues:
[8]         strand[res.name] = strand.get(res.name, 0) + 1
```

Figure 2.4: python script that use reliscript for calculating the amino acid distribution in β -strands

The statistical significance of the frequency can be tested using a d-test (Wilmot 1988) respectively a z-value (Karpen 1992) - both equations produce the same results.

$$d = \frac{(n_{i,j,k}) - (np)}{\sqrt{(npq)}} \quad \text{with} \quad p = \frac{n_{i,p}}{n_{tot,p}} = \frac{\text{total number of residue of type } i \text{ in proteins}}{\text{total number of all residues in proteins}},$$

$$\text{and} \quad q = 1 - p$$

$$\text{and} \quad n_{i,j,k} = \text{no. of residue of type } i \text{ at pos. } j \text{ in turn } - \text{type } k$$

$$\text{and} \quad n = \text{number of particular turn type}$$

$$z = \frac{n_{i,j,k} - e_{i,j,k}}{\sqrt{e_{i,j,k}(n_{tot,j,k} - e_{i,j,k})/n_{tot,j,k}}} \quad \text{with} \quad e_{i,j,k} = \frac{n_{i,p}}{n_{tot,p}} * n_{tot,j,k}$$

If $|z| \geq 1.96$ (5% significance level) the number of the specific residues at this position deviates significantly from the expected value, negative values of z indicate under representation, positive values over representation.

2.4 Relibase Substructure Search

The Relibase webinterface is a powerful tool for analysing protein-ligand complexes. In particular the sketcher option could be used for identifying specific interactions. It provides access to 2D/3D ligand substructures and non-bonded protein-ligand interaction searches. The Relibase code was extended, so that Secbase constraints could be used during substructure searches. For example, only protein substructures could be considered that are within a α -helix and more specific at the N_{Cap} position of a α -helix with a length of ten residues. A couple of possible Secbase constraints are the following:

- helix type: α -helix, π -helix or 3_{10} -helix with a minimum and maximum length
- helix position: N_{Cap}, N_{One}, N_{Two}, N_{Three}, C_{Cap}, C_{One}, C_{Two}, C_{Three} and other
- sheet type: parallel, antiparallel or mixed β -sheet with a minimum and maximum number of β -strands
- at a kink among a α -helix or a β -strand
- specific turn families and types

Secbase constraints can be easily set within the sketcher for substructure searches.

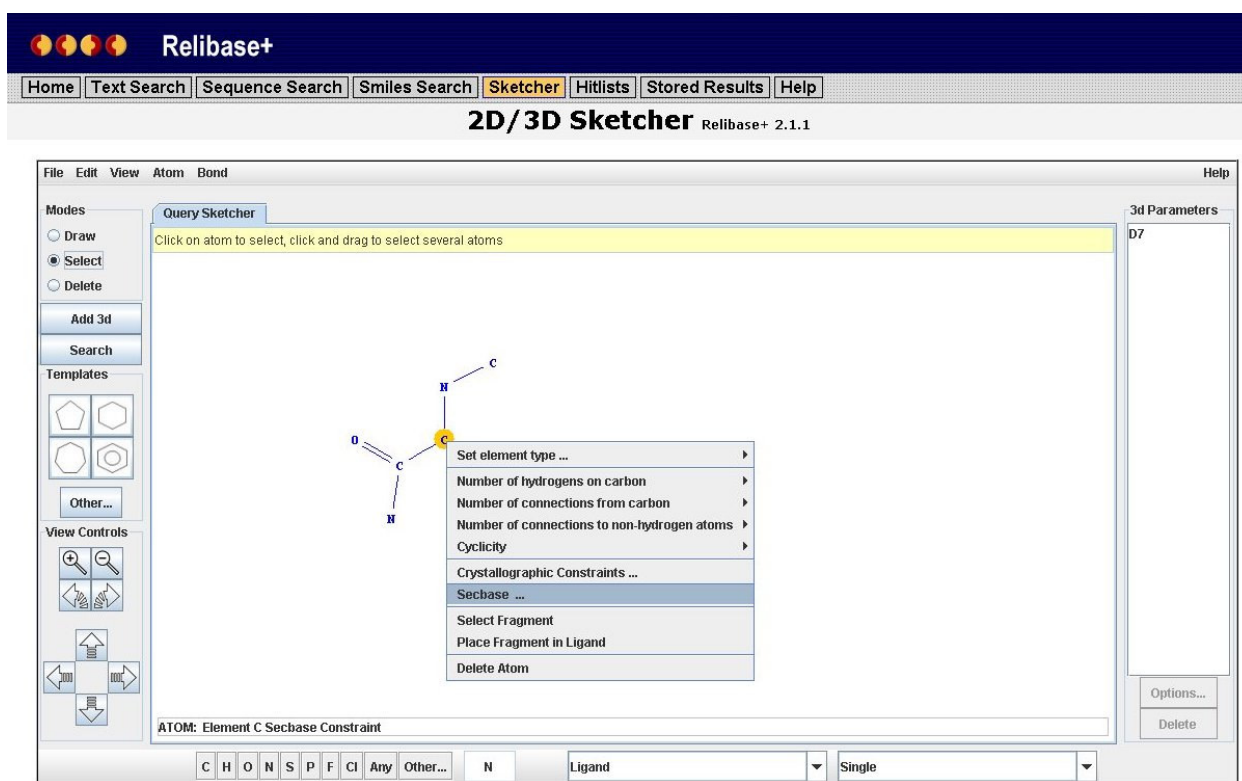


Figure 2.5: Relibase sketcher with additional secbase constraints

2.5 Geometric Description

For α -helices and β -strands geometric descriptions are available that are expressed in terms of vectors together with algorithms that detect kinks and bends.

2.5.1 α -Helices and kinks

Block (Block 2002) described a method for the calculation of helix vectors. The helix is assumed to be a cylinder and the $C\alpha$ -atoms are lying on the surface of this cylinder (Figure 2.6). The distances between the $C\alpha$ -atoms are equal, both in horizontal and vertical direction.

The first step is a projection of two points on a plane with $C\alpha_2$ and perpendicular to the helix axis (Figure 2.6 a)). The point S_1 is lying exactly at half distance between $C\alpha_1$ and $C\alpha_3$ and S_2 is the extension of the vector between $C\alpha_4$ and $C\alpha_3$, whereas $C\alpha_3$ is the midpoint between S_2 and $C\alpha_4$. The three points $C\alpha_2$, S_1 and S_2 are spanning a plane that is perpendicular to the helix axis and for this reason the normal vector \vec{n} of the plane is parallel to the helix axis. Vector \vec{n} is calculated as the vector product of the two vectors $\overrightarrow{S_1C\alpha_2}$ and $\overrightarrow{S_2C\alpha_2}$ (Figure 2.6 b)) and using the normal vector and a parallel translation of the points $C\alpha_1$ and $C\alpha_3$, leads to a triangle that is perpendicular to the helix axis and whose vertices $C\alpha_2$, P_1 and P_2 are lying on the surface of the helix cylinder (Figure 2.6 c)). The perpendicular bisectors of the sides intersect each other in one point and this point is equal to the midpoint of the helix axis that is based on the four $C\alpha$ atoms (Figure 2.6 d)). An example for a calculated midpoint is shown in Figure 2.7 a).

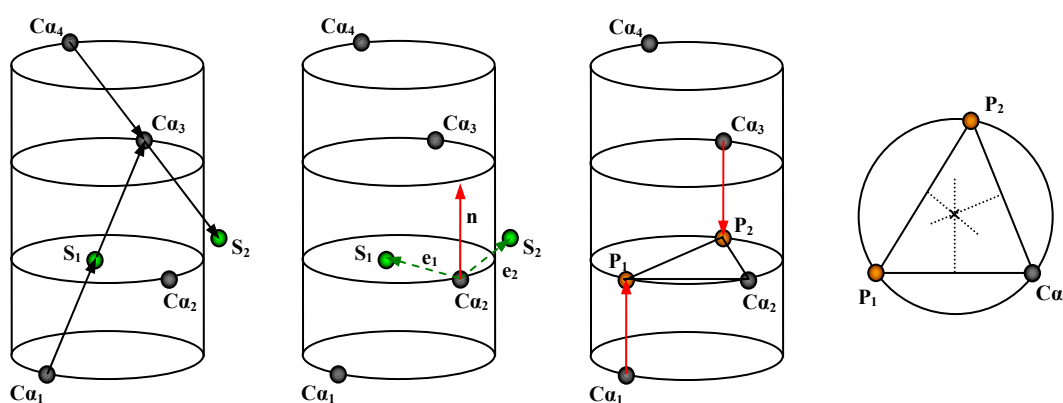


Figure 2.6: Midpoint calculation of four $C\alpha$ of an α -helix

- a) projection of two help points on a plane with $C\alpha_2$
- b) calculation of the normal vector
- c) parallel translation of two points on the surface of the helix
- d) midpoint calculation

Proceeding with this algorithm along the whole helix with overlapping four $C\alpha$ atom section leads to a description of the helix through midpoints (Figure 2.7 b)) and finally, a vector could be drawn through this midpoints (Figure 2.7 c)).

Additionally, kinks within a helix are identified. For each midpoint the angle between the vector to the previous midpoint and the vector to the following midpoint is calculated. If this angle is more acute than a predefined threshold, the helix is kinked. Normally, the threshold angle is 150° , but could be assigned manually.

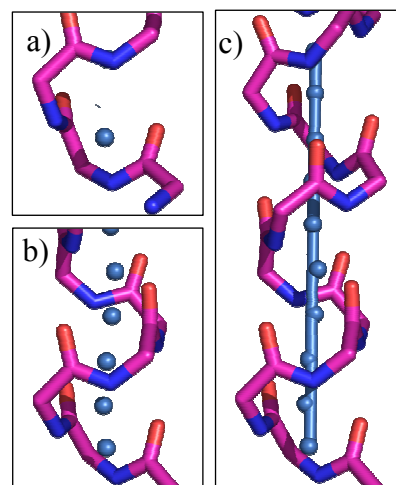


Figure 2.7: Geometric description of α -helices

2.5.2 β -Strands and kinks

The geometric description of β -strands is an easier problem than identifying the kinks and bends of a β -strand (Koch 2004). The vector that is used to describe a β -strand is drawn through the $C\alpha$ -atoms of the β -strand. To identify kinks and a bended β -strand the main chain atoms C, $C\alpha$ and N of one residue and N of the next residue are considered.

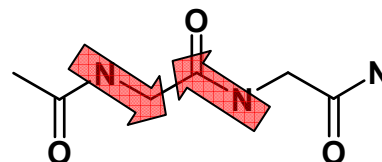


Figure 2.8: Vectors for identifying β -strand conformation

The angle between the vectors through $N_i/C\alpha_i$ and N_{i+1}/C_i shed light on the conformation of the β -strand (Figure 2.8). If the angle is smaller than 86° , the β -strand is assigned to be kinked at this position (Figure 2.9 b). Although the linear part should be expected to have the largest angle, the β -strand is bended, if the angle is bigger than 122° (Figure 2.9 c). Only parts of the polypeptide chain with an angle in between these two cut-offs show an appropriate conformation to build a linear β -strand (Figure 2.9 a).

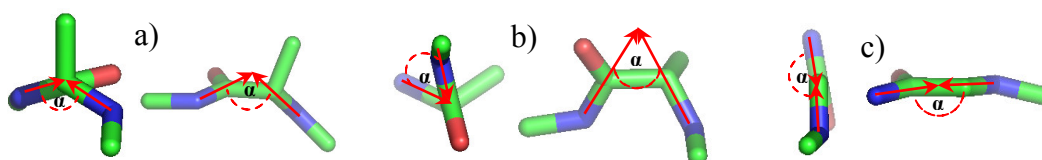


Figure 2.9: a) Linear, b) kinked and c) bended parts of polypeptide chain

The information about kinks and bends is used to divide the β -strand into different parts for each a particular vector is calculated. On the one hand, a β -strand is divided, if a kink has been identified (Figure 2.10 a)) and on the other hand, if a long bended part is recognized. For giving a good approximation of the bended part through a geometric description, it is divided into different parts of similar lengths (Figure 2.10 b)).

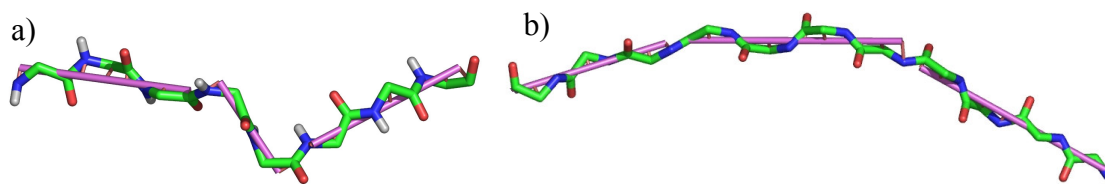


Figure 2.10: Geometric description (purple sticks) of a) a kinked β -strand and b) a bended β -strand

An analysis (data not shown) reveals the specific values of the angles. All possible angles for both cut-offs were used to calculate an rmsd-value between the corresponding geometric description and the main-chain atoms. Within this analysis, a minimum for the rmsd-values and the number of vectors could be recognized at the cut-off value of 86° for kinked and 122° for bended conformations. Similarly to the helix cut-off value for a kink, these threshold values could also be assigned manually.

3 Secbase: Applications

3.1 Conformational parameter

For the calculation of the conformational parameter (see section 2.3) of α -helices a non-redundant dataset was used that is based on the pdb-select list (Hobohm 1994) and contains 2257 protein chains and 9163 α -helices with a minimum length of 8 residues. The minimum length was chosen so that each terminal position is unique without overlapping. In smaller α -helices residues in the middle of the α -helix could be designated with atoms of the opposing terminal end as either N- and C-terminus residues. For example in an α -helix with 6 residues the 3rd residue ($N_{\text{two}}/C_{\text{three}}$) and the 4th residue ($N_{\text{three}}/C_{\text{two}}$) are belonging both to the N- and C-terminus. Table 3.2 shows the result for α -helices with a length of 6 residues. Comparing the values in both tables reveals different amino acid propensities and therefore the data extracted from smaller helices should be neglected with respect to a general conformational analysis, since the N- and C-terminus are not independent.

Table 3.3 compares these results to a previous survey of Penel et al. (Penel 1999) including 2102 helices from 298 protein structures and a survey of Aurora et al (Aurora 1998) including 1316 helices from 274 polypeptide chains. Penel et al. (Penel 1999) analysed the N-terminus and the middle of α -helices and Aurora et al. (Aurora 1998) both termini. Both datasets contain only chains with a sequence homology smaller than 25%. In general, the results for the N-terminus and residues within the α -helix from Penel et al. (Penel 1999) agree well with our results, except the cysteine propensity is found to be higher in our case. In contrast, the results from Aurora et al. differ frequently from this study and the investigation of Penel et al. at the N-terminus. With respect to the analysis of the C-terminus our results also show differences to the Penel study. Possibly this results from the much lower count of helices considered in the older analysis.

The analysis of the N-terminus reveals no surprises; most amino acid propensities are now well understood (Penel 1999): The high propensities of Asp and Asn at the N_{Cap} position are based on their hydrogen bond potential to interact with the free NH group of the N_{two} position. Ser and Thr could be hydrogen bonded to the N_{three} position. Glu, Gln and Asp at the N_{two} position can form a hydrogen bond to their own backbone NH groups and Glu and Gln at N_{three} position can form a hydrogen bond to the free NH group at N_{Cap} position. Additionally,

sidechain-sidechain interactions are favourable between certain residues with high occurrence propensities ($N_{\text{cap}} - N_{\text{three}}$: Ser/Thr - Glu/Asp/Gln, $N_{\text{cap}} - N_{\text{two}}$: Ser/Thr - Glu/Ser and Arg - Asp). The mostly low occurrence of Gly, Ile, Leu, Met and Val can be explained by their non-polarity and the usually high solvent exposure at N-terminus.

Ala shows a high propensity throughout the whole helix, which is due to the short sidechain which shows minimum loss of configurationally entropy during folding (Penel 1999). In contrast, Gly shows a low propensity throughout the whole helix, which could be explained by the flexible backbone and the higher loss of backbone conformational entropy compared to other residues (Cochran 2001). The only exception of this is found at the C_{cap} with the highest propensity for Gly, since Gly is capable to adopt unusual backbone torsion angles. This unusual conformation is needed to provide a free CO group for hydrogen bonding to the free NH group of the following residues (Doig 1995). This special conformation at the end of an α -helix are called “capping motif”, since it achieves stabilizing effects at the end of an α -helix (e.g.: the Schellman motif or the α L-motif (Schellmann 1980; Baker 1984)). The low occurrence of Thr, Val and Ile at the C-terminus could be explained by the bulky, β -branched sidechain which leads to a poorer solvation of the free CO groups.

Pro is not capable to form a hydrogen bond and only the ϕ angle is constrained to helical conformation, the ψ angle shows an extended, non-helical conformation (Penel 1999). The high occurrence at N_{one} position can be explained due to these properties.

The higher amino acid propensity of Cys found in this study is not surprising since Cys is capable to form a hydrogen bond to free NH-groups at the N-terminus (see section 1.1.2) and shows good N_{Cap} preferences in experimental studies (Doig 1995).

In general, a α -helix is stabilized both enthalpically and entropically. This is due to the formations of backbone hydrogen bonds, favourable van der Waals interactions compared with random coil parts of the polypeptide chain or electrostatic interactions such as salt bridges to achieve enthalpic contributions and the minimum loss of configurationally entropy during folding for the short Ala sidechain (the difference between the Ala and Gly propensities can be explained by a large reduction in conformational space available when the sidechain H of a Gly is replaced by a CH_3 in Ala) (Koehl 1999).

aa	Ncap		None		Ntwo		Nthree		other	
ALA	343	< 0.5	766	> 1.11	858	> 1.25	848	> 1.23	7025	> 1.51
ILE	149	< 0.27	469	< 0.87	304	< 0.56	413	< 0.76	4784	> 1.3
LEU	331	< 0.4	834	1.01	528	< 0.64	770	< 0.93	8004	> 1.43
MET	92	< 0.54	130	< 0.77	112	< 0.66	187	1.1	1592	> 1.39
PHE	191	< 0.49	347	< 0.89	267	< 0.69	358	0.92	2758	> 1.05
PRO	518	> 1.28	994	> 2.46	374	0.92	199	< 0.49	240	< 0.09
VAL	195	< 0.31	530	< 0.83	396	< 0.62	607	0.95	4443	> 1.03
ARG	295	< 0.64	487	1.05	434	0.94	363	< 0.79	4078	> 1.3
ASP	1319	> 2.46	525	0.98	853	> 1.59	794	> 1.48	2506	< 0.69
GLU	375	< 0.61	862	> 1.4	1567	> 2.54	1234	> 2.0	4693	> 1.13
LYS	348	< 0.63	613	> 1.11	580	1.05	448	< 0.81	4186	> 1.12
ASN	887	> 2.18	221	< 0.54	331	< 0.81	256	< 0.63	2058	< 0.75
CYS	146	0.98	67	< 0.45	63	< 0.42	93	< 0.62	837	< 0.83
GLN	236	< 0.66	416	> 1.16	475	> 1.33	587	> 1.64	3174	> 1.31
HIS	222	0.99	163	< 0.72	204	0.91	209	0.93	1256	< 0.82
SER	1406	> 2.54	456	< 0.82	575	1.04	455	< 0.82	2507	< 0.67
THR	1093	> 2.15	389	< 0.77	416	< 0.82	518	1.02	2560	< 0.75
TRP	69	< 0.5	152	1.11	113	< 0.82	133	0.97	1008	> 1.08
TYR	193	< 0.58	311	0.93	231	< 0.69	319	0.95	2176	0.96
GLY	742	> 1.18	390	< 0.62	451	< 0.72	323	< 0.52	1749	< 0.41
OTH	13	< 0.31	41	0.99	31	0.75	49	1.18	350	> 1.25
ALL	9163	1	9163	1	9163	1	9163	1	61984	1

aa	Cthree		Ctwo		Cone		Ccap	
ALA	1179	> 1.71	986	> 1.43	1012	> 1.47	738	> 1.07
ILE	609	> 1.12	586	1.08	386	< 0.71	248	< 0.46
LEU	1431	> 1.73	1194	> 1.45	1268	> 1.54	865	1.05
MET	257	> 1.52	209	> 1.23	213	> 1.26	169	1
PHE	411	1.06	323	< 0.83	358	0.92	360	0.93
PRO	32	< 0.08	12	< 0.03	1	< 0.0	1	< 0.0
VAL	486	< 0.76	546	< 0.86	372	< 0.58	314	< 0.49
ARG	605	> 1.31	692	> 1.5	611	> 1.32	502	1.09
ASP	358	< 0.67	353	< 0.66	352	< 0.66	394	< 0.74
GLU	700	> 1.14	870	> 1.41	800	> 1.3	562	< 0.91
LYS	712	> 1.29	876	> 1.58	788	> 1.42	677	> 1.22
ASN	271	< 0.67	297	< 0.73	435	1.07	667	> 1.64
CYS	163	1.09	101	< 0.68	119	< 0.8	141	0.94
GLN	425	> 1.19	524	> 1.46	436	> 1.22	440	> 1.23
HIS	173	< 0.77	232	1.03	246	1.09	332	> 1.47
SER	351	< 0.63	422	< 0.76	589	1.06	648	> 1.17
THR	275	< 0.54	341	< 0.67	469	0.92	347	< 0.68
TRP	131	0.95	116	0.84	97	< 0.71	67	< 0.49
TYR	323	0.97	261	< 0.78	364	1.09	317	0.95
GLY	204	< 0.33	165	< 0.26	182	< 0.29	1335	> 2.13
OTH	67	> 1.62	57	> 1.38	65	> 1.57	39	0.94
ALL	9163	1	9163	1	9163	1	9163	1

Table 3.1: Amino acid propensities for α -helices with at least 8 residues ('<'/>' representing statistical significance, the two highest/smallest propensities and all propensities smaller than 0.5 or bigger 1.5 are shown in bold)

aa	Ncap		None		Ntwo / Cthree		Nthree / Ctwo		Cone		Ccap	
ALA	54	< 0.71	83	1.09	88	1.15	69	0.9	75	0.98	93	> 1.22
ILE	21	< 0.35	51	0.85	31	< 0.52	72	1.2	68	1.13	42	< 0.7
LEU	69	< 0.75	96	1.05	77	0.84	104	1.13	174	> 1.9	97	1.06
MET	14	0.74	17	0.9	11	0.58	30	> 1.6	29	> 1.54	20	1.06
PHE	23	< 0.53	49	1.14	40	0.93	80	> 1.85	66	> 1.53	35	0.81
PRO	76	> 1.69	118	> 2.63	39	0.87	17	< 0.38	2	< 0.04	0	< 0.0
VAL	39	< 0.55	65	0.92	43	< 0.61	83	1.17	82	1.16	54	< 0.76
ARG	36	< 0.7	51	0.99	46	0.9	44	0.86	62	1.21	50	0.97
ASP	120	> 2.02	54	0.91	89	> 1.5	48	0.81	37	< 0.62	58	0.98
GLU	49	< 0.72	89	> 1.3	166	> 2.43	99	> 1.45	45	< 0.66	59	0.86
LYS	35	< 0.57	48	0.78	73	1.19	70	1.14	61	0.99	83	> 1.35
ASN	76	> 1.68	21	< 0.46	38	0.84	28	< 0.62	23	< 0.51	65	> 1.44
CYS	26	> 1.57	17	1.03	16	0.97	25	> 1.51	18	1.09	24	1.45
GLN	34	0.86	25	< 0.63	44	1.11	55	> 1.38	39	0.98	36	0.91
HIS	30	1.2	19	0.76	29	1.16	25	1.0	29	1.16	26	1.04
SER	108	> 1.76	69	1.12	60	0.98	50	0.81	46	< 0.75	73	1.19
THR	76	> 1.35	57	1.01	29	< 0.52	41	< 0.73	63	1.12	41	< 0.73
TRP	13	0.85	27	> 1.77	14	0.92	18	1.18	21	1.38	5	< 0.33
TYR	29	0.78	31	0.84	39	1.05	46	1.24	50	> 1.35	39	1.05
GLY	86	> 1.24	28	< 0.4	42	< 0.6	13	< 0.19	22	< 0.32	115	> 1.65
OTH	3	0.65	2	0.44	3	0.65	0	< 0.0	5	1.09	2	0.44
ALL	1017	1.0	1017	1.0	1017	1.0	1017	1.0	1017	1.0	1017	1.0

Table 3.2: Amino acid propensities for α -helices with 6 residues ('<'/>' representing statistical significance, the two highest/smallest propensities and all propensities smaller than 0.5 or bigger 1.5 are shown in bold)

aa	Ncap	P	A	N1	P	A	N2	P	A	N3	P	A
ALA	0.5	0.5	0.67	1.11	1.2	1.10	1.25	1.3	1.39	1.23	1.3	1.43
ILE	0.27	0.2	0.78	0.87	0.9	1.06	0.56	0.5	0.64	0.76	0.6	<u>1.18</u>
LEU	0.4	0.5	0.79	1.01	0.8	0.84	0.64	0.6	<u>0.91</u>	0.93	0.8	<u>1.52</u>
MET	0.54	0.5	<u>0.98</u>	0.77	0.9	0.90	0.66	0.5	<u>1.10</u>	1.1	1.0	<u>1.68</u>
PHE	0.49	0.4	<u>0.96</u>	0.89	0.9	0.90	0.69	0.7	<u>1.00</u>	0.92	1.0	1.10
PRO	1.28	1.5	1.12	2.46	2.3	1.67	0.92	1.0	0.94	0.49	0.5	0.15
VAL	0.31	0.3	0.67	0.83	0.8	0.76	0.62	0.6	0.70	0.95	0.9	<u>1.14</u>
ARG	0.64	0.6	0.76	1.05	0.9	1.05	0.94	0.8	0.95	0.79	0.8	<u>1.33</u>
ASP	2.46	2.4	1.58	0.98	0.8	1.14	1.59	1.8	1.64	1.48	1.6	<u>0.90</u>
GLU	0.61	0.6	<u>0.94</u>	1.4	1.3	2.30	2.54	2.6	2.07	2.0	2.3	1.70
LYS	0.63	0.5	0.84	1.11	1.0	1.08	1.05	0.9	0.80	0.81	0.8	0.82
ASN	2.18	2.0	1.28	0.54	0.5	0.72	0.81	0.8	0.67	0.63	0.6	0.55
CYS	0.98	0.5	0.37	0.45	0.6	0.26	0.42	0.6	0.52	0.62	0.6	0.52
GLN	0.66	0.4	<u>1.05</u>	1.16	1.3	1.31	1.33	1.4	1.60	1.64	1.8	1.43
HIS	0.99	1.0	0.83	0.72	0.7	0.83	0.91	1.0	<u>1.36</u>	0.93	0.8	0.66
SER	2.54	2.4	1.25	0.82	1.0	0.81	1.04	1.1	<u>0.69</u>	0.82	0.7	0.61
THR	2.15	2.1	1.41	0.77	0.8	0.77	0.82	0.7	0.92	1.02	1.0	<u>0.75</u>
TRP	0.5	0.6	<u>0.94</u>	1.11	1.2	1.26	0.82	0.9	<u>1.10</u>	0.97	1.1	<u>1.68</u>
TYR	0.58	0.6	0.82	0.93	1.0	0.99	0.69	0.5	0.73	0.95	0.9	<u>0.65</u>
GLY	1.18	1.3	0.98	0.62	0.7	0.55	0.72	0.7	0.56	0.52	0.6	0.37

aa	other	P	C3	A	C2	A	C1	A	Ccap	A
ALA	1.51	1.6	1.71	1.73	1.43	1.33	1.47	1.87	1.09	1.19
ILE	1.3	1.3	1.12	1.15	1.08	<u>1.58</u>	0.71	0.90	0.49	0.61
LEU	1.43	1.5	1.73	1.80	1.45	1.63	1.54	1.65	1.06	<u>1.36</u>
MET	1.39	1.6	1.52	2.21	1.23	1.76	1.26	1.35	1.01	<u>1.35</u>
PHE	1.05	1.1	1.06	1.35	0.83	<u>1.22</u>	0.92	0.67	0.91	<u>1.20</u>
PRO	0.09	0.1	0.08	0.07	0.03	0.07	0.0	0.03	0.0	0.10
VAL	1.03	1.1	0.76	0.94	0.86	<u>1.08</u>	0.58	0.51	0.5	0.46
ARG	1.3	1.3	1.31	1.24	1.5	1.39	1.32	1.66	1.09	1.45
ASP	0.69	0.7	0.67	0.68	0.66	0.60	0.66	0.91	0.77	0.72
GLU	1.13	1.1	1.14	1.16	1.41	1.43	1.3	1.88	0.9	<u>1.27</u>
LYS	1.12	1.1	1.29	1.22	1.58	1.71	1.42	1.63	1.24	1.45
ASN	0.75	0.8	0.67	0.70	0.73	0.64	1.07	<u>0.70</u>	1.61	1.33
CYS	0.83	0.8	1.09	<u>0.63</u>	0.68	0.44	0.8	0.33	1.04	<u>0.44</u>
GLN	1.31	1.3	1.19	<u>0.88</u>	1.46	1.37	1.22	1.24	1.21	1.43
HIS	0.82	0.8	0.77	0.76	1.03	1.02	1.09	0.89	1.42	1.55
SER	0.67	0.7	0.63	0.65	0.76	0.42	1.06	<u>0.71</u>	1.16	1.02
THR	0.75	0.7	0.54	0.46	0.67	0.57	0.92	<u>0.50</u>	0.69	0.82
TRP	1.08	1.1	0.95	<u>1.57</u>	0.84	1.00	0.71	<u>1.00</u>	0.5	0.58
TYR	0.96	1.0	0.97	1.10	0.78	<u>1.02</u>	1.09	<u>0.73</u>	0.96	1.06
GLY	0.41	0.4	0.33	0.32	0.26	0.20	0.29	0.33	2.06	<u>0.74</u>

Table 3.3: Comparison of α -helix results to former analysis of Penel et al. (P, including N-terminus and other residues) and Aurora et al. (A, including N- and C-terminus). Large deviation are indicated by underlining (Aurora et al. compared to both other analysis) or in bold (Penel et al. compared to this analysis)

3.2 Ligand to backbone amide interactions

3.2.1 Methods and Materials

An analysis of the interaction geometry between water, ligand-oxygens bound to P or C or water-mediated ligand binding to a backbone amide was made using Relibase+ and Secbase (Figure 3.1) to retrieve information about the general influence of secondary structure elements and the influence of cooperative effects in particular.

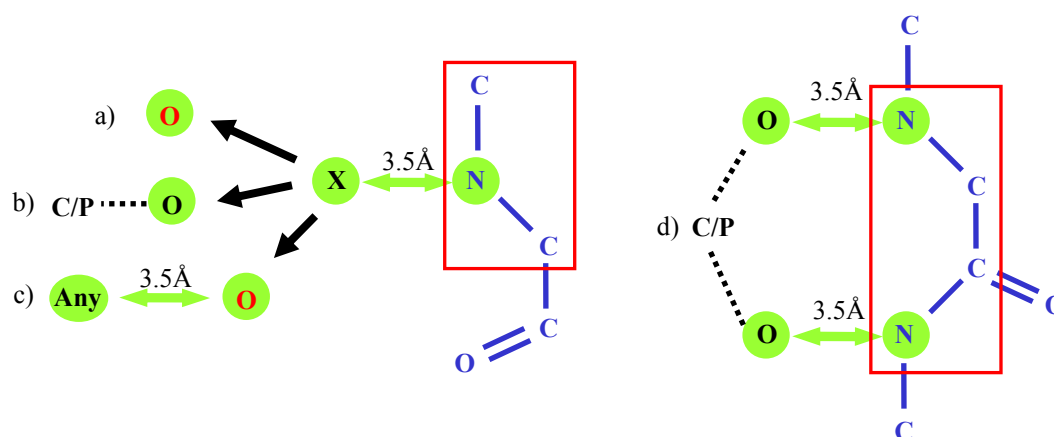


Figure 3.1: Analysed interaction to protein backbone amide with a maximum distance of 3.5 Å between atoms (green arrows: maximum distance of 3.5 Å, blue: protein part, black: ligand part, red: water, dotted lines: any type of bond, red rectangle: atoms for superposition):

a) Water interacting with N; b) Oxygen bound to C or P; c) Water-mediated ligand binding (any ligand atom); d) Two oxygens connected via C or P interacting with two neighbored backbone amides

The following Secbase constraints were used on backbone amides to obtain the data with respect to the type of secondary structure element:

For random chain residues (RCRs):

- helix type: *must not be in any helix*
- sheet properties: *must not be in any sheet*

For α helix:

- helix type: *alpha helix*
- N_{cap} properties: *each with one of: N_{cap} , N_{one} , N_{two} or N_{three}*

For β -sheets:

- sheet properties: *antiparallel* or *parallel*
- position: both *first* and *last* strand allowed

For further post-processing steps, the resulting superimposed structures were used in combination with Python and Reliscript.

The DSSP energy function (Kabsch 1983) was integrated into Reliscript, so that Reliscript could be used to identify the number of cumulative hydrogen bonds to the corresponding peptide CO group (see Figure 3.2). During the analysis, hydrogens of backbone amides were manually set according to McDonald and Thornton (McDonald 1994).

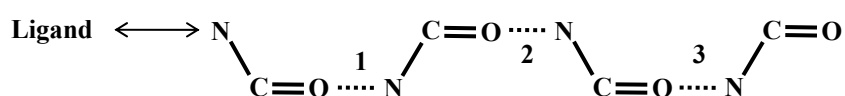


Figure 3.2: Three cumulative hydrogen bonds to the peptide group which is involved in ligand binding

A self-written Python interface for Gnuplot⁵ with standard diagram types was used to automatically generate diagrams shown in this section.

⁵ <http://www.gnuplot.info>

3.2.2 Water – Backbone Amide

Forsyth showed (Forsyth 2002) the lack of a correlation between the different pK_a values at the helix N-terminus and an increasing magnitude of the helix macrodipole. However, it is of interest, whether the identified cooperative effects in α -helices and parallel β -sheets (see chapter 5) have any effect on bonding geometry. For this reason Relibase+ equipped with Secbase was used to retrieve all water molecules in a range of 3.5 Å next to a backbone amide with respect to the secondary structure element (Figure 3.1). Of course, searching the whole Relibase+ will provide biased results as different amount of proteins with high sequence similarities are given. However, this survey analyses local interaction patterns in a rather exhaustive way. Thus, obtained hits should reveal interpretable results independent from protein similarity.

It is quite evident, that the number of N-terminus position is equal to the number of α -helices, but the number of waters differs among different N-terminus positions (Table 3.4). The N_{one} position seems to be the most favourable one for water-binding and solvation. In contrast, N_{three} position is the most unfavourable one. Supposedly, these observations results from the fact that the N_{three} backbone amide is at an excellent position serving as sidechain interaction partner. This is in agreement with the described high propensity residues at the N_{cap} position (Ser and Thr, see section 3.1) which forms a hydrogen bond to the free backbone amide of the N_{three} position. In contrast, the N_{one} position seems to lack this interaction partner, since it shows high frequency for interacting with waters. Therefore, this water is more suitable for replacement by a ligand interaction. Another interesting fact is the pronounced difference between parallel and antiparallel β -sheets which still seeks a reasonable explanation. Possibly, the described differences in cooperative effects between both could serve as an explanation (see chapter 5). A more simple explanation could be that in general the solvent-accessibility of backbone amides is higher for antiparallel β -sheets than for parallel β -sheets, but has to remain open in this context.

	hits	chains	families
RCR	1841292	51135	8331
N_{cap}	39430	13614	1686
N_{one}	70580	15589	1978
N_{two}	54983	14410	1728
N_{three}	20622	9857	1216
antiparallel	58679	13912	1883
parallel	9804	4430	604

Table 3.4: Number of hits for waters near a backbone amide with respect to secondary structure elements and the number of different chains and different families. (RCR: random chain residues)

The amino acid distribution of water-backbone interactions are shown in Table 3.4. The N-terminus values are normalized with the percentage of expected amino acid at this distribution which is based on the amino acid propensity (section 3.1).

$$N = \frac{h_{i,pos}}{h_{pos}} \bigg/ \frac{n_{i,pos}}{n_{pos}}$$

$h_{i,pos}$ = number of hits for residue i at position pos

h_{pos} = overall number of hits at position pos

$n_{i,pos}$ = number of residue i at position pos

n_{pos} = overall number of residues at position pos

	RCR		N_{cap}			N_{one}			N_{two}			N_{three}			anti		para	
	no	[%]	no	[%]	N	no	[%]	N	no	[%]	N	no	[%]	N	no	[%]	no	[%]
ALA	150177	8.2	2759	7.0	1.8	8660	12.3	1.5	7682	14.0	1.5	3171	15.4	1.7	2963	5.0	499	5.1
ILE	74521	4.0	1038	2.6	1.4	3440	4.9	1.0	1531	2.8	0.8	722	3.5	0.7	4066	6.9	803	8.2
LEU	124054	6.7	2081	5.3	1.3	6422	9.1	1.0	3063	5.6	0.9	1707	8.3	1.0	5117	8.7	1004	10.2
MET	26764	1.5	548	1.4	1.3	1217	1.7	1.2	817	1.5	1.2	356	1.7	0.8	1015	1.7	155	1.6
PHE	68457	3.7	1074	2.7	1.3	2929	4.1	1.1	1595	2.9	0.9	684	3.3	0.8	2786	4.7	775	7.9
PRO	45598	2.5	781	2.0	0.3	2490	3.5	0.3	295	0.5	0.1	163	0.8	0.4	474	0.8	44	0.4
VAL	105750	5.7	1434	3.6	1.5	4515	6.4	1.1	2350	4.3	1.0	775	3.8	0.5	6659	11.3	1539	15.7
ARG	88537	4.8	1535	3.9	1.2	3317	4.7	0.9	2472	4.5	1.0	982	4.8	1.2	2756	4.7	310	3.2
ASP	140365	7.6	5010	12.7	0.9	4968	7.0	1.2	5625	10.2	1.1	1571	7.6	0.9	2291	3.9	474	4.8
GLU	119367	6.5	2016	5.1	1.2	6672	9.5	1.0	7916	14.4	0.9	2213	10.7	0.8	3642	6.2	392	4.0
LYS	104168	5.7	1446	3.7	1.0	4365	6.2	1.0	3621	6.6	1.0	1045	5.1	1.0	3216	5.5	302	3.1
ASN	97141	5.3	2974	7.5	0.8	1897	2.7	1.2	1986	3.6	1.0	677	3.3	1.2	1830	3.1	405	4.1
CYS	22778	1.2	411	1.0	0.6	590	0.8	1.0	413	0.8	0.8	104	0.5	0.4	1342	2.3	195	2.0
GLN	61711	3.4	1167	3.0	1.1	3201	4.5	1.1	2790	5.1	1.0	743	3.6	0.6	2085	3.6	212	2.2
HIS	45967	2.5	936	2.4	1.0	1366	1.9	1.1	1120	2.0	0.9	429	2.1	0.9	1405	2.4	290	3.0
SER	141272	7.7	4885	12.4	0.9	3816	5.4	1.0	3594	6.5	1.0	1127	5.5	1.1	4416	7.5	533	5.4
THR	124215	6.7	2430	6.2	0.5	3061	4.3	1.0	2560	4.7	1.1	1180	5.7	1.1	6379	10.9	780	8.0
TRP	28514	1.5	527	1.3	1.5	1613	2.3	1.3	616	1.1	0.8	418	2.0	1.3	1569	2.7	107	1.1
TYR	60941	3.3	1313	3.3	1.5	2258	3.2	0.9	1310	2.4	0.9	781	3.8	1.0	2101	3.6	575	5.9
GLY	210993	11.5	5065	12.8	1.6	3783	5.4	1.3	3627	6.6	1.4	1774	8.6	2.6	2567	4.4	410	4.2

Table 3.5: Amino acid distribution for water near the free backbone amide group (RCR: random chain residues, α -helix N-terminus: N_{cap} , N_{one} , N_{two} , N_{three} , anti: antiparallel and para: parallel β -sheet; general high degree of variation is shown in bold)

The results exhibit no surprises, the small amino acid Ala and Gly seem to be a good interaction partner for water molecules, since they lack a big sidechain. Pro cannot establish a hydrogen bond to water and shows therefore a general low interaction frequency. Interestingly, Ser and Thr show only an expected contribution, although they are a favourable interaction partner for oxygens of ligand substructures (see section 3.2.4).

For the analysis of interaction geometries, the amino acids were split into two groups. Gly was analysed separately from other amino acids, since it showed a completely different behaviour.

Table 3.6 shows the mean hydrogen bond length and mean hydrogen bond angle for strong ($\text{N-H}\cdots\text{O}$ angle: $180^\circ < \alpha < 135^\circ$) and weak ($\text{N-H}\cdots\text{O}$ angle: $135^\circ < \alpha < 90^\circ$) hydrogen bonds for all amino acids except Gly with respect to the secondary structure element and the underlying number of cumulative hydrogen bonds (see Figure 3.1 a). The backbone amide hydrogens were set accordingly to McDonald and Thornton (see Materials and Methods). There is no difference between the hydrogen bonding geometry regarding the secondary structure element or the number of cumulative hydrogen bonds. Only the N_{cap} position of a α -helix and residues outside an SSE (*RCR*: random chain residue) show a small change in hydrogen bonding geometry. The mean distance increases slightly and the mean angle decreases for weak H-bond geometries. On the basis of the similarity of N_{cap} to RCRs it is presumably more a conformational ‘fixing’ through the additional H-bond. The expected cooperative effect in α -helices and β -sheets was not seen. Of course, water molecules in protein X-ray structures have to be handled with care, since they are not always well-defined. At least, this study shows the minimum conclusion that the analysis of water interaction geometry gives no indication of an influence of a cooperative effect in α -helices or parallel β -sheets.

In contrast, Figure 3.4 shows that α -helix N-terminus positions exhibit an additional interaction to bonded water above the peptide N-H bond of the next residue. All positions reveal an expected water cluster directly above the peptide N-H bond. Additionally, the N_{cap} , N_{one} and N_{two} position (Figure 3.4 a-c)) show an additional water cluster, which is in the range of the water above the N-H bond of the next residue. In contrast to the α -helix N-terminus, β -sheets do not exhibit this secondary water cluster. Comparing the backbone conformations of these elements provides an explanation (see Figure 3.3). In contrast to the helix

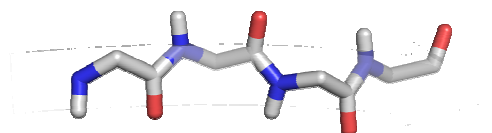


Figure 3.3: β -strand with alternating peptide N-H bond directions

backbone, where all peptide N-H bond vectors are pointing in the same direction, the peptide N-H bonds within strands show alternating directions (Figure 3.3). Furthermore, within the kinked helix backbone the neighbouring peptide N-H show a small distance which favours this behaviour. The protein chain outside any secondary structure element could also show a kinked geometry similar to the kinked helix backbone and therefore, this additional secondary water interaction can also occur outside the helix N-terminus.

		0 H-bond		1 H-bond		2 cum. H-bond		5 cum. H-bond	
Ncap	d _w	2645	2,7 ± 0,35	715	2,82 ± 0,30	1111	2,85 ± 0,29	703	2,83 ± 0,27
	α _w		112,34 ± 13,54		107,65 ± 12,10		106,68 ± 11,08		107,88 ± 10,53
	d _g	10858	2,02 ± 0,20	1560	2,02 ± 0,20	1830	2,03 ± 0,20	1413	2,03 ± 0,21
	α _g		159,59 ± 10,74		160,88 ± 10,33		161,91 ± 10,31		161,21 ± 10,28
None	d _w	1879	2,87 ± 0,28	2074	2,9 ± 0,27	3864	2,9 ± 0,26	2651	2,89 ± 0,27
	α _w		106,09 ± 9,77		104,44 ± 10,26		105,03 ± 9,29		105,54 ± 9,64
	d _g	2605	2,04 ± 0,19	3373	2,05 ± 0,20	6394	2,04 ± 0,19	4660	2,04 ± 0,19
	α _g		161,61 ± 10,41		160,2 ± 10,46		161,55 ± 10,11		162,38 ± 9,97
Anti	d _w	2479	2,86 ± 0,29	1576	2,79 ± 0,34	976	2,74 ± 0,35	187	2,75 ± 0,35
	α _w		107,76 ± 11,70		111,67 ± 13,40		113,05 ± 13,38		113,28 ± 13,60
	d _g	5798	2,03 ± 0,20	12075	2,01 ± 0,19	9256	2,02 ± 0,19	1948	2,03 ± 0,19
	α _g		162,16 ± 10,10		163,57 ± 9,80		163,4 ± 9,74		163,62 ± 9,84
Para	d _w	395	2,82 ± 0,31	191	2,77 ± 0,31	136	2,72 ± 0,34	46	2,82 ± 0,29
	α _w		110,39 ± 12,19		111,95 ± 12,74		114,01 ± 12,71		112,92 ± 12,42
	d _g	1168	2,01 ± 0,20	1401	2,01 ± 0,18	1068	2,01 ± 0,18	736	2,01 ± 0,17
	α _g		163,2 ± 9,96		163,42 ± 9,39		163,46 ± 9,75		163,64 ± 9,50
RCR	d _w	140036	2,74 ± 0,34	78916	2,84 ± 0,32	38588	2,84 ± 0,32	12481	2,84 ± 0,31
	α _w		111,24 ± 13,11		106,92 ± 12,47		107,86 ± 12,37		107,77 ± 12,09
	d _g	429817	2,03 ± 0,21	236726	2,02 ± 0,20	129835	2,03 ± 0,20	40815	2,04 ± 0,19
	α _g		159,45 ± 10,79		161,48 ± 10,35		161,12 ± 10,38		161,1 ± 10,40

Table 3.6: Hydrogen bond geometry for water molecules interacting with all amino acids except Gly with respect to the secondary structure element and the number of cumulative H-bonds. Mean and standard deviation shown for the N-H ... O_{water} distance (d) and the N-H ... O_{water} angle (α) for strong (g: 180° < α < 135°) and weak (w: 135° < α < 90°) hydrogen bond geometries. N_{two} and N_{three} position of a α-helix show similar geometries compared to N_{one}

Water bound to a Gly shows a less distinctive distribution (Figure 3.5). A closer look makes differences apparent between the centre of the water clusters among α-helices and β-sheets. In α-helices water clusters seem to be shifted in direction of the hydrogen that is found at the position of the missing sidechain. The biggest displacement appears at the N_{cap} position, the centre of the water cluster lies at a position shifted approximately by -1.6 Å and 1.0 Å in the projection plane, whereas the water cluster for all other residues are located at the origin (0.0/0.0) (Figure 3.4 a)). The other position at the N-terminus also show a displacement of the centre of the water cluster in the projection plane compared to the water cluster for all other residues:

- N_{one}: -1.0 Å / 0.0 Å (Figure 3.5 b) compared to 0.0 Å / -0.2 Å (Figure 3.4 b)
- N_{two}: -0.6 Å / -0.8 Å (Figure 3.5 c) compared to -0.2 Å / -0.8 Å (Figure 3.4 c)
- N_{three}: -0.6 Å / -1.0 Å (Figure 3.5 d) compared to -0.2 Å / -0.8 Å (Figure 3.4 d)

Interestingly, water near the backbone of an antiparallel β -sheet (Figure 3.5 f) shows a broader distribution into direction of the hydrogen that replaces the missing sidechain, but the centre of the cluster lies above the N-H bond.

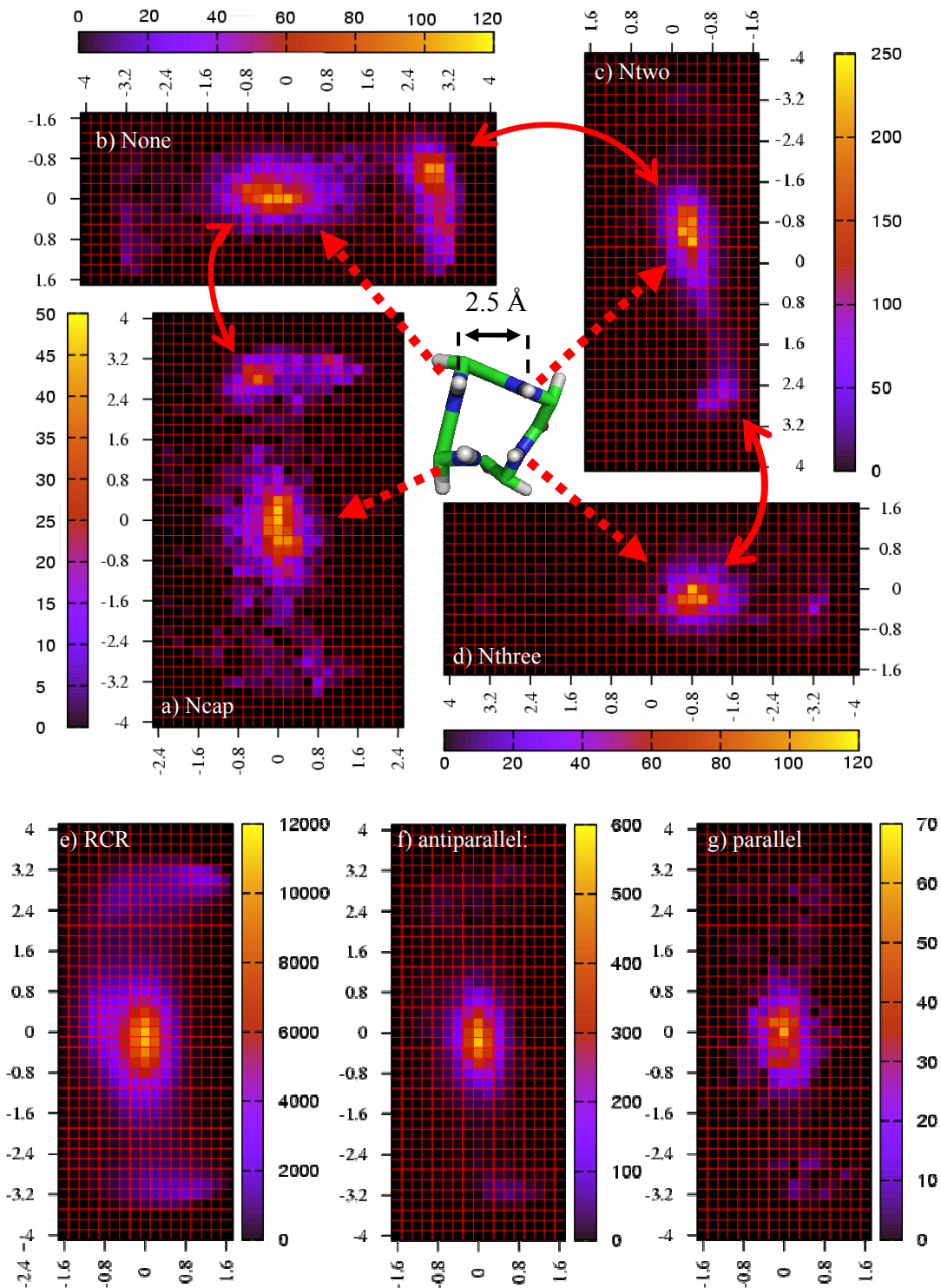
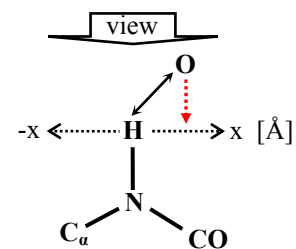


Figure 3.4: Projection of water-O onto a plane perpendicular to the N-H bond for all amino acids except Gly

a) N_{cap} : 2731 hits, b) N_{one} : 6445 hits, c) N_{two} : 5976 hits, d) N_{three} : 2398 hits,; e) RCR: 374102 hits, f) antiparallel β -sheet: 14443 and g) parallel β -sheet: 1659 hits

(H: origin; positive x-axis points into N-C direction; axes: distance to H in Å; α -helical N-terminus position rotated to fit picture of α -helical N-terminus in the middle, SSE data is shown for one cumulative H-bond)



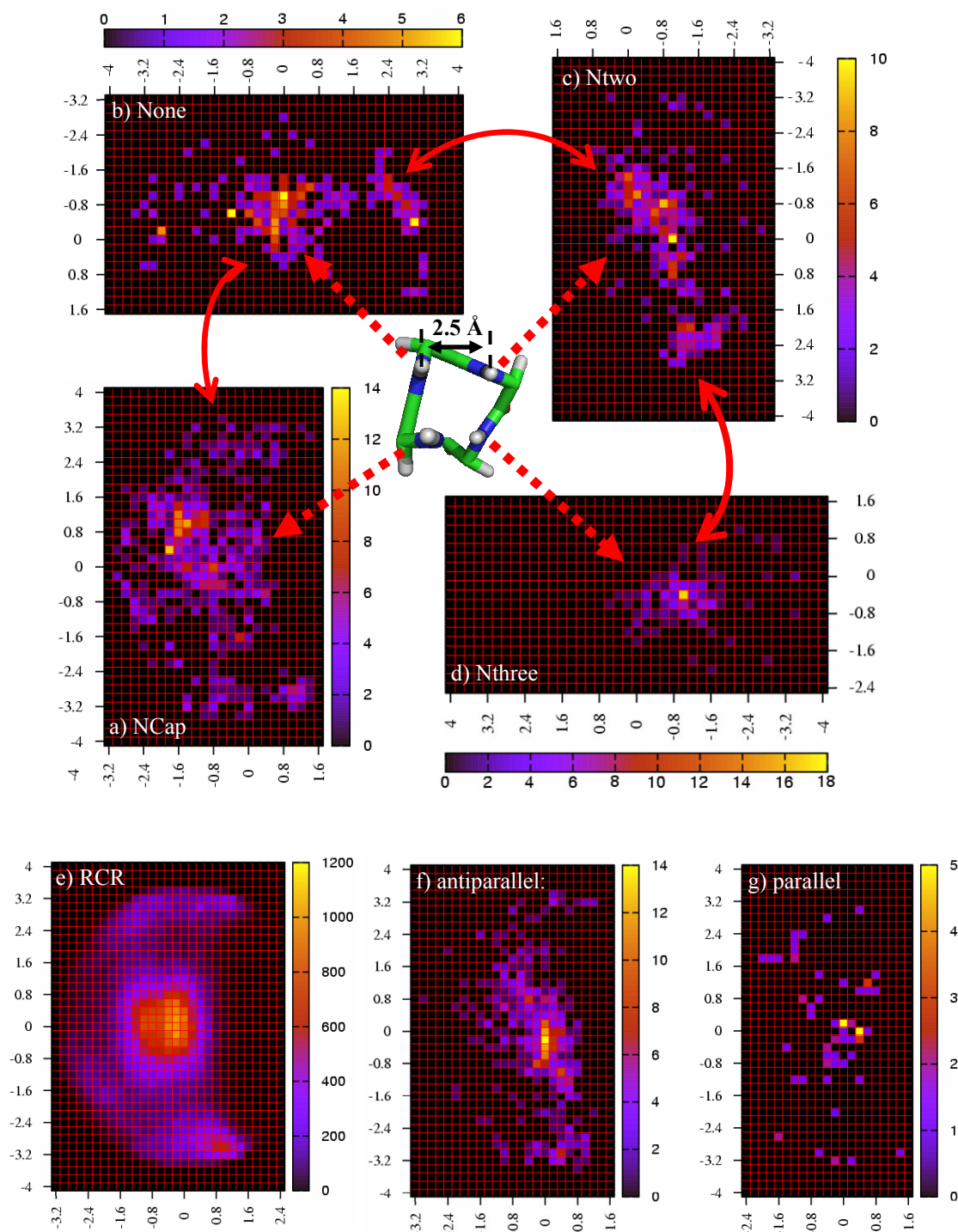
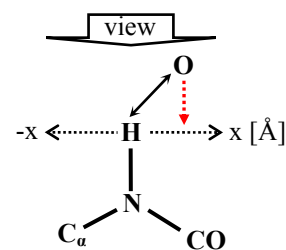


Figure 3.5: Projection of water-O onto a plane perpendicular to the N-H bond for Gly residues

a) N_{cap} : 665 hits, b) N_{one} : 274 hits, c) N_{two} : 353 hits, d) N_{three} : 147 hits,; e) RCR: 34231 hits, f) antiparallel β -sheet: 567 hits and g) parallel β -sheet: 70 hits

(H: origin; positive x-axis points into N-C direction; axes: distance to H in Å; α -helical N-terminus position rotated to fit picture of α -helical N-terminus in the middle, SSE data is shown for one cumulative H-bond)



3.2.3 Water mediated ligand binding to a backbone amide

This analysis deals with waters that are involved in a water-mediated ligand-protein interaction via the backbone amide (Figure 3.1 c). Throughout all analysed elements (Table 3.7), Gly shows the highest occurrence of residues that are involved in this water-mediated ligand binding. The highest Gly occurrence appears at the N_{cap} position with over 77% followed by the N_{three} position with ~55%. Interestingly, the position of waters that act as mediators at the N_{cap} position are shifted off from the position directly above the peptide N-H bond (Figure 3.6 a), see also section 3.2.2. This explains the high occurrence next to Gly, because among all other residues there would be a steric hindrance due to sidechains. Comparing these results with Figure 3.4 a) a water involved in ligand binding is responsible for most of the hits found near a Gly amide backbone. The occurrence of water near a Gly that is not involved in ligand binding seems to be very low. As a possible conclusion, this water replaces the missing sidechain at this position. This highlights the role of waters as a versatile particle to fill empty gaps in protein structures.

The projection of water mediated ligand binding to the N_{one} position of all residues except Gly exhibits that mainly waters bound to the adjacent residue are interacting with the backbone functionality at this position. Considering the high number of directly formed interactions at this position by water molecules, shown in Table 3.4, a water at the N_{one} position might be easily replaceable. Therefore, this position appears to be well accessible for ligand binding. In consequence, at the N_{one} position it is rather unlikely to involve a water molecule as mediator for ligand binding. In contrast, water molecules near the analysed positions show similar projections as described in Figure 3.4 and Figure 3.5 respectively (data not shown). In these cases, water molecules are interacting with the amide backbone and mediate ligand binding. Additionally, interaction to the previous backbone amide could stabilize ligand binding.

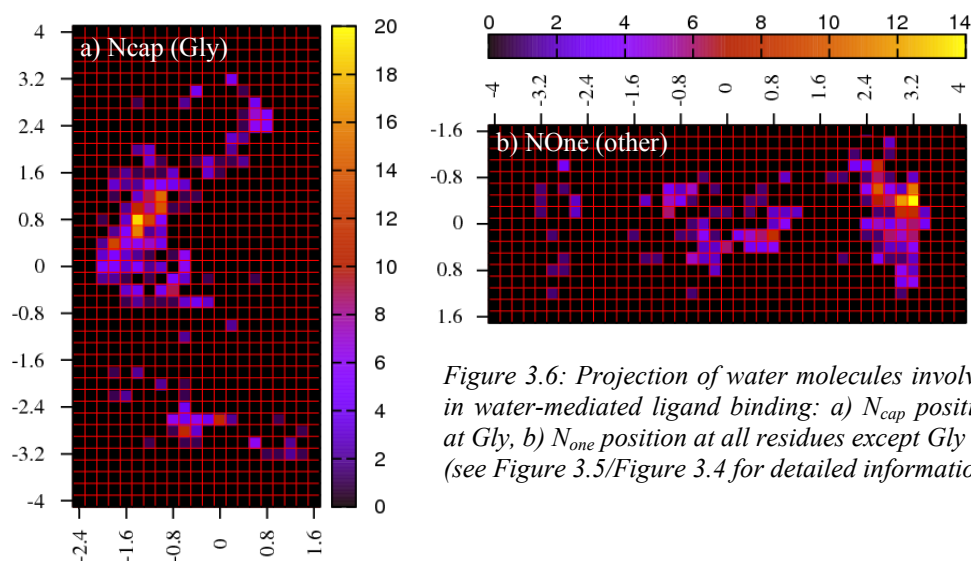


Figure 3.6: Projection of water molecules involved in water-mediated ligand binding: a) N_{cap} position at Gly, b) N_{one} position at all residues except Gly (see Figure 3.5/Figure 3.4 for detailed information)

	RCR		N_{cap}		N_{one}		N_{two}		N_{three}		anti		para	
ALA	6993	8.2%	9	1.5%	14	3.2%	74	11.7%	44	11.6%	152	9.3%	10	6.1%
ILE	4230	5.0%	1	0.2%	10	2.3%	21	3.3%	0	0.0%	27	1.7%	0	0.0%
LEU	3768	4.4%	7	1.2%	10	2.3%	13	2.1%	16	4.2%	112	6.9%	29	17.6%
MET	1383	1.6%	1	0.2%	18	4.1%	18	2.8%	8	2.1%	42	2.6%	5	3.0%
PHE	2899	3.4%	5	0.8%	25	5.7%	28	4.4%	15	4.0%	28	1.7%	14	8.5%
PRO	1174	1.4%	5	0.8%	21	4.8%	0	0.0%	0	0.0%	12	0.7%	0	0.0%
VAL	4693	5.5%	5	0.8%	14	3.2%	29	4.6%	6	1.6%	157	9.6%	12	7.3%
ARG	4091	4.8%	17	2.8%	57	13.0%	22	3.5%	0	0.0%	96	5.9%	3	1.8%
ASP	5420	6.4%	24	4.0%	22	5.0%	30	4.7%	7	1.8%	52	3.2%	4	2.4%
GLU	2306	2.7%	2	0.3%	4	0.9%	16	2.5%	2	0.5%	62	3.8%	0	0.0%
LYS	2271	2.7%	5	0.8%	2	0.5%	9	1.4%	1	0.3%	87	5.3%	0	0.0%
ASN	4102	4.8%	6	1.0%	10	2.3%	82	12.9%	0	0.0%	81	5.0%	18	10.9%
CYS	1219	1.4%	3	0.5%	1	0.2%	26	4.1%	7	1.8%	44	2.7%	0	0.0%
GLN	2264	2.7%	0	0.0%	4	0.9%	2	0.3%	2	0.5%	41	2.5%	0	0.0%
HIS	1700	2.0%	1	0.2%	1	0.2%	30	4.7%	2	0.5%	84	5.1%	1	0.6%
SER	7562	8.9%	34	5.6%	52	11.8%	67	10.6%	22	5.8%	151	9.2%	8	4.8%
THR	6181	7.3%	9	1.5%	76	17.3%	27	4.3%	26	6.9%	159	9.7%	8	4.8%
TRP	1124	1.3%	2	0.3%	12	2.7%	4	0.6%	0	0.0%	9	0.6%	1	0.6%
TYR	2199	2.6%	1	0.2%	4	0.9%	35	5.5%	14	3.7%	43	2.6%	18	10.9%
GLY	19575	23.0%	469	77.4%	82	18.7%	101	15.9%	207	54.6%	194	11.9%	34	20.6%
ALL	85154	100.0%	606	100.0%	439	100.0%	634	100.0%	379	100.0%	1633	100.0%	165	100.0%

Table 3.7: Distribution of residues that are involved in water-mediated ligand binding

3.2.4 Ligand – Backbone Amide

The interaction of free backbone amide groups to ligand oxygen atoms are the focus of this final analysis (Figure 3.1 d). C and P were chosen as possible covalently bonded partners of oxygen since they could be part of a negatively charged ligand-functional group (e.g. carboxylate or phosphate groups) that are postulated to be stabilized by the α -helix macrodipole. Table 3.8 shows the oxygen Sybyl atom type⁶ distribution for all oxygens bound to a carbon atom. N_{one} and N_{two} show the highest percentage of oxygens involved in carboxylate groups (O.co2), N_{cap} and N_{three} the highest percentage for O.3 and sheets for O.2. It seems as if the N-terminus of the α -helix macrodipole can accommodate negatively charged groups. In contrast, the specific structure of the helix N-terminus can give an alternative explanation for this high occurrence of negatively charged groups. As described (see section 3.2.3), secondary interactions from neighbored backbone amide groups can have additional stabilizing effects. Furthermore, both oxygens of a carboxylate group can interact with two neighbored backbone amide groups (Table 3.9). Since this interaction does not appear to be present next to antiparallel and parallel β -sheets it could be postulated, that the kinked backbone of the α -helix, present at the terminal end of this secondary structural element, is needed to provide an appropriate pattern of two parallel NH groups to recognize both oxygens of a carboxylate group.

	O.2		O.3		O.co2	
	no	%	no	%	no	%
RCR	14840	34.48	18271	42.45	9930	23.07
N_{cap}	309	19.36	864	54.14	423	26.5
N_{one}	567	30.67	604	32.67	678	36.67
N_{two}	158	18.72	355	42.06	331	39.22
N_{three}	34	14.47	139	59.15	62	26.38
antiparallel	659	52.93	385	30.92	201	16.14
parallel	73	52.9	61	44.2	4	2.9

Table 3.8: Sybyl atom type distribution for oxygen bound to a carbon

⁶ http://www.tripos.com/mol2/atom_types.html

	$N_{\text{cap}}-N_{\text{one}}$	$N_{\text{one}}-N_{\text{two}}$	$N_{\text{two}}-N_{\text{three}}$	antiparallel	parallel	RCR
no of hits	48	126	40	0	0	1097

Table 3.9 Number of hits of two oxygens of a carboxyl group to two neighbored backbone amides (Figure 3.1 d))

The high number of Lys, Ser, Thr and Gly involved in binding of phosphates (Table 3.10) is expected, since they are known to be part of highly conserved phosphate binding motifs (Hirsch 2007). Interestingly, Ser and Thr occur throughout most of all analysed positions and Lys only occurs at the N_{one} position in a high number of cases.

Presumably, this Lys at the N_{one} position is part of the phosphate-binding loop (P-Loop) (Saraste 1990) with the highly conserved sequence GXXXXGK[T or S]. In contrast, Ser or Thr are found on different structural elements among proteins. They seem to be a more general interaction partner for phosphates without the need for a specific structural environment like the P-Loop. The general interaction pattern for these structures looks similar for all different positions (Figure 3.7). In most cases, the oxygen interacts with both, the backbone amide and the Ser oxygen. Furthermore, this interaction could also be found for carboxylate oxygens (Ser: 1009 hits, Thr: 784 hits).

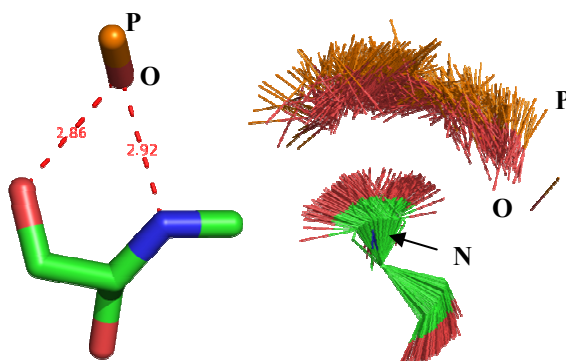


Figure 3.7: Interaction pattern for PO sub-structure with Ser at α -helix N_{one} position left: representative geometry, right: superposition of 557 entries on C_{n-1} , N_n , C_{α_n} (see also Figure 3.1 b))

	RCR		N _{cap}		N _{one}		N _{two}		N _{three}		anti		para	
ALA	2111	6.0%	84	2.5%	207	4.3%	293	8.9%	55	6.8%	15	9.3%	1	9.1%
ILE	769	2.2%	69	2.1%	235	4.9%	218	6.6%	9	1.1%	0	0.0%	0	0.0%
LEU	717	2.0%	42	1.3%	111	2.3%	112	3.4%	40	4.9%	1	0.6%	0	0.0%
MET	456	1.3%	24	0.7%	94	2.0%	68	2.1%	7	0.9%	1	0.6%	0	0.0%
PHE	712	2.0%	16	0.5%	211	4.4%	52	1.6%	12	1.5%	26	16.1%	0	0.0%
PRO	178	0.5%	12	0.4%	95	2.0%	1	0.0%	0	0.0%	0	0.0%	0	0.0%
VAL	1282	3.7%	65	1.9%	231	4.8%	169	5.1%	29	3.6%	0	0.0%	0	0.0%
ARG	2031	5.8%	105	3.1%	261	5.4%	41	1.2%	12	1.5%	9	5.6%	0	0.0%
ASP	928	2.6%	117	3.5%	53	1.1%	25	0.8%	17	2.1%	0	0.0%	0	0.0%
GLU	645	1.8%	28	0.8%	103	2.1%	25	0.8%	33	4.1%	5	3.1%	0	0.0%
LYS	2047	5.8%	195	5.8%	1010	21.0%	27	0.8%	6	0.7%	1	0.6%	6	54.5%
ASN	942	2.7%	85	2.5%	108	2.3%	85	2.6%	8	1.0%	0	0.0%	4	36.4%
CYS	192	0.5%	35	1.0%	32	0.7%	51	1.6%	49	6.0%	1	0.6%	0	0.0%
GLN	538	1.5%	18	0.5%	82	1.7%	25	0.8%	9	1.1%	5	3.1%	0	0.0%
HIS	516	1.5%	13	0.4%	33	0.7%	58	1.8%	1	0.1%	10	6.2%	0	0.0%
SER	4531	12.9%	228	6.8%	557	11.6%	733	22.3%	180	22.2%	49	30.4%	0	0.0%
THR	5020	14.3%	211	6.3%	447	9.3%	974	29.6%	317	39.0%	15	9.3%	0	0.0%
TRP	93	0.3%	1	0.0%	12	0.3%	30	0.9%	0	0.0%	0	0.0%	0	0.0%
TYR	435	1.2%	61	1.8%	75	1.6%	31	0.9%	14	1.7%	0	0.0%	0	0.0%
GLY	10980	31.3%	1948	58.0%	843	17.6%	270	8.2%	14	1.7%	23	14.3%	0	0.0%
ALL	35123	100.0%	3357	100.0%	4800	100.0%	3288	100.0%	812	100.0%	161	100.0%	11	100.0%

Table 3.10: Involvement of specific residue in binding of PO ligand functional groups (RCR: random chain residues, N_{cap} – N_{three}: α -helix N-terminus, anti: antiparallel and para: parallel β -sheets)

The analysis considering two oxygens bound to a phosphorus atom interacting with two neighbored backbone amide groups reveals a high frequency of hits at the α -helix N-terminus (Table 3.11). Therefore, the already mentioned interaction to two neighbored backbone amide groups and the putative dependence on a kinked backbone also applies to bidentate interactions with phosphate groups.

	RCR-N _{cap}	N _{cap} -N _{one}	N _{one} – N _{two}	N _{two} – N _{three}	antiparallel	parallel
no of hits	1076	1789	1414	229	0	0

Table 3.11 Number of two oxygens bound to a phosphorus atom interacting with two neighbored backbone amides (Figure 3.1 d))

Although the interaction to a free backbone amide represents a more general contact, it is interesting to see whether certain positions at α -helix N-terminus are involved in binding of specific ligands. The hits were analysed with respect to the type of ligand (Table 3.12). An interaction from N_{three} backbone amide to ADP, GDP and ATP is highly favoured as reflected by these interactions, where Thr is involved. In contrast, Ser at N_{three} position is more specific to recognize PLP and FAD. Thr at N_{cap} position is highly involved in binding PLP, whereas Ser is more frequently involved in binding FAD and NDP. PLP has been demonstrated to bind to the so-called CaNN structural motif (Hirsch 2007). This CaNN motif is known to

contain the previously described interaction pattern of oxygen that interacts with the backbone and the sidechain of Thr.

Presumably, a general interaction such as a hydrogen-bond formation to a free backbone amide group at the α -helix N-terminus also depends on an additional interaction to the functional groups of the side chain.

	NCap			None			NTwo			NThree		
	Lig	no	[%]	Lig	no	[%]	Lig	no	[%]	Lig	no	[%]
overall	ADP	846	12.65	FAD	946	10.05	FAD	784	11.93	ADP	356	21.93
	GDP	578	8.64	NAD	862	9.16	ADP	756	11.5	GDP	246	15.16
	NAD	388	5.8	ADP	782	8.31	GDP	512	7.79	ATP	184	11.34
	PO4	386	5.77	PLP	700	7.44	NAD	478	7.27	GNP	106	6.53
	PLP	370	5.53	PO4	514	5.46	PLP	466	7.09	PLP	92	5.67
	FAD	348	5.2	FAD	946	10.05	ATP	340	5.17	FAD	76	4.68
Ser	FAD	120	26.32	FAD	142	12.77	GDP	270	18.44	PLP	68	18.89
	NDP	80	17.54	FMN	136	12.23	ADP	268	18.31	FAD	56	15.56
	NAD	44	9.65	PLP	98	8.81	FAD	148	10.11	ATP	42	11.67
	NAP	36	7.89	PO4	94	8.45	GNP	110	7.51	GDP	34	9.44
	PO4	34	7.46	ADP	74	6.65	PLP	74	5.05	GNP	32	8.89
	FAD	120	26.32	DUT	50	4.5	ATP	62	4.23	ADP	26	7.22
Thr	PLP	108	25.59	ADP	130	14.54	ADP	332	17.11	ADP	174	27.49
	FAD	40	9.48	A3P	104	11.63	GDP	210	10.82	GDP	118	18.64
	FAB	32	7.58	FAD	102	11.41	FAD	146	7.53	ATP	88	13.9
	NAP	26	6.16	FMN	86	9.62	PLP	96	4.95	GSP	22	3.48
	PAL	24	5.69	ATP	78	8.72	FMN	90	4.64	ANP	21	3.32
	ATP	16	3.79	PO4	44	4.92	PO4	84	4.33	POP	20	3.16

Table 3.12: Distribution of specific ligands interacting with backbone amide group (overall and for the residues Ser and Thr)

Important 3-letter-codes: ATP: Adenosine-5'-Triphosphate, ADP: Adenosine-5'-Diphosphate, GDP: Guanosine-5'-Diphosphate, NDP: Dihydro-Nicotinamide-Adenine-Dinucleotide Phosphate (NADPH), PLP: Pyridoxal-5'-Phosphate; FAD: Flavin-Adenine Dinucleotide

As mentioned before, this analysis is biased by a higher frequency of certain classes of proteins. Nevertheless, half of all known proteins interact with structural motifs containing phosphates (Hirsch 2007). Accordingly, comparing this high amount of local interaction pattern should lead to statistical significant and equally populated results.

3.2.5 Discussion & Conclusion

The overall α -helix macrodipole as a putative feature to stabilize charges at the α -helix termini is often consulted as a possible explanation for various electrostatic opportunities. So far, there is still no evidence provided to correlate the strengths of this macrodipole with the deviating α -helix lengths.

As described, the high occurrence of polar and charged amino acids at the α -helix N-terminus can be explained due to specific hydrogen bond patterns to free backbone amide groups at the α -helix N-terminus. This seems to be a more convenient explanation than a general influence of the overall helix macrodipole. This is supported by the study of Porter et. al (Porter 2006) who showed the lack of a correlation between carboxyl pK_a shifts at helix N-terminus and an increasing magnitude of the overall α -helix macrodipole. They also explained the pK_a -shifts with a greater number of hydrogen bond interactions (especially to the helix terminus backbone) which is independent from the length of an α -helix and the magnitude of the macrodipole.

The analysis of the interaction geometry of water molecules bound to the amide backbone group also reveals no correlation in terms of bond lengths due to possible cooperative effects and the length of the α -helix or a macrodipole of increasing strength. Of course, this result has to be handled with care and give only an advice since water molecules are problematic within X-ray proteins structures. They are often not well-defined and small changes of the binding geometry are difficult to detect, but this result fits to the other described properties.

As shown in this study, multiple additional features, exemplified in specific ways, can stabilize negatively charged groups at the α -helix N-terminus: Amides of the kinked backbone reveal secondary interactions to atoms bound to the amide backbone of the next residue. Carboxylates and phosphatates can interact with two neighboured backbone amide groups, which explain their high occurrence at the N-terminus. The N-terminus exhibits the kinked backbone that is needed to present two neighboured NH groups in parallel orientation. Furthermore, specific sidechain interactions can also stabilize groups at the α -helix N-terminus.

Ion channels are another prominent example, where the α -helix dipole is used as a general explanation for functionality. Examining the structure of the ClC chloride channel (Dutzler 2002) in more detail, reveals again specific hydrogen bonding and stabilizing patterns at the α -helix N-terminus. The Cl^- ion is among others interacting with two backbone amide groups

of one α -helix N-terminus (Ile 356 and Phe 357) and a Ser hydroxyl group (Ser 107). This hydroxyl group caps the end of a second α -helix and is forming a hydrogen bond to the backbone amide group of Ile 109. Dutzer et al. (Dutzler 2002) suggest that a full positive charge would cause a Cl^- to bind too tightly and partial positive charges provided at the α -helix N-terminus are stabilizing a Cl^- in a way that still allows rapid ionic diffusion rates. Thinking further, this could also provide the explanation why the helix terminus is a widely used motif in protein structures for stabilizing charged ligand substructures instead of using locally placed charged amino acid sidechains.

However, two theoretical studies of charged residues at the α -helix termini involved in capping motifs confirm these observations (Aqvist 1991; Tidor 1994). Both analysis come down to the conclusion, that the first/last turn is the main source for charge stabilising effects, mainly by providing hydrogen bonds. The next turn has only minor influence in terms of dipolar groups and all other turns have no effect. Therefore, this charge stabilizing effects is independent from the length of the α -helix and the overall macrodipole.

To summarize, most observations at α -helix termini could be explained by the specific structure of the first or last turn and the available free backbone amide and carbonyl groups. As a conclusion, the specific structure of the first turn of the α -helix provides a suitable explanation for stabilizing charged structures. Of course, the free backbone and carbonyl groups produces a positively or negatively charged field around the termini, but the magnitude of the overall helix macrodipole seems to have no or only a minor effect.

4 Secbase: Summary

A new module of Relibase is presented which integrates the information about secondary structure elements. The information about helices and β -sheets were retrieved from the PDB and a new uniform classification of all turn families that is based on recent clustering methods, is integrated. Additionally, geometric description for β -strands and helices are provided.

The first part of this study describes the design and performance of Secbase. The second part demonstrates some application to show the scope of analysis easily applicable in combination with the existing functionality of Relibase. The obtained results provide some interesting new insights into the involvement of secondary structure elements into ligand binding.

A survey of water molecules next to the N-terminus was presented to show their involvement in ligand binding. Additionally, the NH-groups of the α -helix N-terminus reveals secondary interactions to ligand substructures bound to the NH group of the next residue. The kinked backbone shows interactions between two neighboured backbone amide groups and carboxylate or phosphate groups, respectively. Furthermore, a more general interaction pattern shows different specificity to the binding of some endogenous ligands depending on the position in secondary structure elements.

Secbase in combination with the functionality of Relibase can be used for knowledge discovery, protein structure prediction and for the comparison of ligand binding to secondary structure elements present in special patterns, motifs or binding domains. In this context, a new and uniform classification of turns gives the opportunity to analyse turns with respect to their role as binding elements to ligands.

Secbase and its easy access to structural data about secondary structure elements will hopefully find its impact on structure-based drug design and protein structure prediction.

5 Cooperative Effects in Hydrogen Bonding

This chapter is already published (Koch 2005) and the content match the original publication. The abstract was removed and the manuscript was slightly adapted to fit into this thesis.

5.1 Introduction

Helices and β sheets are the most frequently occurring secondary structural elements in proteins.(Kabsch 1983) The most abundant helical structures can be subdivided in α helices (backbone angles $\phi, \psi \cong -57^\circ, -47^\circ$), characterised by a repeat unit of 3.6_{13} with a rise of 5.4 \AA per turn and the more elongated 3_{10} helices ($\phi, \psi \cong -49^\circ, -26^\circ$) with a repeat unit of 3.0_{10} and a height of 6.0 \AA per turn. The more densely packed π helix ($\phi, \psi \cong -57^\circ, -70^\circ$) is characterised by a repeat unit of 4.4_{16} and a height of 5.2 \AA per turn and has only rarely been observed in proteins deposited up to now in the Protein Data Bank (PDB).(Berman 2000)

Helices are stabilised mainly by intrastrand hydrogen bonds formed between the carbonyl groups of the peptide backbone as hydrogen bond acceptors and the backbone amide N-H groups operating as H-bond donors. The sidechains attached to the C_α atoms are pointing towards the outside off from the helix axis thus avoiding steric clashes with the backbone and sidechain atoms of the residues forming the next turn. Glycine exhibits no sidechain atoms at C_α , accordingly a larger conformational space is accessible to this residue and it is frequently observed at the end of ordered helical structures next to turn-type structural elements. Proline, lacking the free N-H group, is rarely found in the central part of helices because it is known to break the helical geometry due to its incompetence to form intrahelical hydrogen bonds. Additionally, there are also position-dependent propensities for all other amino acids.(Richardson 1988; Chakrabarti 2001; Engel 2004). Across an α helix, the H-bonds are arranged in such a way that the carbonyl bond of the n th residue arranges parallel to the helix axis towards the facing peptide N-H bond of the $(n+4)^{\text{th}}$ residue. Overall, this geometry results in a nearly ideal linear H-bond geometry with all C=O and N-H bond vectors in identical orientation obviously piling up to a composite overall dipole moment coinciding with the helix axis (Figure 5.1).(Hol 1981; Branden 1999)

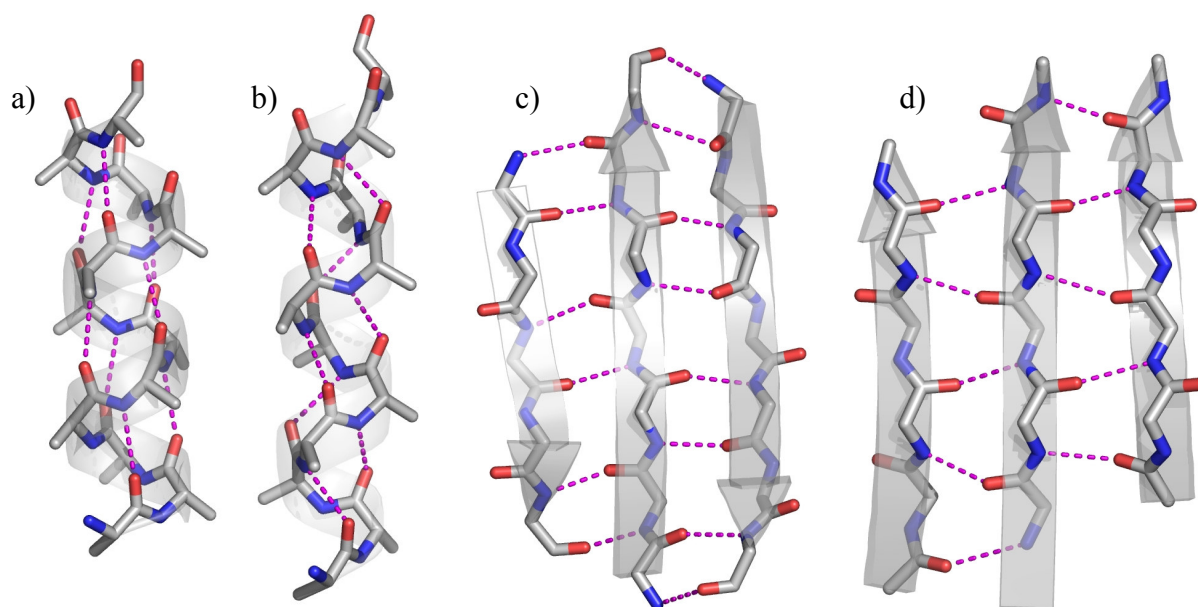


Figure 5.1: Schematic view on the hydrogen-bonding network along a right handed α helix (a), 3_{10} helix (b), antiparallel β sheet (c) and parallel β sheet (images created with PYMOL (DeLano 2002)).

It has been discussed whether this permanent dipole moment influences the H-bond strength and cooperativity in helix formation. In the 3_{10} helix the H-bonds are formed between the n^{th} and $(n+3)^{\text{th}}$ amino acid. In this structural motif, the carbonyl groups are inclined off from the helix axis and the H-bonds are tilted from the linear geometry (Figure 5.1). Theoretical calculations based on model helices such as poly Ala_n (Topol 2001; Wu 2001) predict short 3_{10} helices of one or two turns to be more stable than corresponding α helices of similar size, but beyond a minimal size of two turns, the α helix is calculated to be more stable in solution. In comparison to unfolded random coil structures, helices are stabilized due to cooperative effects.

Another common secondary structure is the β -pleated sheet. In this structural element the successive sidechains of the polypeptide chain are oriented alternating to the opposing sides of the sheet and the peptide plane extends linearly in a rippled geometry. The backbone hydrogen bonds are formed not within the same strand but across two neighbouring strands, connecting the peptide carbonyl oxygen of the first strand with an opposing peptide NH-group of the neighboring strand. The corresponding NH group of the first strand is either connected with the carbonyl group of the same amino acid in the neighboring strand if it shows opposite orientation (antiparallel β sheet) or with the adjacent amino acid if both strands exhibit the same orientation (parallel β sheet, Figure 5.1). The first arrangement results in dual pairs of parallel narrow spaced H-bonds that alternate with widely spaced pairs. The latter parallel sheet shows evenly spaced hydrogen bonds that are tilted in an alternating pattern. (Branden 1999) In proteins β sheets are observed more frequently with antiparallel

topology, characterised by alternating directions of the backbone of the individual strands. Mutually the strands are connected by turns or short loops. Less abundant are β sheets with parallel topology. Their contiguous arrangement along the polymer chain requires connection of the opposing ends via extended crossover helices (eg. TIM barrel) or loops. β sheets formed by L-amino acid residues exhibit a characteristic right handed twist, which is more pronounced in the parallel topology. Mixed sheets are rarely observed, since the strand geometry of two adjacent β -strands facilitates the association of a third strand in corresponding parallel or antiparallel fashion due to a given preorientation. Theoretical calculations on β sheets showed that the observed twist is a consequence of nonbonded interactions experienced between neighboring sidechains which distorts the interchain β sheet H-bonds (Bour 2004).

The different spatial arrangement of the peptide bond and the formed H-bonding network in the various secondary structure elements should give rise to deviating properties. This has been indicated by theoretical calculations, however to support and confirm such simulations we analysed the H-bonding geometry in helices and β sheets across the entire PDB. The present study reveals statistically significant differences in H-bonding patterns for the most frequently occurring secondary structures, namely α helix, 3_{10} helix and parallel or antiparallel β sheets. The analysis have been performed using Relibase+ (Bergner 2001; Hendlich 2003), recently equipped with the additional module Seabase which allows to addresses each protein atom in terms of its associated secondary structure.

5.2 Methods

The presented analysis has been carried out using Secbase, a modular extension of Relibase+. Relibase is an object-orientated data management system and contains the three-dimensional structural information of protein-ligand complexes deposited in the PDB. Secbase integrates the information about secondary structure elements exhibited by individual protein structures with additionally derived data. A Python-based toolkit allows to analyse the data via the object-orientated scripting language Python. Details about the implementation of Secbase and the Python-based toolkit will be described in detail elsewhere.

Two different strategies were used to calculate the hydrogen bonds in helices and sheets.

In α helices the hydrogen bonds are formed across the backbone between the carbonyl oxygen of the n^{th} residue and the hydrogen atom at the amino group of the $(n+4)^{\text{th}}$ residue. In a 3_{10} helix it is formed between O_n and HN_{n+3} . Since the positions of protons are not determined in most X-ray structures, only the distances between the corresponding non-hydrogen atoms can be evaluated. Accordingly, for this analysis the distance between O_n/N_{n+4} and O_n/N_{n+3} respectively was calculated.

Hydrogens bonds are not explicitly deposited in Relibase. To identify possible hydrogen bonds a functionality defined in the Python-based toolkit was used which allows to assign putative hydrogen-bonding partner. All atoms were retrieved coinciding within a hemisphere around the oxygen and nitrogen, respectively. The maximal distance between adjacent partner atoms was requested to be below 3.5 Å and the $C=O \cdots N$ angle had to fall into a range between 90° and 180° (Baker 1984).

A dataset of 1581 protein chains comprising 9055 α helices, 3686 3_{10} helices, 14125 β strands was analysed considering at the same time only database entries for which diffraction data up to a resolution of 1.5Å or better had been recorded.

A reduced dataset solely considering non-redundant proteins was selected using the PDB-select list (Hobohm 1994). It revealed 358 protein chains with a sequence homology beyond 25%. This unique set showed in all analyses similar trends but at minor statistical significance. Accordingly, it was decided to base the final analysis on the entire data pool. On a first glance, the assumption was made that in this evaluation the H-bonding pattern found in the individual secondary structure elements is not systematically perturbed by the overall folding pattern adopted by the structures considered in our analysis.

5.3 Results

The H-bonding distance was evaluated between peptide carbonyl O and peptide N observed in crystal structures deposited in the PDB with respect to the embedding secondary structure.

5.3.1 Helical structures

The average H-bond length in α helical and 3_{10} helical secondary structure elements was compared; π helical structures are too scarcely populated to provide statistically significant correlations. To investigate putative cooperativity in different H-bonding patterns with respect to their surrounding environment, each H-bonding geometry was individually correlated with local descriptors. Accordingly, all α helices composed by one (3 H-bonds), two (7 H-bonds), three (11 H-bonds), four (15 H-bonds) and five turns (19 H-bonds) were individually compiled (Table 5.1), helices with more than five turns occur only rarely and do not change the correlation significantly. Similarly 3_{10} helices were subdivided into groups of one (2 H-bonds), two (5 H-bonds), three (8 H-bonds), four (11-H-bonds) and five turns (14 H-bonds). Facing the results of both evaluations, shown in Figure 5.2, makes completely opposing trend apparent for both types of helices. In α helices the mean lengths of hydrogen bonds are successively shrinking with increasing length and turn numbers, whereas in 3_{10} helices the corresponding H-bond distance is systematically extended. The angle between the carbonyl group C=O and the facing peptide N remains virtually unchanged at $153^\circ \pm 2^\circ$ in α helices of various lengths, whereas it drops from 122° to $106^\circ \pm 6^\circ$ in 3_{10} helices with increasing length of the helix.

These distinct trends in the molecular dimensions of H-bonds is in good agreement with the theoretical calculations predicting a similar more pronounced shortening of H-bond distances in α helices of increasing length.

		<u>alpha helices</u>					<u>310 helices</u>							
		number of hydrogen bonded turns					number of hydrogen bonded turns							
h-bonds		1	2	3	4	5	h-bonds	1	2	3	4	5		
1		3.011	3.014	3.006	3.010	3.002	1	3.003	3.035	2.994	2.995	3.140		
2		2.984	2.962	2.969	2.970	2.950	2	3.095	3.054	3.070	3.160	3.147		
3		3.054	2.994	2.999	2.999	3.017	3		3.116	3.152	3.229	3.154		
4			2.994	2.975	2.972	2.961	4		3.056	3.140	3.195	3.214		
5			2.983	2.969	2.960	2.966	5		3.102	3.123	3.174	3.228		
6			2.988	2.976	2.967	2.989	6			3.091	3.233	3.278		
7			3.015	2.977	2.984	2.980	7			3.086	3.229	3.260		
8				2.982	2.960	2.953	8			3.114	3.273	3.233		
9				2.988	2.967	2.957	9				3.225	3.269		
10				2.985	2.987	2.984	10				3.221	3.276		
11				3.014	2.981	2.979	11				3.050	3.225		
12					2.963	2.955	12					3.256		
13					2.970	2.967	13					3.285		
14					2.998	2.986	14					3.254		
15					2.993	2.977	15							
16						2.963	16							
17						2.999	17							
18						2.990	18							
19						2.967	19							
h-bond length		Mean					h-bond length	Mean						
Average		3.016	2.993	2.986	2.979	2.976	2.990	Average	3.049	3.073	3.096	3.180	3.230	3.126
Stdev		0.04	0.02	0.01	0.02	0.02	0.02	Stdev	0.07	0.03	0.05	0.08	0.05	0.06
datapoints		3351	13122	21375	16730	17834		3053	1477	240	149	557		

Table 5.1: H-bond length analysis in α -helices and 310-helices according to position and length

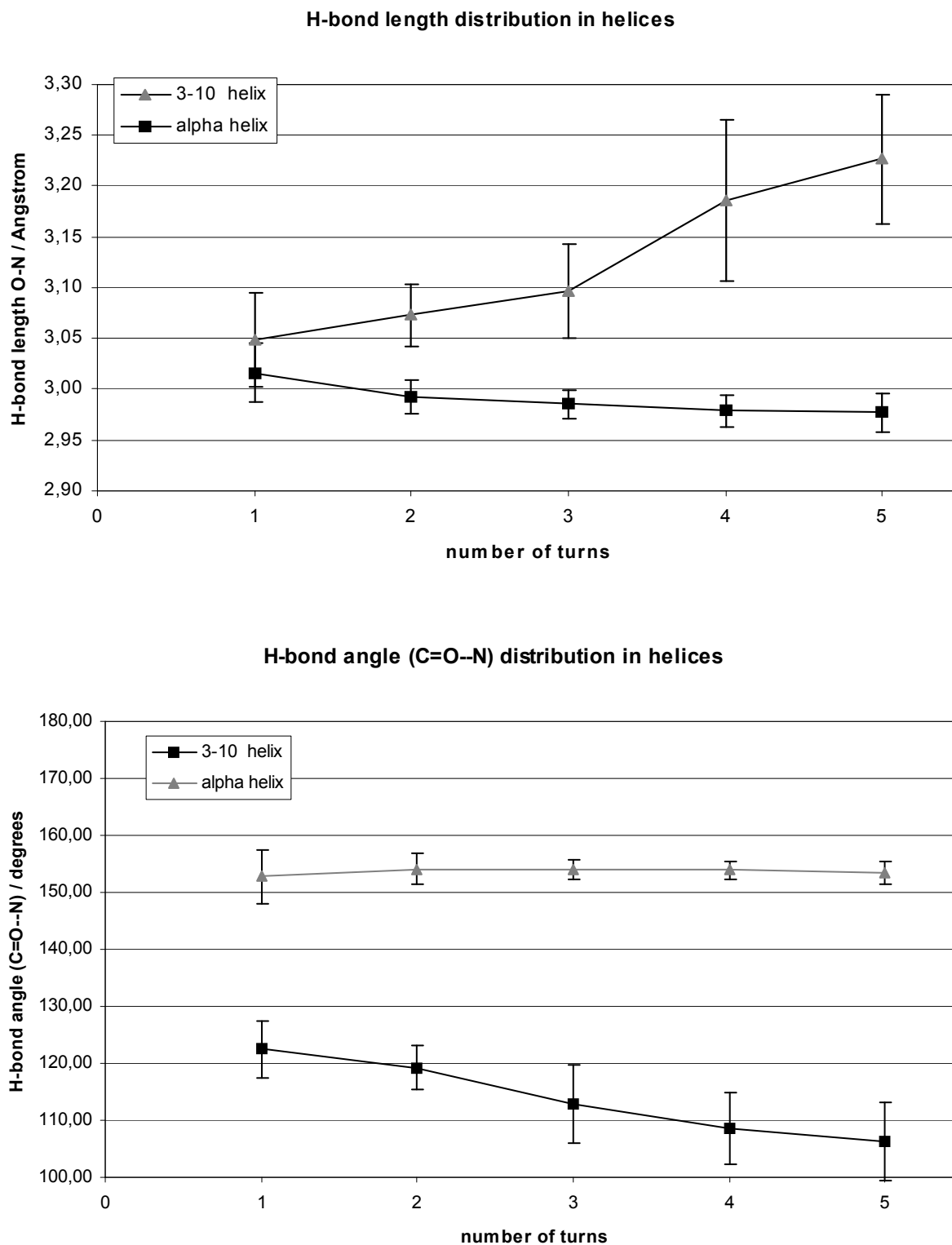


Figure 5.2 Correlation of the mean H-bond length (top) and H-bond angle (bottom) with respect to the helix length expressed by the number of turns present in α helices and 3_{10} helices as observed in crystal structures of proteins with a resolution better than 1.5\AA . The bars indicate the observed standard deviations for the different molecular dimensions.

5.3.2 β sheet structures

Similar analyses were performed for parallel and antiparallel β sheet structures.

Data analysis shows that the mean H-bond length is $2.94 \text{ \AA} \pm 0.05 \text{ \AA}$ for parallel and $2.94 \text{ \AA} \pm 0.03 \text{ \AA}$ for H-bonds in antiparallel β sheets, accordingly for the overall means no difference are apparent. Compared to the average H-bond length of $2.99 \text{ \AA} \pm 0.02 \text{ \AA}$ in α helical structures this indicates, with some care, slightly shorter means. Other analyses of hydrogen bonding in secondary structure elements (Thomas 2001), which described the median H-bond lengths, shows similar results, in fact 3.0 \AA for α helices and 2.93 for β sheets.

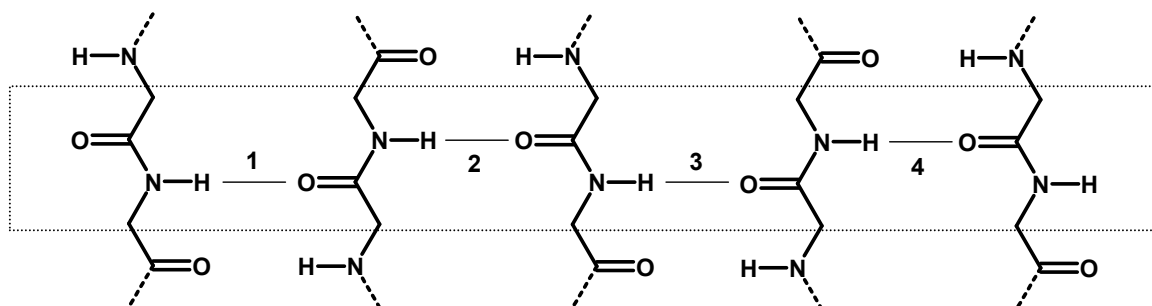


Figure 5.3 Schematic drawing of a parallel β sheet with four hydrogen bonds accumulated perpendicular to the strand direction that build a "hydrogen-bond chain"

However, considering hydrogen bonds formed perpendicular to the strand direction (Figure 5.3), distinct trends were found with respect to the number of hydrogen bonds interconnecting accumulated β -strands. In parallel β sheets, the mean length of hydrogen bonds is successively shrinking from $2.95 \text{ \AA} \pm 0.02 \text{ \AA}$ to $2.89 \text{ \AA} \pm 0.04 \text{ \AA}$ with each additional hydrogen bond considering an array of two up to seven strands (Table 5.2). In contrast, the corresponding value is systematically extended from $2.92 \text{ \AA} \pm 0.01 \text{ \AA}$ to $2.99 \text{ \AA} \pm 0.04 \text{ \AA}$ in antiparallel β sheets of similar size (Table 5.2).

The angle between the carbonyl group C=O and the facing peptide N remains virtually unchanged in antiparallel β sheets at 153° , only the standard deviation increases from 0.02° to 5.28° with the number of piling H-bonds. In contrast it increases from $157^\circ \pm 0.12^\circ$ to $162.3^\circ \pm 3.1^\circ$ in parallel β sheets with increasing number of hydrogen bonds.

<u>parallel β sheets</u>							<u>antiparallel β sheets</u>						
number of accumulated h-bonds							number of accumulated h-bonds						
pos.	2	3	4	5	6		pos.	2	3	4	5	6	
1	2,924	2,911	2,924	2,921	2,865	h-bond length	1	2,900	2,898	2,910	2,926	3,018	h-bond length
2	2,928	2,898	2,893	2,896	2,917		2	2,908	2,906	2,914	2,938	2,942	
3		2,924	2,893	2,913	2,859		3		2,905	2,926	2,953	2,946	
4			2,919	2,895	2,848		4			2,916	2,912	2,912	
5				2,891	2,877		5				2,906	2,990	
6					2,910		6					3,034	
						mean							mean
Average	2,926	2,911	2,907	2,903	2,879	2,905	Average	2,904	2,903	2,917	2,927	2,974	2,925
Stdev	0,003	0,013	0,016	0,013	0,028	0,015	Stdev	0,005	0,004	0,007	0,019	0,048	0,017
datapoints	2070	2028	1802	1108	173			9414	7342	2770	1060	563	

Table 5.2: H-bond length analysis in parallel and antiparallel β -sheets according to position and length

5.4 Discussion and Conclusions

As confirmed by this study, right handed α helices of increasing length exhibit statistically significant trends towards a systematic shortening of the hydrogen-bonding distance along the helix axis. This suggests that some sort of energetic determinant is responsible to result in a cooperative stabilisation of α helices.

The strong polarising properties of the backbone carbonyl groups in α helical structures has recently been observed experimentally by X-ray crystal structure analysis of cholesterol oxidase at atomic resolution (Lario 2003). The electron density maps of this high resolution study show that the carbonyl groups, located in α helical portions, exhibit a characteristically increased electron density at the oxygens indicating a more pronounced negative partial charge and thus stronger polarization of the C=O bond. Due to the increasingly polarised carbonyl bond the oxygen should display stronger H-bond acceptor properties thus enhancing the H-bonding strength and stability in extended α helices.

Quantum mechanical calculations on infinite α helical structures applying periodic boundary conditions suggest an increasing H-bond strength with growing helical length (Ireta 2003). This observation most probably results from an increasing polarisation of the growing permanent dipole built-up along the α helix (Improta 2001). Calculations on growing chains of H-bonded formamides suggest a stabilization of the interconnecting H-bonds between individual formamides with growing chain length (Kobko 2003). H-bonds closer to the center of a H-bonded chain are calculated shorter than at the rim. This agrees well with our observed pattern (Figure 4, top) in X-ray structures and results indicated by a NMR study performed on helical proteins (Jaravine 2001). Molecular mechanical (CHARMM) calculations performed on polyalanine model helices (Elstner 2000; Cui 2001) is qualitatively in agreement with the present statistical analysis of X-ray data. The computed data also indicates a shortening of hydrogen bonds in α helices of increasing length. It was observed that hydrogen bonds in α helices are simulated as systematically too long by general purpose molecular mechanical force fields compared to quantum mechanical models of polyalanine helices. Advanced and specifically adapted force fields considering polarizability are a promising alternative (Rick 2002) to handle cooperative electrostatic effects by simple molecular mechanics calculations.

The analysis of parallel β sheets with respect to the number of hydrogen bonds bridging across accumulated strands shows results that are surprisingly comparable to the trends observed in α helices. In fact a shortening of the hydrogen bond distance with increasing number of piling H-bonds and shorter hydrogen bonds at the center of an H-bonded chain (Figure 5.4, bottom) are observed. Interestingly, the opposite is indicated for antiparallel sheets.

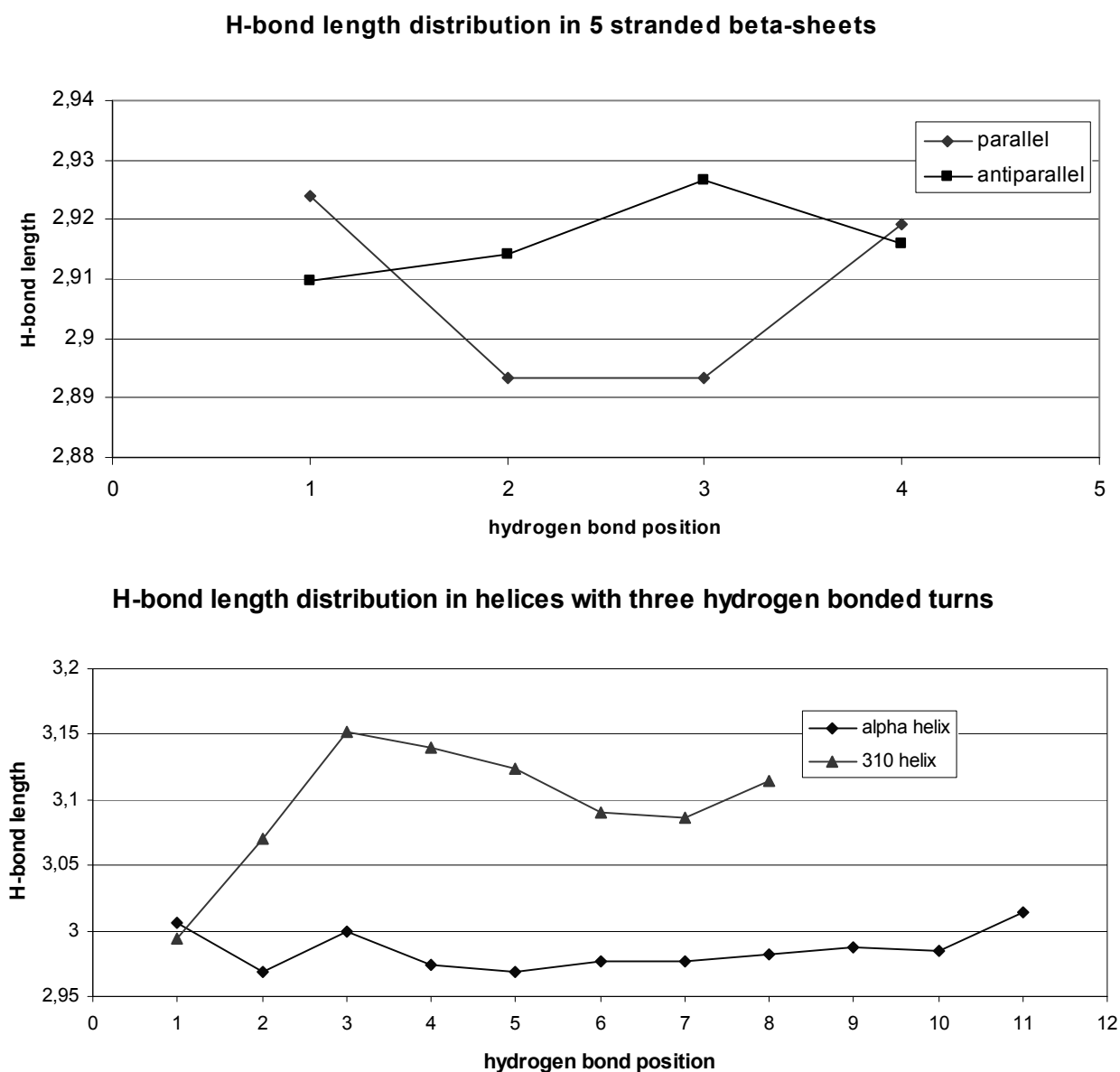


Figure 5.4 : Trends of the mean H-bond distances with respect to the position within a helix with 3 turns (top) or 5 stranded β sheets with a hydrogen bond chain of 4 H-bonds perpendicular to strand direction (bottom) as observed in protein crystal structures with a resolution better than 1.5Å.

It is in question whether similar features operate in helices and sheets resulting in such trends. In helices all bond dipole vectors accumulate along the same direction (Figure 5 a,c), whereas in sheets a pattern of alternating rows of vectors pointing in opposing directions is given (Figure 5.5 b,d).

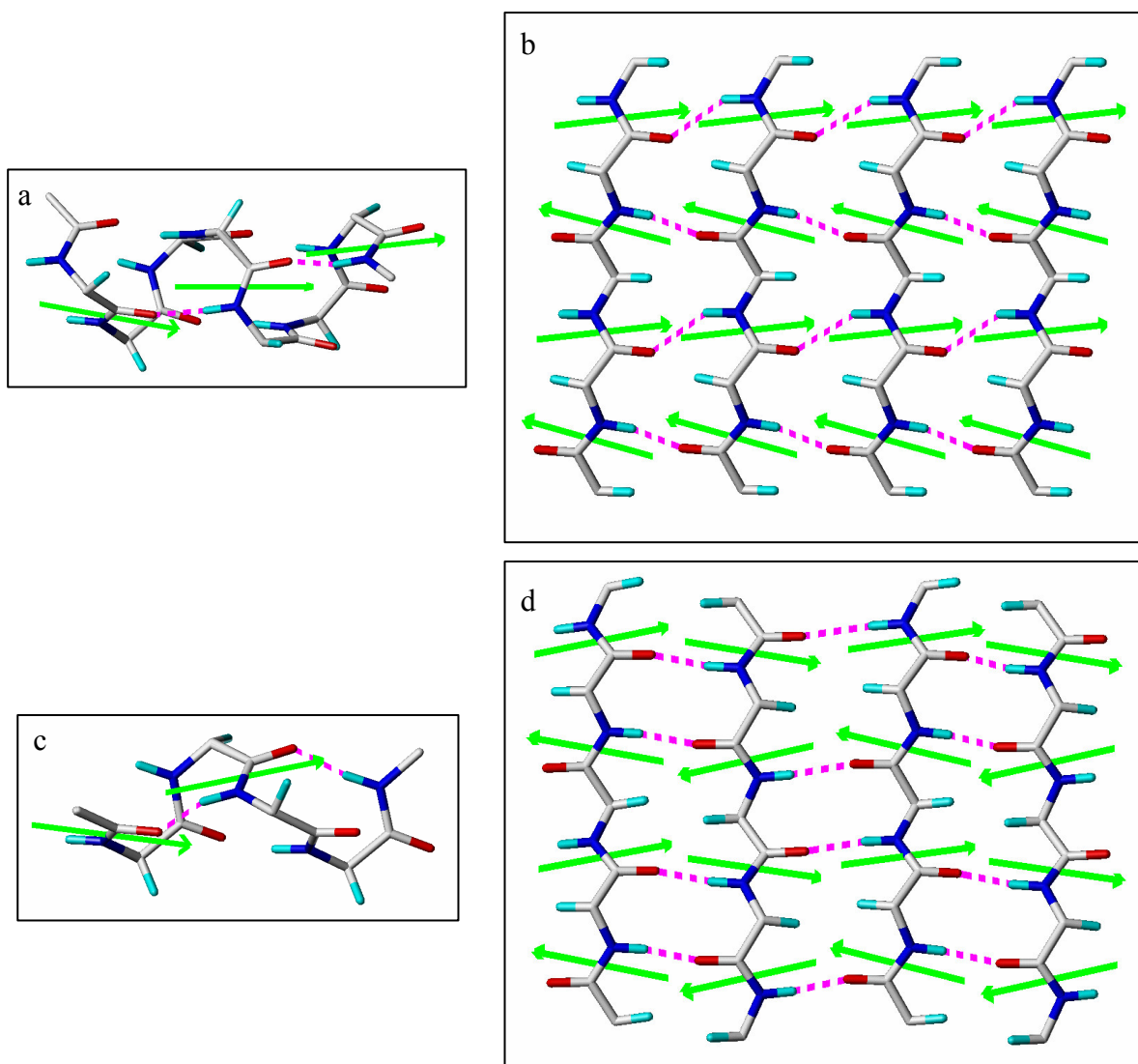


Figure 5.5: Hydrogen bonds (dashed lines) along the $C=O \cdots H-N$ arrays and peptide dipole vectors (arrows) in an idealized α helix (a), parallel β sheet (b), 3_{10} helix (c) and antiparallel β sheet (d). In both helices all dipole vectors orient in the same direction, whereas in the sheets alternating rows of piling vectors are given. Interestingly in α helices and along a row in parallel β sheets the bond dipole vectors and the $C=O \cdots H-N$ adopt similar directionality, whereas in 3_{10} helices and in a row of antiparallel β sheets an inclination across the accumulating bond dipole vectors and the H -bonded $C=O \cdots H-N$ arrays seems to be given. Possibly this alignment difference is responsible for the production of cooperative polarizing effects in α helices and parallel β sheets, leading to a systematic shortening of H -bond distances with increasing size of the secondary structure element (images created with SYBYL 7.0).

As discussed, in α helices the close directional alignment of bond dipole vectors and H -bonded $C=O \cdots H-N$ arrays serves as explanation for the experienced cooperative effects resulting in the decreasing H -bond distances. Also in 3_{10} helices all bond dipole vectors align

in the same direction, however the significant inclination between individual bond dipoles and the C=O \cdots H-N array direction obviously disrupts cooperative effects at least to a larger extend. Instead a gradual expansion of the H-bond distances is detected.

If similar properties hold for β sheets, the following can be suggested. Obviously, the alternating orientation of piling H-bonds and bond dipole vectors perpendicular to the stretch of the strands is not detrimental to the creation of cooperative polarizing effects across parallel sheets. As major difference between parallel and antiparallel sheets remains, as in the case of 3_{10} - and α helices, the deviating inclination angle between bond dipole vectors and H-bonded C=O \cdots H-N arrays. They align well (as in α helices) in parallel β sheets, and deviate in antiparallel ones (Figure 5 b,d), in particular if in the latter a mutual strand shear is given (Ho 2002).

Hopefully, this systematic analysis of hydrogen-bond length variations in secondary structures as evident from crystal data (resolution <1.5 Å) assists to improve computational tools to correctly reproduce experimental information. Understanding the nature of the secondary structure stabilisation is essential for the theory of protein folding and has to be considered in protein structure prediction. As shown by mutagenesis studies, the placement of correctly charged residues at the capping ends of extended helices takes significant impact on the stability of proteins. Furthermore, the apical position at the helical terminus exhibits a favorable binding site for charged groups in ligands bound to proteins (Hol 1985) and stabilizes pKa- shifts of amino acids involved in catalysis (Calabrese 2004). Furthermore ion channels such as the potassium or chloride channel exploit the properties of extended helices to achieve their unique function (Dutzler 2002; Jiang 2003) These are further indications for strong polarizing effects piled up along α helices.

In addition, the hydrogen bond properties of the other secondary structure elements and the putative conclusion should be focus for further theoretical studies.

6 Turn Classification: Existing turn families and types

6.1 δ -turn

The δ -turn family is the smallest of the different turn families and, in contrast to the well-known β and α -turn families, the intraturn hydrogen bond is formed between the main-chain NH_i and CO_{i+1} . A putative occurrence of the δ -turn was first described by Toniolo (Toniolo 1980) and later confirmed by theoretical conformational studies on three linked peptide units containing a cis-Pro (Table 6.1). In this work Nagarajaram et al. (Nagarajaram 1992) concluded that a 1 \rightarrow 2 hydrogen bond is generally possible for tripeptides with a cis peptide at the second position. They described possible dihedral angles and two conformations in cyclic pentapeptides with a weak 1 \rightarrow 2 hydrogen bond.

To the best of our knowledge, in protein structures δ -turns are not yet described.

tripeptide	ϕ_i	ψ_i	ϕ_{i+1}	ψ_{i+1}	type
Gly-cis-Pro	71	76	-71	-42	1
L-Ala-cis-Pro	63	76	-72	-43	1
Gly-cis-Pro	-70	-48	-59	-49	2
L-Ala-cis-Pro	-74	-40	-60	-46	2
Gly-cis-Pro	162	84	-66	128	3
L-Ala-cis-Pro	-146	58	-84	170	4

Table 6.1: Torsion angles for tripeptides with a 1 \rightarrow 2 hydrogen bond (Nagarajaram 1992)

6.2 ϵ -turn

The ϵ -turns were only theoretically proposed by Toniolo (Toniolo 1980). Here, they form a ‘reverse’ hydrogen bond between the main-chain NH_i and CO_{i+2} , which is comparable to the δ -turns. To the best of our knowledge, ϵ -turns have not been detected in proteins or peptides or during theoretical conformational analysis.

6.3 γ -turn

The γ -turn consists of three consecutive residues with an intraturn hydrogen bond formed between the main-chain CO_i and NH_{i+2} and was first described by Nemethy and Printz (Némethy 1972) and Matthews (Matthews 1972). Toniolo (Toniolo 1980) characterised two

different peptide conformations as γ -turns, an equatorial and an axial conformation, whereas Rose et al. (Rose 1985) named the two mentioned conformations as classic and inverse γ -turn with specific dihedral angle ranges (Table 6.2). They also note the importance of hydrogen bonding for the stability of γ -turns in peptides which is supported by the conclusion of Baker and Hubbard (Baker 1984) who characterised the hydrogen bonding in γ -turns of proteins. They pointed out that the hydrogen bonds are highly non-linear but still with an acceptable geometry for a stabilising effect. Additionally, Guruprasad et al. (Guruprasad 2000) described average values for the two conformations.

	ϕ_{i+1}	ψ_{i+1}	author
equatorial	-75°	50°	Toniolo (Toniolo 1980)
axial	75°	-50°	
'classic'	70° to 85°	-60° to -70°	Rose et al. (Rose 1985)
'inverse'	-70° to -85°	60° to 70°	
'classic'	$\bar{\phi} 70,97^\circ$	$\bar{\psi} -57,63^\circ$	Guruprasad et al. (Guruprasad 2000)
'inverse'	$\bar{\phi} -82,92^\circ$	$\bar{\psi} 75,59^\circ$	

Table 6.2: Specified torsion angles for γ -turn types

Later analyses (Milner-White 1988; Milner-White 1990) showed that classic γ -turns are mostly found at the top ends of β -hairpins which connect two antiparallel strands. In contrast, inverse γ -turns are normally not responsible for a polypeptide chain reversal but lead to a kink in the chain. Inverse γ -turns with a weak hydrogen bond are found in the middle of a β -strand. On contrast, inverse γ -turns with a strong hydrogen bonding occur at the end of β -strands, inducing the β -strand termination, or at the end of helices.

Consecutive inverse γ -turns, often situated within β -strands, were described as a 2.2_7 helix or as a "compound gamma turn" (Milner-White 1990). They support the hydrogen bonding network within β -strands, rather than forming a unique secondary structure. Studies on the tripeptide Ala-Phe-Ala (Motta 2005) showed that, depending on the temperature, the preferred conformation is an inverse γ -turn or a β -strand-like conformation. This observation seems to confirm the postulation of Milner-White (Milner-White 1990) that compound γ -turns could be intermediates of β -strands during folding into a β -sheets and that temperature is a regulatory possibility.

In addition to structural functions, γ -turns could also play a role within protein function. Aspartate 102, which is part of the catalytic triad in serine proteases such as trypsinogen and chymotrypsin, is residue $i+1$ of a strong γ -turn (Milner-White 1990). Additionally, γ -turns could have important biological functions within small, endogenous peptides (Motta 2005;

Tyndall 2005), e.g. vasopressin and the related desmopressin, bradykinin and angiotensin II show γ -turn conformation.

6.4 β -turn

The β -turn is the most analysed turn family; it consists of four residues with an intraturn hydrogen bond between the main-chain carbonyl group of the first residue and the main-chain amide group of the fourth residue. Theoretical calculations on three linked peptide units (Venkatachalam 1968) were originally used to classify β -turns into 6 different types (Table 6.3) based on the dihedral angles of the inner residues ($\phi_2, \psi_2, \phi_3, \psi_3$). Type I and II show a very similar non-helical conformation, only a peptide flip of the C=O direction of the second peptide unit occurs. In contrast, type III describes a helical turn conformation which could be part of a 3_{10} helix.

	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
I	-60°	-30°	-90°	0°
I'	60°	30°	90°	0°
II	-60°	120°	80°	0°
II'	60°	-120°	-80°	0°
III	-60°	-30°	-60°	-30°
III'	60°	30°	60°	30°

Table 6.3: Torsion angles for β -turns described by Venkatachalam (Venkatachalam 1968) (I, II, III, and the inverse conformations I', II', III')

β -turn classification has been subject to frequent changes. Lewis et al. (Lewis 1973) extended the number of turn-types up to 11 (I/I', II/II', III/III', IV, V/V', VI, VII; Table 6.4), whereas type VII describes more a kink than a real chain reversal. In contrast to Venkatachalams definitions, the latter authors used a distance criterion to identify the turn conformation in protein structures. The distance between $C\alpha_i$ and $C\alpha_{i+3}$ had to be smaller than 7\AA . Only 57% of the observed turn conformations showed a $4\rightarrow 1$ hydrogen bond and together with some energy minimisation studies, which also showed a lack of hydrogen bonding, they concluded that a $4\rightarrow 1$ backbone hydrogen bond is not necessary for the existence of a β -turn conformation. This turn definition was also used by Chou and Fasman (Chou 1977) who identified 459 turns in 29 protein structures with an average frequency of 32% as compared to helices (38 %) and β -sheets (20%).

	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}	Ψ_{i+2}
I	-60°	-30°	-90°	0°
I'	60°	30°	90°	0°
II	-60°	120°	80°	0°
II'	60°	-120°	-80°	0°
III	-60°	-30°	-60°	-30°
III'	60°	30°	60°	30°
IV	types I up to III' with two or more angles differing by at least 40°			
V	-80°	80°	80°	-80°
V'	80°	-80°	-80°	80°
VI	cis Pro at position $i+2$			
VII	kink: $\psi_{i+1} \approx 180^\circ$ and $ \phi_{i+2} < 60^\circ$ or $ \psi_{i+1} < 60^\circ$ and $\phi_{i+2} \approx 180^\circ$			

Table 6.4: Torsion angles for β -turns described by Lewis et al. (Lewis 1973)

Hydrogen bonded turns of type II/II' were divided by Némethy et al. (Nemethy 1980) into II₁/II₁' and II₃/II₃' based on the orientation of the third peptide group differentiating a negative or positive value for ψ_{i+2} .

	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}	Ψ_{i+2}
I [I']	< 0	< 0	< 0	> 0
II1 [II1']	< 0	> 0	> 0	> 0
II3 [II3']	< 0	> 0	> 0	< 0
III [III']	< 0	< 0	< 0	< 0

Table 6.5: Torsion angles for β -turn types described by Nemethy et al. (Nemethy 1980)

Later Richardson (Richardson 1981) proposed to include type III within type I, because the only difference is the involvement of type III turns in 3_{10} helical structures with similar dihedral angles. Richardson also introduced a new turn-type Ib, with some showing an overall “L” shape. Additionally, turn-type V/V' and VII were neglected since no examples for type V/V' were found and based on the definition by Lewis et al. (Lewis 1973) turn type VII conformations vary greatly with questionable turn conformations. Furthermore, turn-type VI is divided into VIa/VIb based on the conformation of proline and the existence of a hydrogen bond. This results in 8 different turn-types: I/I', Ib, II/II', VIa/VIb and IV.

Rose et al. (Rose 1985) used a combination of Venkatachalam and Richardson proposed turn-type definitions: I/I', II/II', III/III', VIa/VIb and introduced the term ‘open’ for β -turn conformations without a hydrogen bond. They also point out the importance of visual turn evaluation for turns with a C α distance between the first and last residue bigger than 7Å.

Wilmot and Thornton (Wilmot 1988) used the turn-type definitions of Richardson (Richardson 1981), but they also added a new turn-type VIII (see Table 6.6) for turns in

which the central residues ($i+1$, $i+2$) adopt an $\alpha_R\beta$ conformation, which means residue $_{i+1}$ lies in the right-handed helical region and residue $_{i+2}$ in the β region of the Ramachandran plot. They also recognized that almost half the turns show distorted conformations which do not fall within 30° next to the described standard angles. Therefore, they suggest that β -turn type I', which seems to be disfavoured due to steric hinderance, must be stabilized by additional hydrogen bonds in a β -sheet.

	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}	Ψ_{i+2}
I [I']	-60 [60]	-30[30]	-90[-90]	0 [0]
II [II']	-60 [60]	120 [-120]	80 [-80]	0 [0]
VIa	-60	120	-90	0
VIb	-120	120	-60	0
VIII	-60	-30	-120	120

Table 6.6: Torsion angles for β -turns described by Wilmot and Thornton (Wilmot 1988)

A completely new nomenclature was later proposed by Wilmot and Thornton (Wilmot 1990) based on the dihedral angles of the inner residues and the populated regions in the Ramachandran plot, because 42 % of the described β -turns did not fit into the eight accepted turn types (Wilmot 1988) within a range of $\pm 30^\circ$ around the standard angles. This led to 16 different β -turn types: $\alpha\alpha$, $\beta_P\gamma$, $\alpha\beta_E$, $\beta_E\gamma$, $\gamma\gamma$, $\gamma\alpha$, $\epsilon\alpha$, $\alpha\gamma$, $\beta\alpha$, $\gamma\beta_E$, $\beta_E\epsilon$, $\epsilon\beta_P$, $\epsilon\gamma$, $\alpha\epsilon$, $\beta_P\beta_P$, $\gamma\epsilon$ and 2 types for turns containing cis-proline: $\beta\alpha_R^{\text{cis-pro}}$ $\beta_E\beta_P^{\text{cis-pro}}$ (see paragraph 6.7.3 for further information about the regions)

A conformational study (Nagarajaram 1992) on a system of three linked peptide units with a trans-cis-trans configuration revealed a new hydrogen bonded turn conformation (ϕ_{i+1} : $\sim 10^\circ$, ψ_{i+1} : $\sim 90^\circ$, ϕ_{i+2} : $\sim -69^\circ$, ψ_{i+2} : $\sim -30^\circ$) compared to the already known VIa/VIb turn-type conformations described by Richardson (Richardson 1981). Additionally, an analysis of protein structures displayed five examples in which the cis-peptide unit is a non-proline.

Hutchinson and Thornton (Hutchinson 1994) refined the analysis based on a non-homologous data set; they divided the turn-type VIa into VIa1 and VIa2 and created new dihedral angle values for all used turn-types (I/I', II/II', VIa1, VIa2, VIb, VIII, IV). 43% were classified as the miscellaneous type IV using the criteria of Lewis et al. (Lewis 1973). They also identified 58% of the turns as multiple turns, in which at least one residue occurred in two different turns, with double turns containing 5 residues occurring most frequently.

Chou et al. (Chou 1997) mentioned again the importance of cis-proline in turn types VIa1, VIa2, VIb otherwise the $C\alpha_i-C\alpha_{i+2}$ distance is bigger than 7 Å and the conformation is dissimilar to a turn conformation. Additionally, they described turn-type IV as not being not

an authentic β -turn type because the $C\alpha$ distance for the representative of this turn is 7.15 Å when based on the average angle derived by Hutchinson and Thornton (Hutchinson 1994).

	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}	Ψ_{i+2}	
I [I']	-60 [60]	-30 [30]	-90 [90]	0 [0]	
II [II']	-60 [60]	120 [-120]	80 [-80]	0 [0]	
VIa1	-60	120	-90	0	i+2: cis-Proline
VIa2	-120	120	-60	0	i+2: cis-Proline
VIb	-135	135	-75	160	
VIII	-60	-30	-120	120	
IV	all conformations that do not fit the other turn types				

Table 6.7: Widely accepted β -turns definitions described by Hutchinson and Thornton (Hutchinson 1994)

Later analyses (Chou 2000; Guruprasad 2000; Pal 2002; Fuchs 2005) also used the refined classification of 9 different turn-types derived by Hutchinson and Thornton (Hutchinson 1994), so it can be safely assumed that this is the most widely accepted turn type definition for β -turns.

Recent analysis (Panasik 2005) showed the existence of geometrically identified non-hydrogen bonded β -turns within protein structures deposited in the PDB that could be minimized into hydrogen-bonded β -turns. After minimisation, these turn conformations hardly deviate from the original structure. More precisely, 87% of these geometrically identified open β -turns could be reclassified into hydrogen-bonded β -turns.

A recent review mentioned the importance of β -turns (Brakch 2006). From functional point of view, they are known to play a central role as molecular recognition elements in protein-protein interfaces, in antibody-antigen recognition, and in recognition sites for post-translational modification of inactive precursors.

Additionally, β -turns are involved in interactions between ligands and receptors, since β -turns could present their sidechains in an optimal conformation for biological activity. Several studies have shown (Rose 1985; Kee 2003; Brakch 2006), that small, endogenous peptides with specific functions, e.g. hormones or neurotransmitter, adopt β -turns in bioactive conformations. As example for endogenous peptides with β -turn conformation enkephalin, somatostatin, melaconyte-stimulating hormone (α -MSH), calcitonin related peptide (CGRP), neurotensin (NT), tachykinin peptide family and others are described. This knowledge was exploited in the design and synthesis of modified peptides which mimic β -turn structures that are used as drugs or antigens (Kee 2003; Brakch 2006) e.g. for somastatin analogues. This was successful, since ligands with a β -turn conformation represents a more effective

conformer in receptor-peptide interactions. This was also confirmed for some anticancer drugs where the potent inhibitors show β -turn conformations.

An example of the functional role of β -turns at protein-protein interfaces is the interaction between nerve growth factor (NGF) and its transmembrane tyrosine kinase receptor TrkA. The critical hotspot for the interaction involves various NGF β -turn regions. These are targeted as models for the synthesis of β -turn mimetics (Kee 2003). Furthermore, the interaction between CD4, a cell-surface glycoprotein that is found on T lymphocytes, and the HIV viral envelope glycoprotein gp120 depends on a surface-exposed β -turn of CD4 (Chen 1992). It has been shown that a specific residue within this β -turn has a significant role in binding. Furthermore, binding affinity is dependent on the turn conformation, peptides derived from this turn without the secondary structure did not show significant affinity (Chen 1992).

6.5 α -turn

One of the first detailed analysis (Nataraj 1995) based on a non-redundant dataset with a sequence homology smaller 40%, showed that the majority of α -turns are part of an alpha-helix (96%). The rest of the turns were divided into two major and seven minor groups, which are named after the region in the Ramachandran plot that the torsion angles of the inner residues belongs to (Table 6.8). The largest family (2%) have alpha-helical-like dihedral angles and are isolated alpha helix turns with opposing amino acid propensities when compared to alpha-helices. Additionally, this group can be built up from multiple β -turns.

name	group	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}	ϕ_{i+3}	ψ_{i+3}
$\alpha_R\alpha_R\alpha_R$	F1	-65	-25	-75	-35	-100	-20
$E\alpha_L\alpha_R$	F2	-60	130	75	0	-115	-50
$E\alpha_L\alpha_L$	g1	-65	155	70	30	60	40
$\alpha_LX\alpha_R$	g2	0	95	70	-15	-40	25
$E'\alpha_R\alpha_R$	g3	60	-145	-75	-25	-80	-40
$E'P\alpha_L$	g4	60	-140	-95	75	60	40
α_RPX	g5	-50	-60	-70	105	120	-50
$\alpha_L\alpha_L\alpha_R$	g6	50	30	60	10	-120	-30
$EX\alpha_R$	g7	-65	100	150	-60	-80	-30

Table 6.8: α -turn types described by Nataraj et al. (Nataraj 1995)

A second analysis was published which uses the whole PDB after removing similar turns and turns in α helices. Based on the signs of inner ϕ angles eight groups were defined with an additional group for cis peptides. The conformations were named as “X- α_{YZ} turns” (Table 6.9) and the mean values with standard deviations were given (Table 6.10). They describe the occurrence of β -turns within the α -turns between residues i and $i+3$ and residues $i+1$ and $i+4$, respectively, with a possible hydrogen bond between these residues.

names	I- α_{RS}	I- α_{LS}	II- α_{RS}	II- α_{LS}	I- α_{LU}	I- α_{RU}	II- α_{LU}	II- α_{RU}	I- α_C
Φ_{i+1}	-	+	-	+	-	+	-	+	cis peptide
Φ_{i+2}	-	+	+	-	+	-	-	+	
Φ_{i+3}	-	+	-	+	+	-	+	-	

Table 6.9: Designation for α -turn types (X - α_{YZ}) described by Pavone et al. (Pavone 1996)

$X \rightarrow 'T'$ if $sign(\phi_{i+2}) = sign(\phi_{i+3})$, otherwise 'II'

$Y \rightarrow 'R'$ if $\phi_{i+3} < 0^\circ$ (right-handed chain reversal), otherwise 'L' (left-handed)

$Z \rightarrow 'S'$ if $sign(\phi_{i+1}) = sign(\phi_{i+3})$ (screw-like shape), otherwise 'U' (U shape)

names	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}	Ψ_{i+2}	Φ_{i+3}	Ψ_{i+3}
I- α_{RS}	-60 ± 11	-29 ± 13	-72 ± 14	-29 ± 15	-96 ± 20	-20 ± 17
I- α_{LS}	48 ± 22	42 ± 14	67 ± 9	33 ± 14	70 ± 11	32 ± 12
II- α_{RS}	-59 ± 10	129 ± 15	88 ± 15	-16 ± 19	-91 ± 22	-32 ± 18
II- α_{LS}	53 ± 15	-137 ± 25	-95 ± 12	81 ± 23	57 ± 5	38 ± 8
I- α_{LU}	-61 ± 12	158 ± 15	64 ± 17	37 ± 21	62 ± 12	39 ± 9
I- α_{RU}	59 ± 18	-157 ± 31	-67 ± 17	-29 ± 20	-68 ± 12	-39 ± 12
II- α_{LU}	-65 ± 15	-20 ± 15	-90 ± 17	16 ± 44	86 ± 18	37 ± 27
II- α_{RU}	54 ± 8	39 ± 15	67 ± 13	-5 ± 31	-125 ± 11	-34 ± 32
I- α_C	-103 ± 23	143 ± 4	-85 ± 8	2 ± 6	-54 ± 6	-39 ± 9

Table 6.10: Mean torsion angles for α -turn types described by Pavone et al. (Pavone 1996)

Ramakrishnan et al. (Ramakrishnan 1998) carried out an energy minimization study on glycol-tetra peptides and described thirteen different conformational classes along with corresponding inverse classes. Not all conformational classes were found in proteins and the conformational overlap between them could vary. The missing classes were not observed evidently due to side chains being present in protein structures. They give examples for twelve of these energy minimized classes, whereas for two additional examples no accordant theoretical class exists.

In contrast to the analysis from Nataraj et al. (Nataraj 1995) and Pavone et al. (Pavone 1996), who used a hydrogen bond between the CO_i and NH_{i+4} as a criterion to identify α -turns, Dasgupta et al. (Dasgupta 2004) employed a classification that defined α -turns as being present when the distance between Ca_i and Ca_{i+4} was less than 6.5 \AA . They identify 15 different α -turn types and showed that many of these turn-types contain multiple β -turns; 38% of major α -turns are constructed of multiple β -turns.

designation	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}	Ψ_{i+2}	Φ_{i+3}	Ψ_{i+3}
AAA	-67 ± 15	-30 ± 16	-78 ± 16	-33 ± 17	-103 ± 21	-17 ± 21
PAA	-79 ± 13	168 ± 14	-63 ± 11	-24 ± 14	-88 ± 15	-2 ± 17
AAa	-65 ± 13	-23 ± 13	-93 ± 14	5 ± 11	83 ± 16	11 ± 19
AAE	-78 ± 20	-29 ± 20	-102 ± 20	-25 ± 21	-142 ± 21	142 ± 28
Bra	-136 ± 14	126 ± 16	53 ± 6	41 ± 8	77 ± 11	6 ± 16
BAA	-140 ± 16	179 ± 13	-63 ± 9	-25 ± 14	-96 ± 19	-10 ± 20
PPa	-76 ± 13	161 ± 13	-55 ± 5	133 ± 7	85 ± 14	0 ± 30
AAD	-71 ± 18	-31 ± 14	-89 ± 17	-25 ± 20	-134 ± 17	72 ± 11
rgE	53 ± 5	43 ± 9	84 ± 8	-3 ± 9	-115 ± 15	145 ± 19
BpA	-135 ± 15	121 ± 25	61 ± 7	-125 ± 7	-95 ± 12	8 ± 14
pAE	65 ± 11	-125 ± 8	-94 ± 13	6 ± 15	-108 ± 15	142 ± 18
BPa	-140 ± 18	169 ± 14	-55 ± 9	130 ± 9	85 ± 16	-1 ± 18
AAe	-62 ± 9	-27 ± 13	-98 ± 17	4 ± 17	128 ± 29	180 ± 21
PgA	-54 ± 7	133 ± 9	87 ± 12	-11 ± 13	-125 ± 16	-16 ± 35
pAA	79 ± 18	-172 ± 30	-73 ± 17	-18 ± 18	-97 ± 20	-1 ± 16

Table 6.11: Designation and mean torsion angles for α -turn types described by Dasgupta et al.

(Dasgupta 2004)

'E': extended \rightarrow divided into 'B' (β -strands) and 'P' (polyprolinelike helices)

'A': helical \rightarrow 'R' (right-handed) and 'G' (C-terminal residues in 3_{10} helices)

'D': bridging region between 'E' and 'A'

Nataraj et al. (Nataraj 1995) and Pavone et al. (Pavone 1996) identified mostly hydrophilic residues within isolated hydrogen-bonded α -turns that are mostly exposed to the solvent. Pavone et al. (Pavone 1996) assumed a function as a key with a hook-like shape in lock-key interaction between proteins and they suggest examples of α -turns with a functional role in active site and metal ion coordination. For example, in 1COX α -turn residues are in close contact to the flavin adenine dinucleotide. Additionally, α -turns also occur in small peptides with biological activity, Ilamicin b1, which shows antibiotic activity contains the α -turn-type I- α_C . In contrast to the hydrogen-bonded α -turns, non-hydrogen-bonded α -turns (Dasgupta 2004) have more hydrophobic residues at position i and $i+4$, which could lead to hydrophobic interactions across the turn.

6.6 π -turns

The first detailed analysis of hydrogen bonded π -turns in proteins was described by Rajashankar and Ramakumar (Rajashankar 1996). The arising π -turns were grouped into four classes based on the conformation of the (i+4)th residue, and eleven overall subclasses, based on the position of the dihedral angles of the residues in the Ramachandran plot.

	π_{α_L} -turn	π_{α_R} -turn	π'_{α_L} -turn	π_{β} -turn
subclasses	$\alpha_R\alpha_R\gamma\alpha_L$	$\alpha_R\alpha_R\alpha_R\alpha_R$	$\alpha_L\alpha_L\alpha_L\alpha_R$	$\epsilon\beta\beta\beta$
	$\alpha_L\alpha_R\gamma\alpha_L$	$\beta\alpha_L\alpha_R\alpha_R$		$\gamma\beta\beta\beta$
	$\beta\alpha_L\gamma\alpha_L$	$\alpha_R\alpha_R\beta\alpha_R$		
	$\alpha_R\beta\gamma\alpha_L$	$\alpha_R\gamma\alpha_L\alpha_R$		

Table 6.12: Designation for π -turn types described by Rajashankar and Ramakumar (Rajashankar 1996)

Only the π_{α_R} -turns form consecutive π -turns. They are normally responsible for distortions in the middle or at the end of alpha helices and show different amino acid propensities compared to alpha helices. The $\alpha_R\alpha_R\alpha_R\alpha_R$ conformation is normally stabilized by $6 \rightarrow 1$ and $5 \rightarrow 1$ hydrogen bonds. The π_{α_L} -turn that occurs most frequently in peptide structures shows an additional $5 \rightarrow 2$ hydrogen bond, which is in general a type I β -turn within this π -turn. Previously, only the π_{α_L} turn was described at the C-terminus of alpha helices as the so-called Schellmann motif which terminates helix structures (Schellmann 1980; Baker 1984), but could also appear at the end of β -strands.

6.7 Additional information

6.7.1 Deviations from standard angles

Not only had the turn-type definitions varied over the years but also the acceptable deviations of the standard torsion angles. According to Lewis et al. (Lewis 1973) Chou and Fasman (Chou 1977) classified β -turns as ideal if none of the angles deviates more than $\pm 50^\circ$ from ideal values. Additionally, they accepted turns when a single angle deviates more than 50° as non-ideal. Rose et al. (Rose 1985) only included β -turns with an angle deviation of 30° from standard values. Wilmot and Thornton (Wilmot 1988) extended the deviation of $\pm 30^\circ$ to $\pm 45^\circ$.

6.7.2 Intermolecular hydrogen bonding

Toniolo (Toniolo 1980) specified an alternative notation for turns. In general, the direction of an hydrogen bond is described from donor to acceptor atom which means in the case of turns, an hydrogen bond between the N-H of an amino acid number m and C=O of a residue number n is indicated as $m \rightarrow n$. On the basis of the number of atoms in the ring formed by closing the hydrogen bond, the conformations are called C_{nr} conformations. This leads to alternative designations of the possible turns: C_7 or $3 \rightarrow 1$ hydrogen bond for a γ -turn, C_{10} or $4 \rightarrow 1$ hydrogen bond for β -turns, C_{13} or $5 \rightarrow 1$ for α -turns, C_{16} or $6 \rightarrow 1$ for π -turn, C_8 or $2 \rightarrow 3$ hydrogen bond for δ -turn, C_{11} or $2 \rightarrow 4$ for ϵ -turn, and C_{14} or $2 \rightarrow 5$, C_{17} or $2 \rightarrow 6$, C_{20} or $2 \rightarrow 7$ for the other reverse turn conformations.

6.7.3 Nomenclature based on position in Ramachandran plot

Various publications adapted a nomenclature that is based on specific regions in the Ramachandran plot. In order to obtain comparable results this designation was also used within this work to describe the new turn-types (see also 7.5 for more details).

Wilmot and Thornton (Wilmot 1990) used six different regions to classify the different β -turn types. In more detail, with β_E as an idealized β -strand conformation (-60° , 120°), β_P as polyproline conformation (-120° , 120°), ϵ as only accessible for glycine (120° , $\pm 180^\circ$), α_R for the right-handed α -helical region (-60° , -30°), γ_L and α_L for the left-handed α -helical region (90° , 30°).

Efimov (Efimov 1993) divided the Ramachandran plot into 8 different regions, α for the right-handed α -helical region, β_E and β_P for the β -region in the upper left corner, γ , δ , ε and α_L and γ_L for the left-handed alpha-helical region.

Nataraj et al. (Nataraj 1995) even choose a different designation for each individual position within α -turns. For position $i+1$ five different regions are named: E for the extended region, E' for the inverse extended region, α_R and α_L for both helical regions and B for a bridging region between E and α_R . At position $i+2$ four regions are marked: α_R and α_L , P for a polyproline region and X for some isolated points in the lower, right corner. The last position ($i+3$) has only a α_R and α_L region. Some outliers in the lower right corner were named as X.

Dasgupta et al. (Dasgupta 2004) used 10 different regions to designate α -turns: 'E', a extended conformation which is further divided into 'B' (β -strands) and 'P' (polyproline-like helices); 'A' for the α -helical region which is split into 'R' (right-handed) and 'G' (C-terminal residues in 3_{10} helices); 'D': bridging region between 'E' and 'A' and the symmetry related regions 'e' \rightarrow 'b' and 'p'; 'a' \rightarrow 'r' and 'g'; 'd'.

7 Turn Classification: Methods and Materials

The description of the PDB-Select list (section 2.1.2) and the amino acid propensity respectively the conformational parameter (section 2.3) are already described in the Secbase method and materials section.

7.1 Emergent Self-Organizing Maps (ESOM) and U-matrix

Self-Organizing Maps (SOM) (Kohonen 1997) are widely used for clustering high-dimensional data (Stahl 2000; Yan 2003; Kupas 2004) and could be associated to the artificial neural networks, as a subtype with a single layer network. The SOMs project high-dimensional data in a self-organized process onto a two-dimensional grid of neurons which preserves the neighbourhood relationships and the topological properties, respectively. The emergent self-organizing map (ESOM) is a variant of the self-organizing map with a larger amount of neurons

SOMs are built up from neighbored neurons on a two-dimensional grid that represents a vector of the high-dimensional space. This vector is called the weight vector and after training a neuron should represent an input vector of the high-dimensional data. During training the following algorithm is processed every epoch: For each input vector of the data space, a neuron with the smallest distance between the input vector and its weight vectors is identified, the so-called “best match”. The weight vector of this best match is then drawn closer to the used input vector. Furthermore, the weight vectors of the neurons in the neighbourhood are also drawn closer but with less emphasis. This process of SOM training adapts the weight vectors to the given input vectors and creates a two-dimensional projection.

Since the neighbourhood of neurons on the edges of planar maps is small (Figure 7.1 a), compared to the centre of the map, a two-dimensional toroidal grid with four direct neighbours for a neuron is used within this analysis where the edges of the grid are connected (Figure 7.1 b) to avoid border effects (Ultsch 2003).

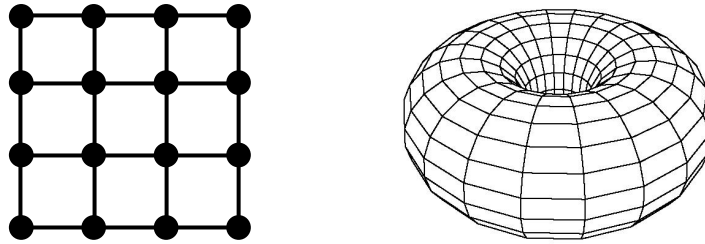


Figure 7.1 Neuron topology⁷ - a) four direct neighbours for each neuron (quadgrid)
b) toroidal topology without borders

For a detailed explanation of SOMs and artificial neural networks see (Kohonen 1997), since the algorithm could only be explained in general.

Normally, SOMs are used only with a few neurons as a kind of k-means clustering (Ultsch 1999). The number of neurons is equal to the number of expected clusters, so that all data that belong to a single neuron are a member of this cluster. In contrast, the Emergent SOMs (Ultsch 1999) have many more neurons than input points and the clusters are not detected by regarding single neurons, but by analysing the overall structure of the whole map. The map exhibits emergence, new properties ('cluster') could be detected based on the cooperation of elementary processes ('neurons'). A well-known example of emergence produced by humans is the so-called "Mexican Wave" (German: 'La-Ola Welle') in a stadium. The elementary processes are the humans, who stand up and throw the arms up in the air. Together, in a cooperative process, they produce a large wave, which rolls through the audience.

For displaying, the U-matrix (Ultsch 2003) is used, which visualizes local distances on each neuron between the high-dimensional data points as a 3D landscape of valleys and mountains. The U-height at each neuron is a sum of distances to the neighbouring weight vectors and is displayed as a coloured contour plot on top of the planar grid (see Figure 7.2). In combination with the underlying ESOM much information can be retrieved easily. For visualization, toroid maps are displayed as a planar map, where the neurons on the top are neighboured by the neurons on the bottom and the right side is neighboured by the left side (like a 'Pacman universe'). The position of the best matches reflects the topology of the input space, which means, that two adjacent data points are also neighboured in data space. Neurons with high U-height represent groups of dissimilar points in data space, whereas low values represent

⁷ <http://databionic-esom.sourceforge.net/user.html>

groups of points that are closely related. Thus, clusters are located within valleys on a U-matrix and the borders between the clusters are represented by mountains.

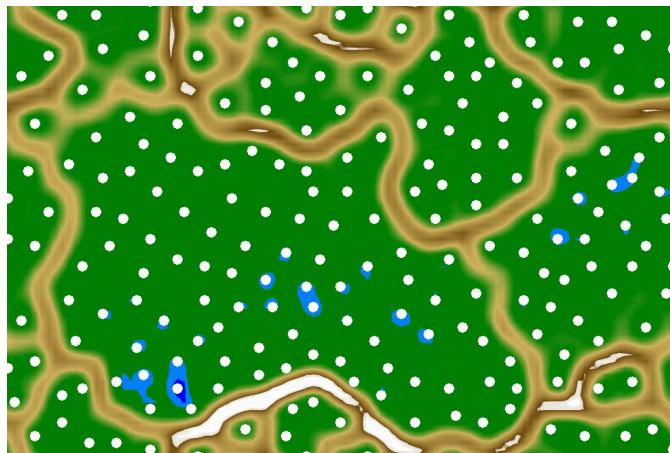


Figure 7.2: Trained ESOM with U-Matrix visualisation data points (“best matches”) are shown as white dots

For identifying structures that cross the edges of the maps a so-called tiled display is used, where the same map is drawn four times (Figure 7.3). For displaying the important features of a map during turn clustering a section of this tiled display is often shown in this work. Thus, the data points and structures could occur more than once (Figure 7.3).

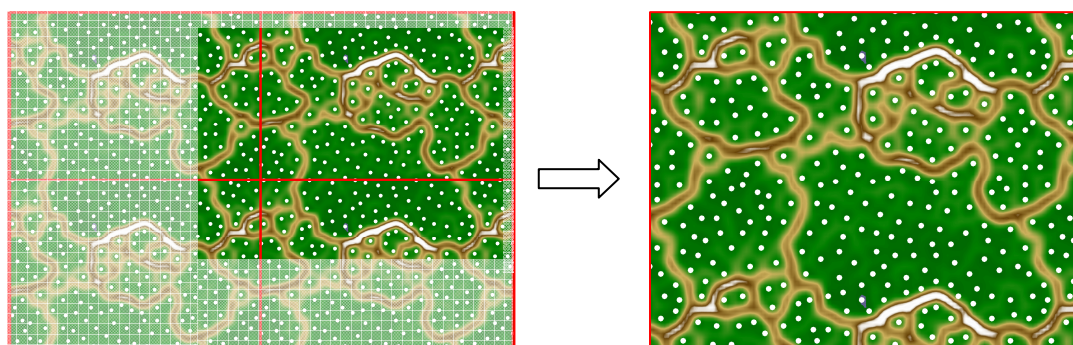


Figure 7.3: Tiled visualizing and the consequential cutout (“asymmetric unit”)

There are several reasons why Emergent Self Organizing Maps were used in this work. In combination with the U-matrix data points are grouped together into several clusters; outliers are easy to detect and can be ignored. Unsupervised, self-organizing learning is another advantage of this approach as information about the number of clusters is not needed before clustering (a priori). In contrast, the popular k-means algorithm always puts the whole data including outliers together into initially detected number of clusters. This makes outlier recognition more difficult. Furthermore, self-organizing maps were already used successfully in clustering structural motifs (Schuchhardt 1996) of the protein backbone or the principles of self-organizing maps were used to cluster so-called protein building blocks (de Brevern 2000). The biggest advantage of ESOM clustering is that the trained maps can be used to

classify other, new data points based on previously identified clusters. A tool is freely available for clustering, visualisation and classification: the Databionic ESOM Tools (Ultsch 2003) (<http://databionic-esom.sourceforge.net>).

The number of neurons that should be used for the ESOM are at most three times the number of datapoints but at least 1000 and this neurons were used to build maps with an amount of columns that is 1.5 times higher than the number of rows (personal communication: Prof. A. Ultsch and member of his group). Table 7.1 shows the number of used neurons for each turn family. Due to computational time, the number of neurons had to be decreased for the bigger dataset. The following training important parameter values were set: *epochs* = 200, *neighbourhood kernel function*: Gaussian, *weight initialisation method*: pca, *data space distance function* = Euclidean, *map grid*: 2-dimensional toroid and *start value for radius*: number of rows/ 2. For all other options, the standard values were used.

turn category	residues	turn structures	multiplicator	final number of neurons
reverse	2	194	30x	5820
reverse	3	114	30x	3420
reverse	4	1957	3x	5871
reverse	5	1340	3x	4020
reverse	6	954	3x	2862
normal	3	20162	1x	20162
normal	4	28650	1x	28650
normal	5	91726	0.5x	45863
normal	6	3995	3x	11985
open	4	137101	0.1x	13710
open	5	19607	1x	19607
open	6	21204	1x	21204

Table 7.1: Used number of neurons for each turn family (based on the number of turn structures)

7.2 Turn Data Collection

The processed data set was based on the PDB select list (see 2.1.2) and includes 1903 protein chains with a sequence homology smaller than 25%. Using Reliscript, possible turn structures were identified using their hydrogen bonding pattern based on the DSSP energy function or the C α -C α distance and divided into 16 different families based on the categories (normal, open, reverse) and the different turn classes (turns with similar lengths, see section 8.1 for definitions). The feature vector for clustering was a list of the torsion angle of the inner residues and after collecting all turn structures, Python programs were used for the calculation of the torsion angles and conversion into the correct file formats for clustering with ESOMs.

7.2.1 C α -C α distance cut-off

As described by Rose et al. (Rose 1985) the chosen cut-off for a C α -C α distance is always a compromise between over-inclusion and over-exclusion. Consequently, it is better to use visual inspection to include chain reversals with a larger separation. Chou and Fasman (Chou 1977) accepted 11 β -turns that did not fit their C α -C α distance criterion but were identified as chain reversals by visual inspections. Dasgupta et al. (Dasgupta 2004) remarked that there are hydrogen bonded α -turns with a C α -C α distance bigger than 6.5 Å, which they choose as a cut-off for their analysis.

To prevent exclusion of possible turn conformations a much larger cut-off of 10 Å for the C α -C α distance criterion was chosen in this work. All C α -atoms within a radius of 10 Å around the starting C α -atom were considered and the C α -atom with the smallest distance to the starting C α was chosen as the end of this turn conformation. Based on the number of residues involved in this open conformation, the turn was assigned to one of the open turn families. Clustered conformations were checked visually to identify open turns and eliminate other conformations. In general, bended conformations that indicates a chain reversal or helical conformations similar to helical hydrogen bonded turns were considered to be a turn. For critical conformations, the information about the C α -C α distance distribution was used to support the decision. If the border of the distribution is clearly within the 10 Å cut-off, these conformations are assigned as a turn cluster, since this indicates clearly separated turn conformations. Furthermore, some of these conformations that are clearly within the 10 Å cut-off are better described as a kinked or a hook-conformation. These conformations are named similar to separate them from the turns.

This special kind of data collection with respect to the hydrogen bonding pattern, $C\alpha$ -distance and number of turns differentiate this analysis from other studies that cluster local parts of the protein chains without considering the secondary structure elements (Schuchhardt 1996; de Brevern 2000; Hunter 2003; Sander 2006). These studies try to identify a small amount of common local motifs with a specific length. In contrast, this study identify all possible turn conformation based on certain criteria, which leads to a more precisely description of the local motifs of the protein chain.

7.2.2 Final data sets

For clustering the torsion angles of the inner residues were used as a feature vector which leads to the following data sets for the different turn families:

	designation	number of			torsion angles used in the feature vector				
		residues	torsion angles	turn structures	i	i+1	...	n-1	n
reverse	δ	2	5	210	φ_i	ψ_i	ω_i	φ_{i+1}	ψ_{i+1}
	ε	3	8	134	φ_i	ψ_i	...	φ_{i+2}	ψ_{i+2}
	-	4	11	1957	φ_i	ψ_i	...	φ_{i+3}	ψ_{i+3}
	-	5	14	1340	φ_i	ψ_i	...	φ_{i+4}	ψ_{i+4}
	-	6	17	954	φ_i	ψ_i	...	φ_{i+5}	ψ_{i+5}
normal	γ	3	4	20198	ω_i	φ_{i+1}		ψ_{i+1}	ω_{i+1}
	β	4	7	28718	ω_i	φ_{i+1}	...	ψ_{i+2}	ω_{i+2}
	α	5	10	91726	ω_i	φ_{i+1}	...	ψ_{i+3}	ω_{i+3}
	π	6	13	3994	ω_i	φ_{i+1}	...	ψ_{i+4}	ω_{i+5}
open	β	4	7	137101	ω_i	φ_{i+1}	...	ψ_{i+2}	ω_{i+2}
	α	5	10	19607	ω_i	φ_{i+1}	...	ψ_{i+3}	ω_{i+3}
	π	6	13	21204	ω_i	φ_{i+1}	...	ψ_{i+4}	ω_{i+5}

For an example picture and the used torsion angle for normal β -turns see Figure 1.4. The torsion angles of the other families are named in a similar way.

7.3 Data Preparation

7.3.1 ω transformation

The problem of using the torsion angles of the residues as the feature vector for clustering is the ω angle of the peptide bond and the distribution of the ω angle around $\pm 180^\circ$ and around 0° (Figure 7.4 a). Although an angle of $+179^\circ$ and -179° results in a similar conformation, the distance in Euclidean space that is used for clustering, is large (Figure 7.4 b). To avoid problems that occurred during clustering, 90° were added to the ω angle which leads to a distribution around -90° and $+90^\circ$ with a small distance between similar conformations (Figure 7.4 c).

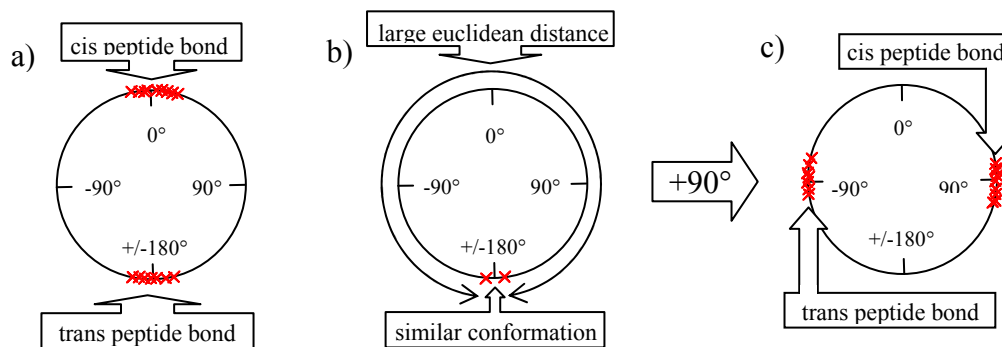


Figure 7.4: Transformation of ω angle

a) distribution of peptide bond ω angle

b) similar conformations with large Euclidean distance ($\pm 179^\circ$)

c) transformed ω angle distributions

Figure 7.5 shows the difference between clustering with (a) and without (b) the omega transformation. In (b) there are more clusters as the algorithm splits similar conformations into separate clusters due to the value of ω crossing the \pm boundary. In this case, both cluster (the small one in the middle and the data points outside this cluster) are divided into four independent clusters, because of two ω angles of γ -turns. Therefore, the result using the transformation is better, since it reduces the number of similar clusters.

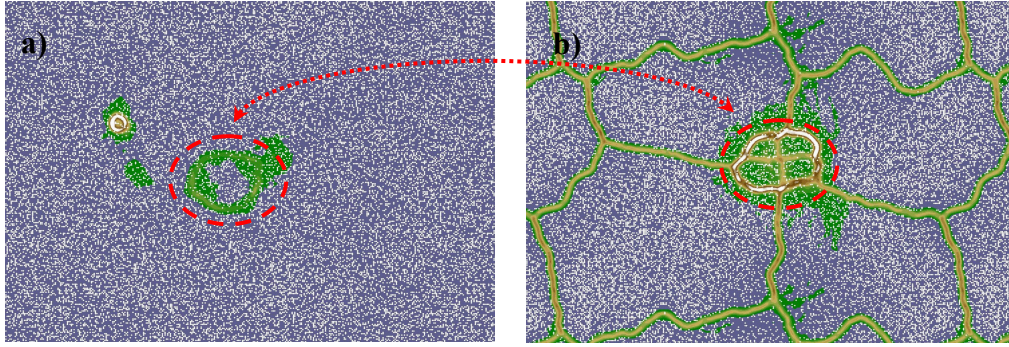


Figure 7.5: Clustering example of γ -turns with (a) and without (b) ω transformation

→ the small cluster and the remaining data points are each divided into four different clusters (combination of the two ω angle that occur within γ -turn feature vector)

In contrast, the distribution of the ψ/ϕ angles are not so clearly distributed among specific regions, so a transformation does not enhance the results. Similar conformations, that traversed the angular boundary of the ϕ/ψ angles and are lying in different clusters, were identified manually afterwards (see 7.4).

7.3.2 Z transformation

The z transformation was used as an essential pre-processing step before using ESOMs for clustering. Torsion angles are normalized so that the mean for the data range is 0 and the variance over that data range is 1. Accordingly, around 66% of the data of a specific torsion angle fall between -1 and 1. With this transformation, the absolute level and the corresponding deviation are standardized and therefore comparable.

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j(x_{ij})} \quad \text{with} \quad s_j(x_{ij}) = \sqrt{\frac{\sum_i (x_{ij} - \bar{x}_j)^2}{n-1}}$$

i = the feature vector of turn structure i

j = one torsion angle j among all turn structures

n = number of turn structures

\bar{x}_j = mean value for one torsion angle j

7.4 Cluster assignment and evaluation

Applying the U-matrix visualisation, the data points have to be assigned manually to a specific cluster and this can be done within the ESOM-tools. Nevertheless, this can be a very time consuming step that greatly depends on the number of data points and the size of the ESOM. For example, the ESOM (Figure 7.6) for the small ε -turn dataset (see 8.3.1) shows clear separated clusters and the different clusters and cluster borders can be identified easily. So, data points can be assigned to clusters easily.

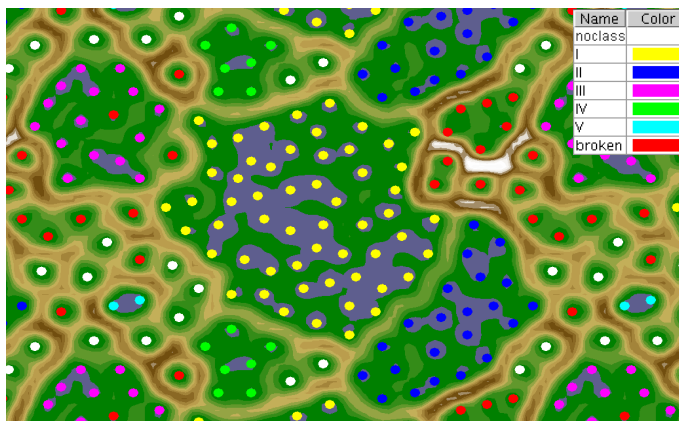


Figure 7.6: ESOM clustering for ε turns

In contrast, bigger datasets produce ESOMs that are more difficult to interpret. In particular, the border regions exhibit turn conformations where differentiation is very difficult between turn conformations that belong to a specific cluster or not.

In addition, large regions within U-matrix borders can also belong to two different conformations with a large number of consistent torsion angle ranges. Figure 7.7 shows the torsion angle distribution of the open α -turn types IXa and X (see 8.5.2). The distributions of all torsion angles are located in similar regions except for ψ_{i+1} . Since nine of the possible torsion angle shows similar values, the distance in high-dimensional data space between these two conformations is rather small. The U-matrix for the whole feature vector does not show a significant border between these two clusters (Figure 7.8 a). To resolve such problems, U-matrix visualisation can be used with subsets of features of the whole feature vector as the distance metric. Figure 7.8 b shows the

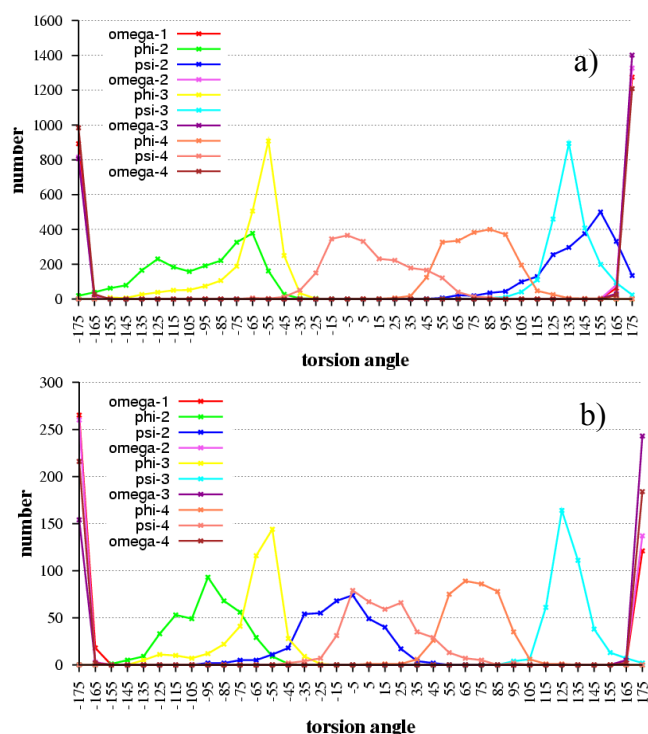


Figure 7.7: Torsion angle distribution for open α turn-type a) IXa and b) X

same ESOM as in Figure 7.8 a, but with a U-matrix distance visualisation that is based only on the ψ_{i+1} torsion angle. Cluster borders can be identified easily. One can see that turn-type IXa and X belong to different types.

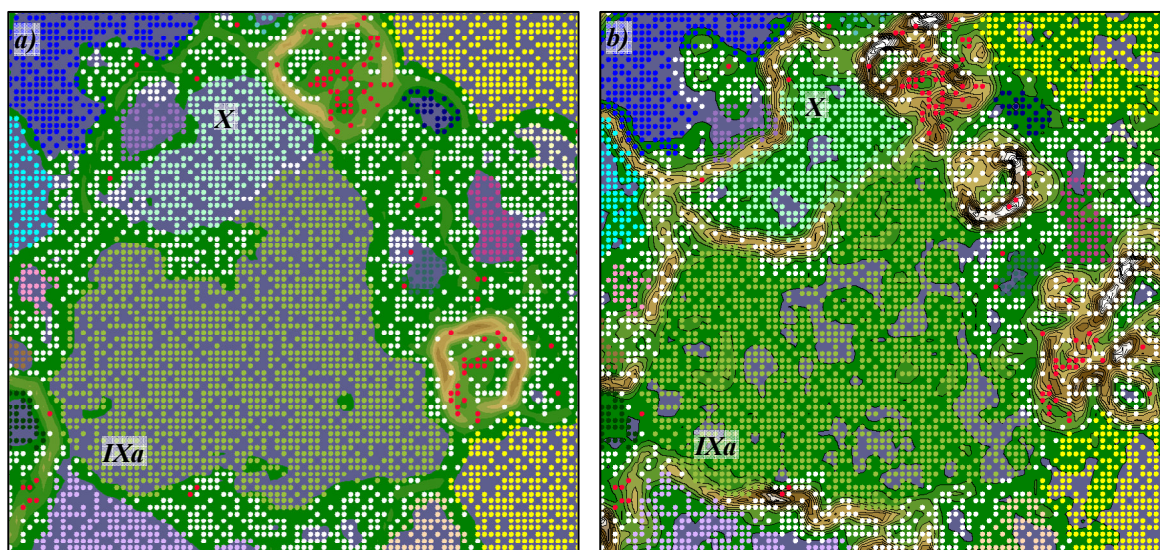


Figure 7.8: Cutout of ESOM clustering for open α -turn types IXa/X with different U-Matrix distances
a) whole feature vector, b) only ψ_{i+1}

Furthermore, clusters exist with similar conformations that show a torsion angle around $\pm 180^\circ$. For example β -turn-type VIIIa and VIIIb differ only within the ψ_{i+2} (ψ_{i-3}) distribution (Figure 7.9). In fact, they describe the same conformation, but a single torsion angle crosses the $\pm 180^\circ$ boundary leading to two different clusters (see 7.3.1 for a detailed explanation). Therefore, they are named with the same Roman numeral and small alphabetic letter to identify similarity.

Data points that do not belong to a specific cluster are shown in white and they belong normally to "cluster" 0 and are designated as 'noclass'.

Turns are mentioned as an exception that lack atom coordinates, because of missing electron density. These structures with a 'broken chain' are reassembled, named as 'broken' and shown in red (see Figure 7.6). They were neither considered for any further analysis nor assigned to a pdb structure.

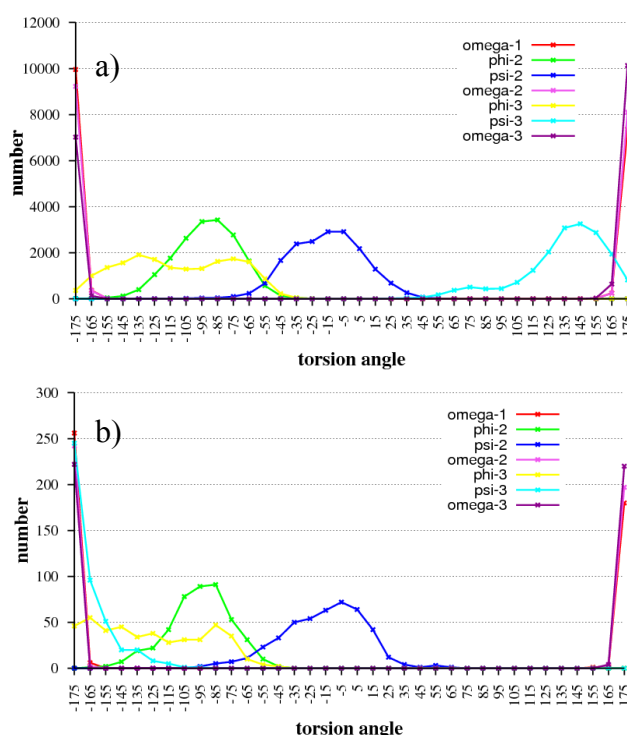


Figure 7.9: Torsion angle distribution for open β -turn type a) VIIIa and b) VIIIb

For these reasons a Pymol extension (Figure 7.10) was written that supports the evaluation of each cluster and the identification of the described problematic turn conformations. It automatically processes the ESOM tool output, combines the specific cluster data, calculates the mean and standard deviation of the different torsion angles and visualizes the data. Additionally, the extension retrieves turn conformations from Relibase using Reliscript, aligns and visualizes them using Pymol.

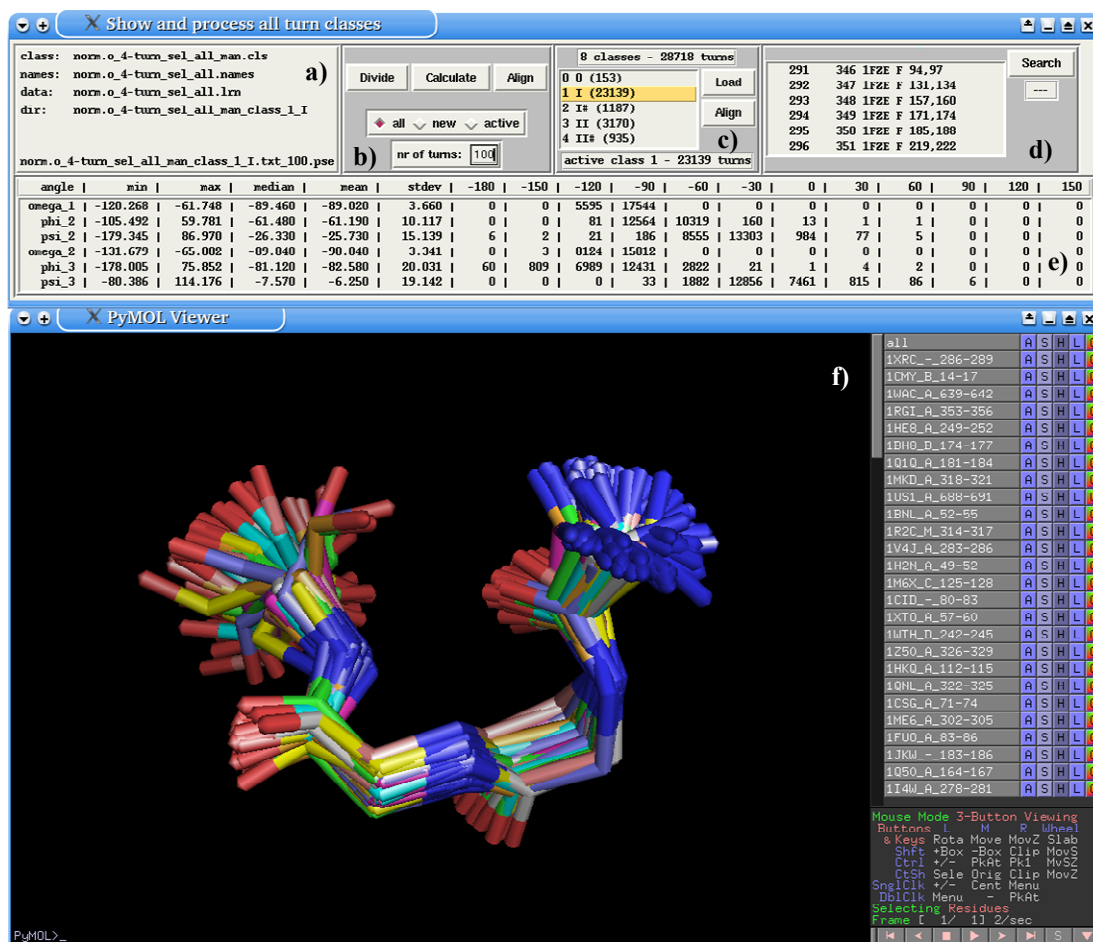


Figure 7.10: Turn Viewer - A Pymol extension for visualisation and evaluation of turn types

- processes ESOM tool output files
 - calculates statistics including mean, standard deviation and a simple histogram
 - retrieves turn conformations from Relibase
 - shows aligned turn conformation of a specific turn-type
- information about used files (cls: ESOM file with specified clusters; names: ESOM file with corresponding pdb names and residues; lrn: ESOM file with torsion angle feature vector; pse: Pymol session file with aligned turns)
 - action window (divide: divide ESOM files and create files with data about specific turn-types; calculate: calculate statistics; align: randomly align a number of turn structures that is specified in the field 'nr of turns')
 - information about the different turn-types (load: load aligned structures of marked turn-type into Pymol; align: align turn structures of marked turn-type)
 - information about turns of currently shown turn-type
 - statistics about turn-type that is currently shown
 - pymol window for visualisation of aligned turn structures

7.5 Conformation of amino acids in protein

The Ramachandran plot is used to describe the main chain conformations of amino acids in proteins and peptides. It was first proposed by Ramachandran & Sasisekharan (Ramachandran 1968). The main chain conformation is described by the two torsion angles ϕ , ψ which are plotted against each other. In general, the different regions within a Ramachandran plot are named after the structural feature of the protein chain (Hovmoller 2002). The α -helical region ($-180^\circ < \phi < 0^\circ$, $-100^\circ < \psi < 45^\circ$) describes amino acids that are mainly located within α -helices, the β -sheet region ($-180^\circ < \phi < 45^\circ$, $45^\circ < \psi < 180^\circ$ and $-180^\circ < \psi < -135^\circ$), a left-handed α -helical region ($0^\circ < \phi < 180^\circ$, $-90^\circ < \psi < 90^\circ$) which is also called turn region, since the number of left-handed α -helices is low. Additionally, a border region ($-180^\circ < \phi < 0^\circ$, $-100^\circ < \psi < 45^\circ$) between the α -helical and β -sheet region is described.

Different amino acids show similar patterns in the Ramachandran plot and can be grouped together. Based on their conformation, whole groups can be over or under represented in specific turn conformations. Therefore, Ramachandran plots can assist in the interpretation of different amino-acid propensities within turn conformations, so the plots for the amino acids are analysed in more detail here.

Chapter 7.5.1 shows the Ramachandran plots for the 20 amino acids within a non-redundant data set of 1973 protein chains, which is similar to an analysis of 1042 protein chain that is described in literature (Hovmoller 2002). Interestingly, the main-chain conformation of a residue can depend on the χ_1 torsion angle of its sidechain (Chakrabarti 1998). Figure 7.11 shows three possible conformation and depending on the conformation the sidechain atoms can influence the conformation of the C=O group or the N-H group. Therefore, χ_1 values around $+60^\circ$ and 60° should have an effect on ψ and χ_1 values around $+60^\circ$ and 180° on ϕ .

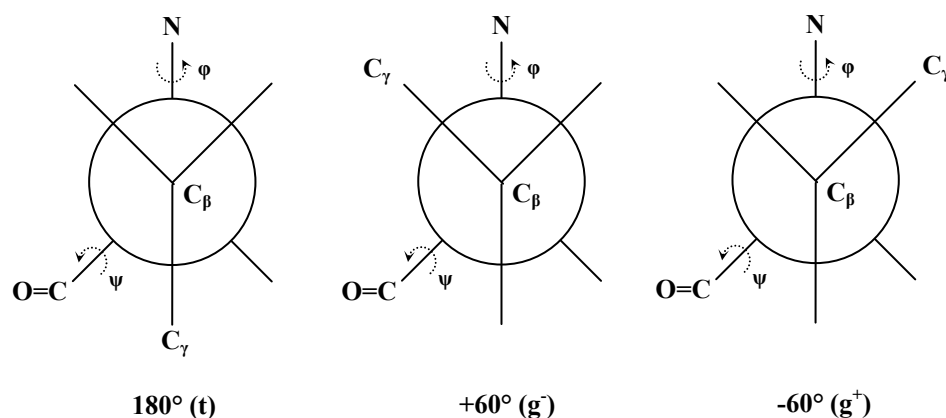


Figure 7.11: Newman projection of three different sidechain conformation

Based on the value of the χ_1 torsion angle, the amino acids can be grouped into four different classes (Chakrabarti 1998). Ser, Cys, Met, Glu, Gly, Lys, Arg (without a branch before the δ position) and Leu belong to one major class. Furthermore, the short polar/acid group with Asp and Asn, the aromatic group with His, Phe, Tyr and Trp and the group with residues that are branched at C_β position: Val, Ile and Thr.

Additionally, the different amino acids can be divided based on the regions of the Ramachandran plot (see section 7.5.1). At first, Pro and Gly that are both unique and show specific regions on the Ramachandran plot. Pro has a five-membered ring that reduces the conformational space. In contrast, Gly lacks a sidechain and steric clashes of the C_β atom are missing. Therefore, Gly conformations run over the borders at -180° and 180° .

The residues Asn, Asp, Thr and Ser show also conformations with ψ around $\pm 180^\circ$ and can be grouped together. These residues share a short sidechain with a possible hydrogen bond donor/acceptor so that a putative stabilizing hydrogen bond could be the explanation for this additional region. Additionally, Asn and Asp show a distinct border region between the α -helical and β -sheet region that should also be explained by sidechain features.

Val, Ile and Thr are branched at C_β position and do not show a distinct left-handed α -region. Steric clashes based on the C_β branch are the explanation for this missing region.

To summarize, the amino acids can be grouped based on conformational similarity:

- i) unique: Pro and Gly
- ii) short polar/acid: Asp and Asn
- iii) aromatic: His, Phe, Tyr and Trp
- iv) C_β branched: Val, Ile and Thr
- v) Ser, Cys, Met, Glu, Gly, Lys, Arg and Leu

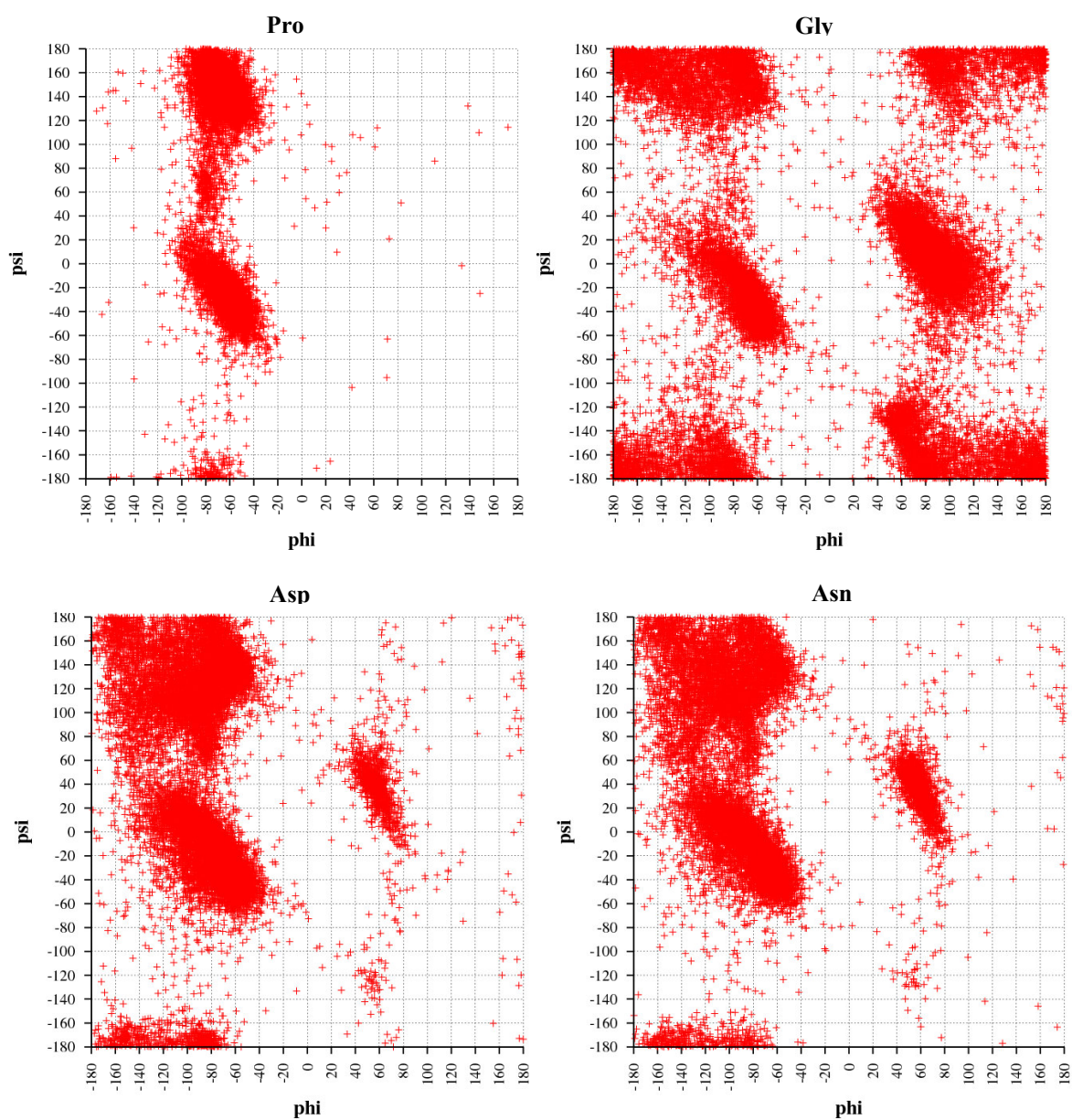
Additionally, Asn, Asp, Thr and Ser share conformations with ψ around $\pm 180^\circ$.

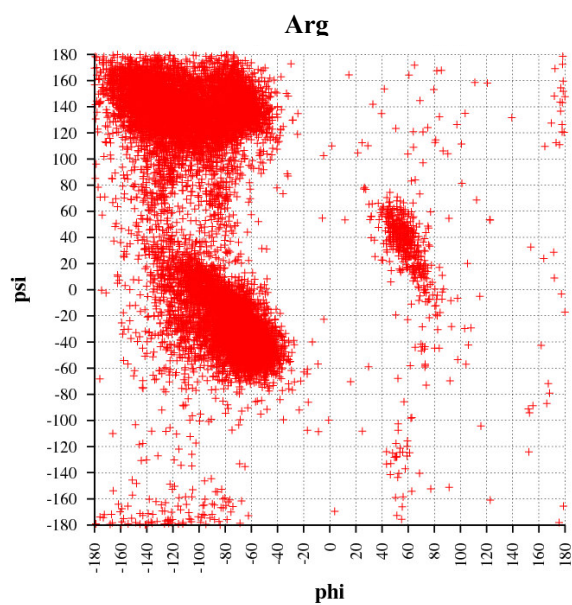
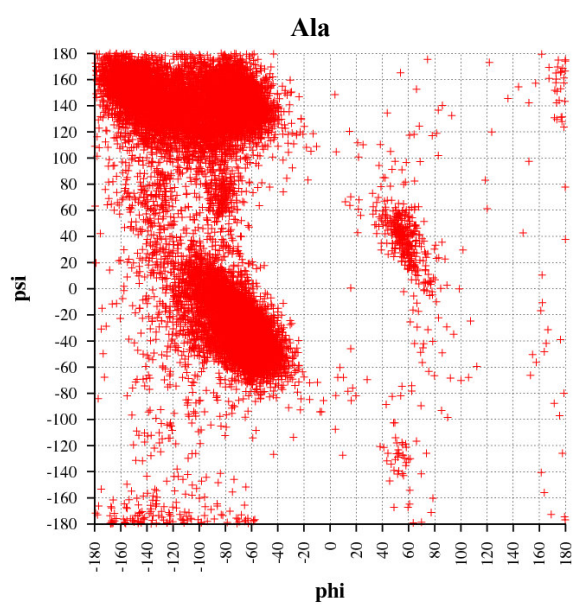
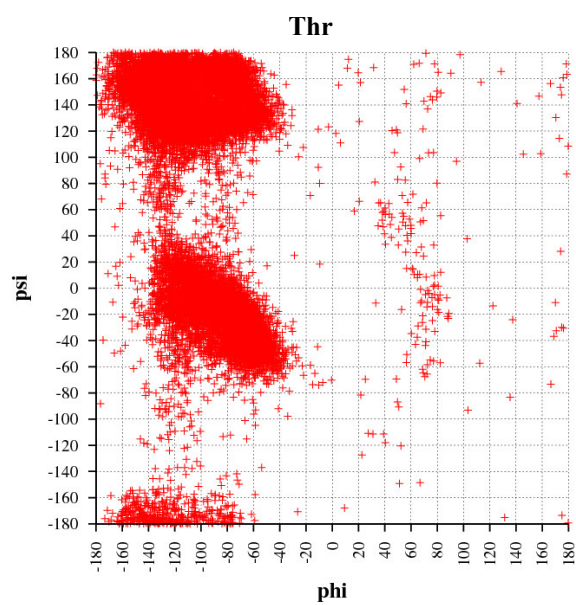
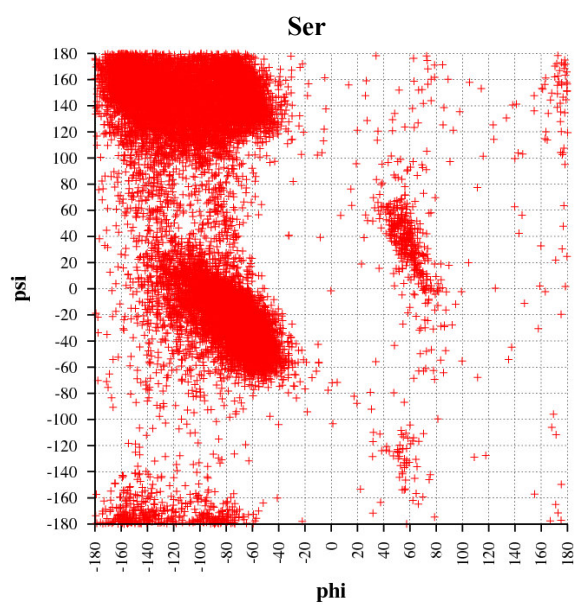
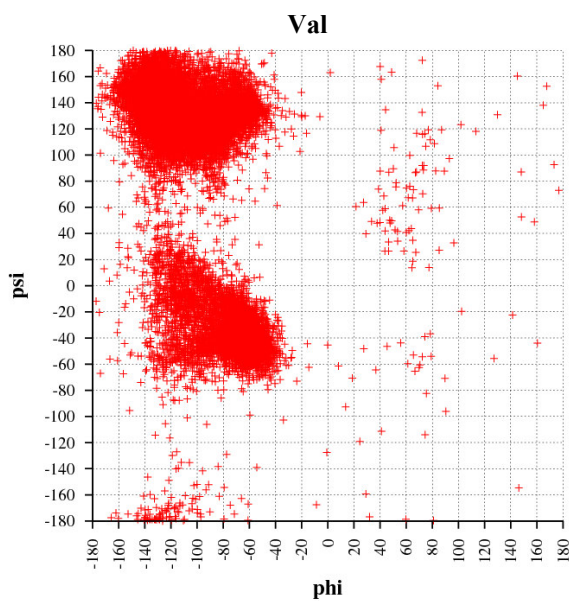
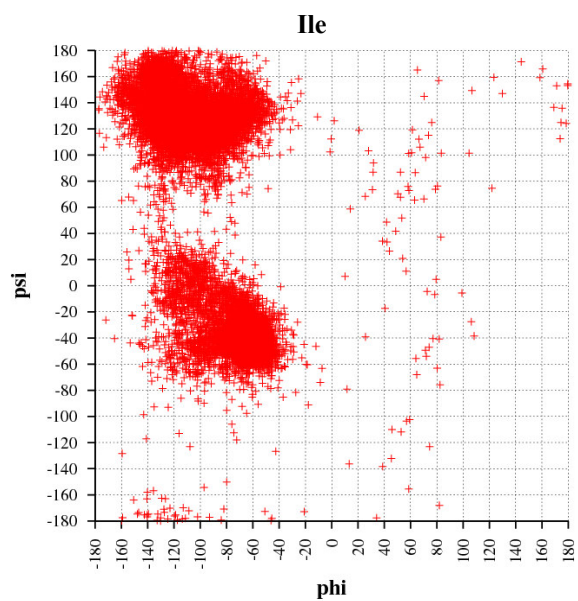
7.5.1 Ramachandran plots

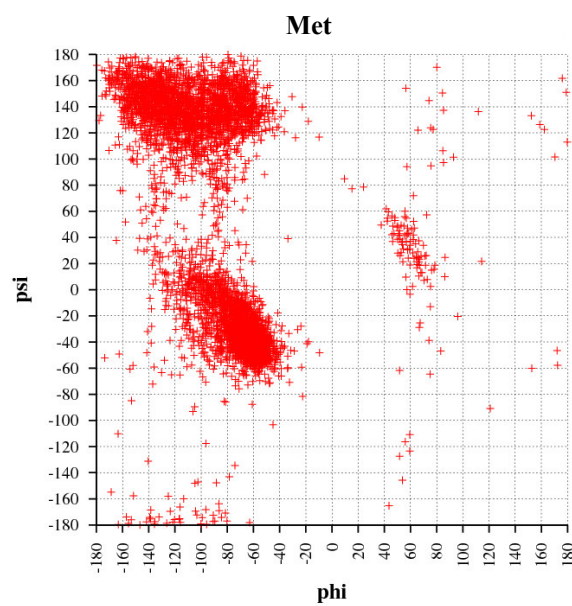
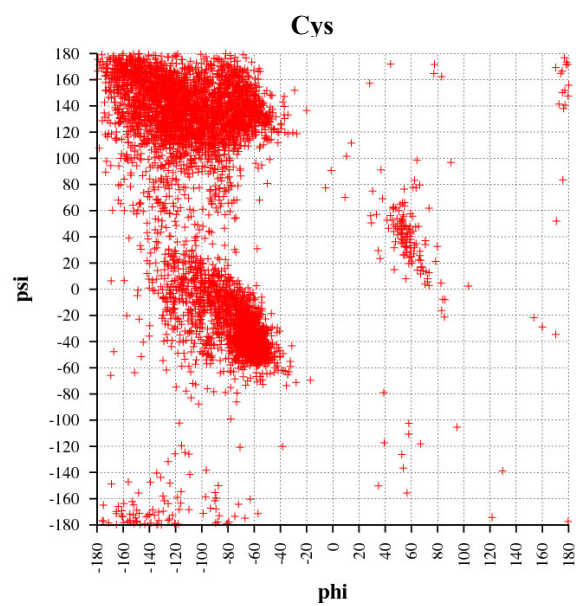
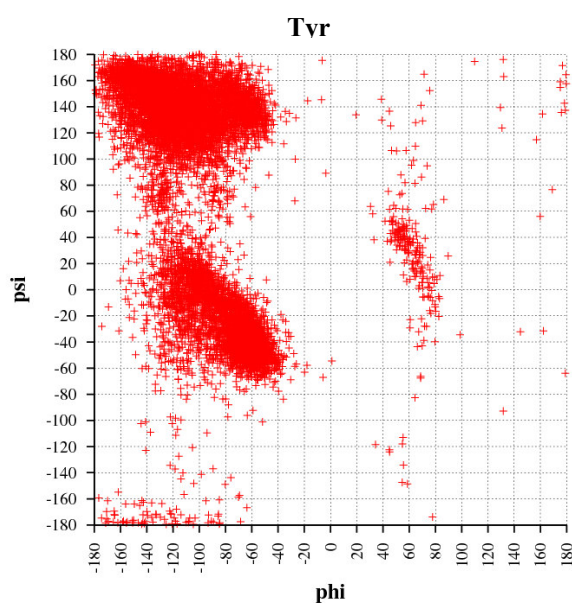
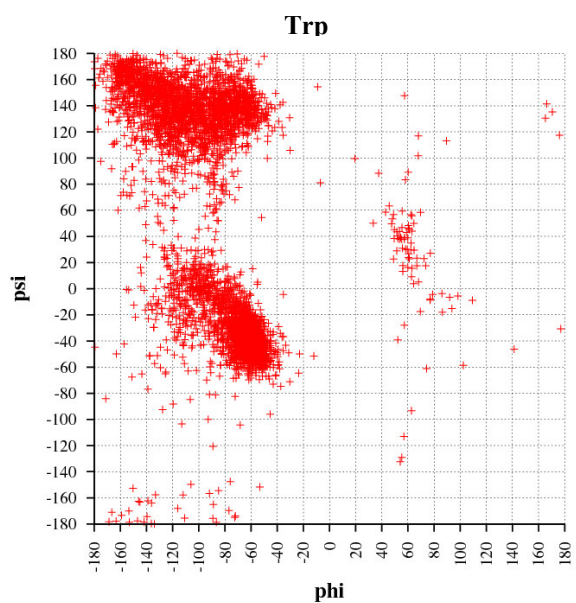
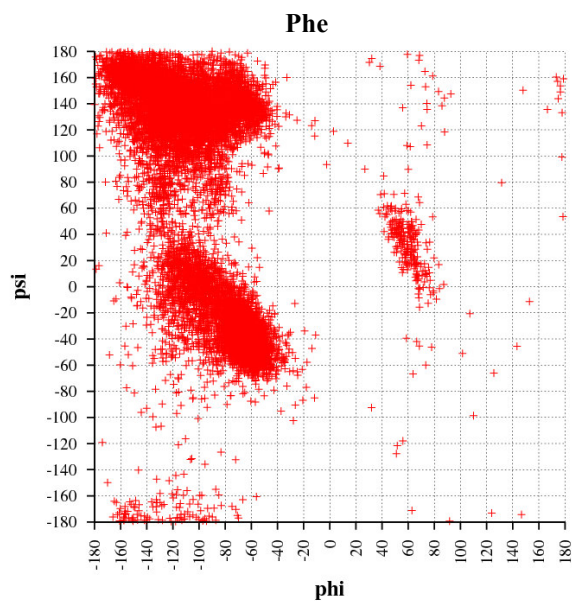
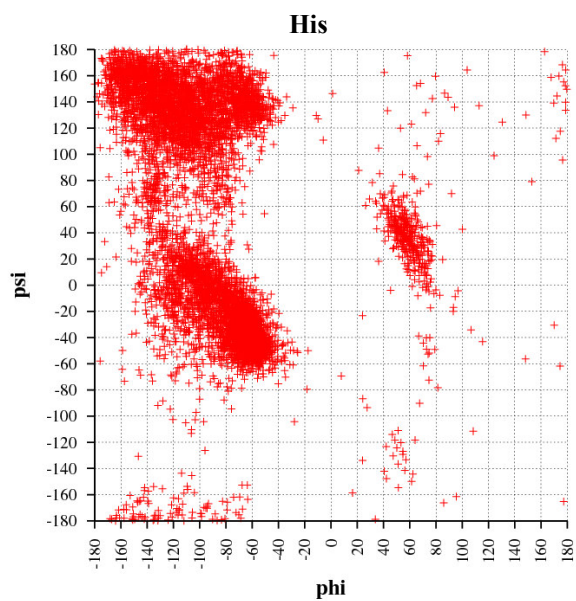
The Ramachandran plots for the 20 amino acids were collected within a non-redundant data set of 1973 different protein chains. They can be grouped into:

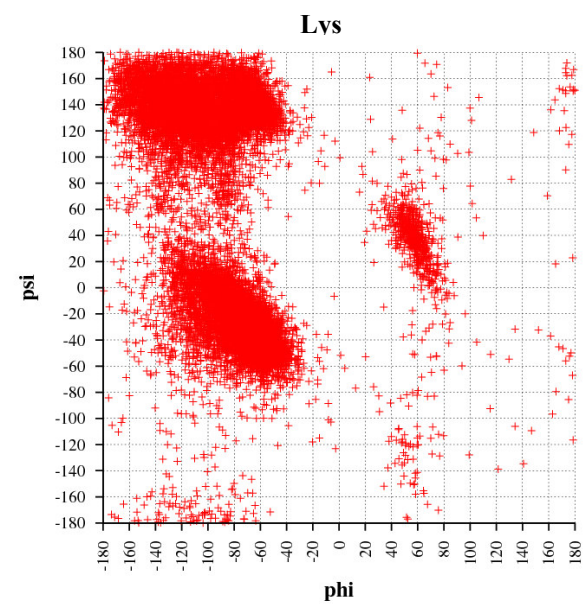
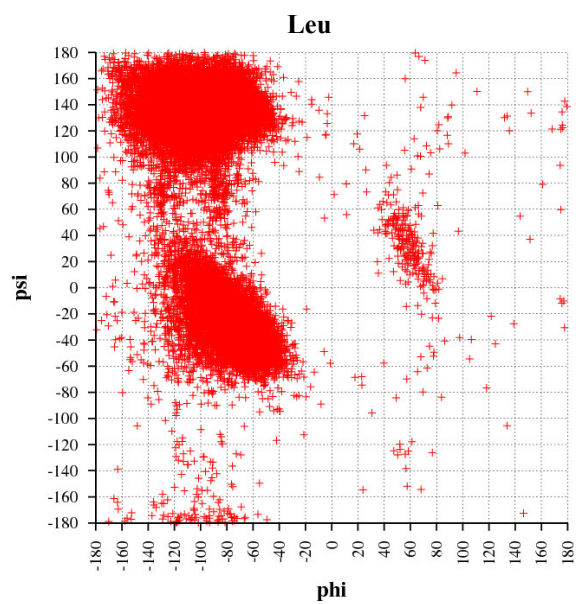
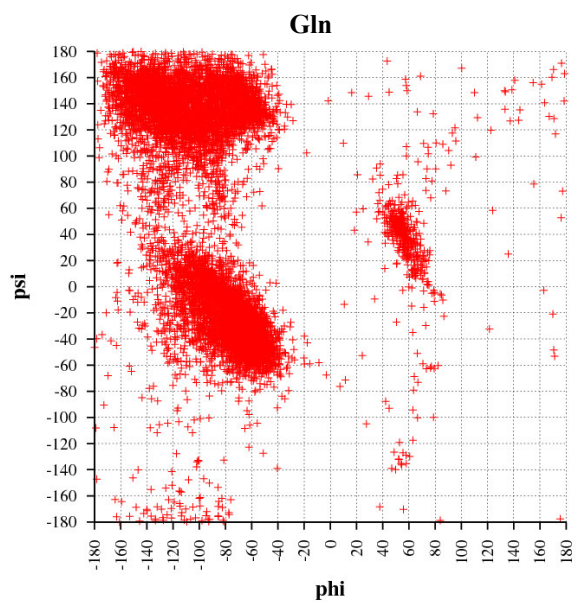
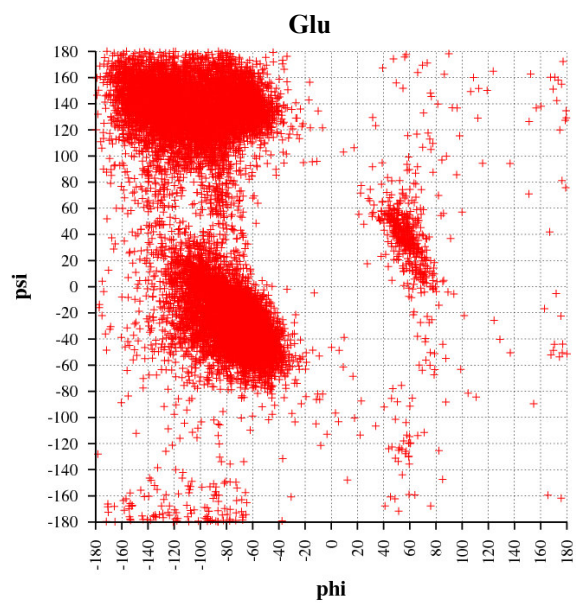
- vi) unique: Pro and Gly
- vii) short polar/acid: Asp and Asn
- viii) aromatic: His, Phe, Tyr and Trp
- ix) C_β branched: Val, Ile and Thr
- x) Ser, Cys, Met, Glu, Gly, Lys, Arg and Leu

Additionally, Asn, Asp, Thr and Ser share conformations with ψ around $\pm 180^\circ$.









8 Turn Classification: Results and discussion

8.1 Introduction

8.1.1 Definitions

The current definition of most turn families is based on the $C\alpha$ - $C\alpha$ distance with a distance smaller than 7\AA or a hydrogen bond between CO_i and NH_{i+n} , but during this analysis three different subcategories for turns based on the hydrogen-bonding pattern between the first and the last residue have been introduced (see Figure 8.1):

- i) A standard or *normal* conformation with a hydrogen bond between CO_i and NH_{i+n} .
- ii) A distorted or *open* turn conformation lacking a hydrogen bond, with a $C\alpha_i$ - $C\alpha_{i+n}$ distance $< 10\text{\AA}$ and a visual inspection of the resulting cluster.
- iii) A *reverse* conformation with a hydrogen bond between NH_i and CO_{i+n} .

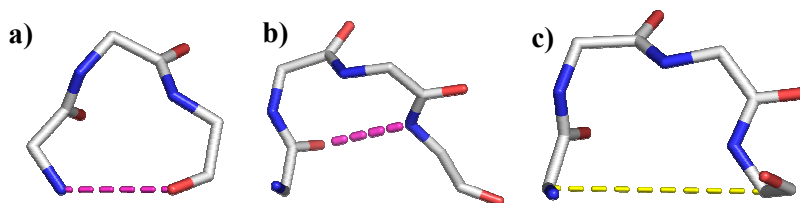


Figure 8.1: Different turn conformation
 a) reverse turn with 3 residues, b) normal β -turn (4 residues),
 c) open β -turn (4 residues)
 magenta: hydrogen bonding, yellow: $C\alpha$ - $C\alpha$ distance

The *inner* residues of a normal or open turn are those turns that lie within the described hydrogen bonded ring (see section 6.7.2). The φ and ψ angles of these residues and the additional ω angles are used for clustering (see section 7).

During the last decade a lot of names and synonyms for the different turn families have been used. In this work the following definitions are applied:

- turn: as described above, between 2 and 6 residues long, intraturn hydrogen bonded or a $C\alpha_i$ - $C\alpha_{i+n}$ distance smaller than a specific cut off
- loop: part of polypeptide chain that connects two secondary structure elements
- β hairpin (turn): a loop that connects two adjacent, antiparallel β -strands

Furthermore, new turn families occurred in this work and the differentiation into families is based on the following definitions:

- category: based on the hydrogen-bond pattern/ $C\alpha$ - $C\alpha$ -distance (open, normal, reverse)
- class: turns with same lengths (e.g. four residue turns)
- family: one specific type of turn (e.g. the normal α -turns or the open π -turns)
- turn type: a specific conformation within one turn family

8.1.2 Description of an example ESOM map

Figure 8.2 shows an ESOM map of normal β -turn clustering. The yellow points outside the region of valleys and mountains are belonging to cluster 1 and turn-type I. The number of entries is so high and in combination with the specific features of ESOMs it only looks like if they are not surrounded by a mountain. The small extension reveals a closer look to unclassified data points. On the one hand, the white points that are lying on border regions and therefore do not belong to any clusters. On the other hand, the red points indicate a broken chain and these conformations also do not belong to a specific turn-type and cluster.

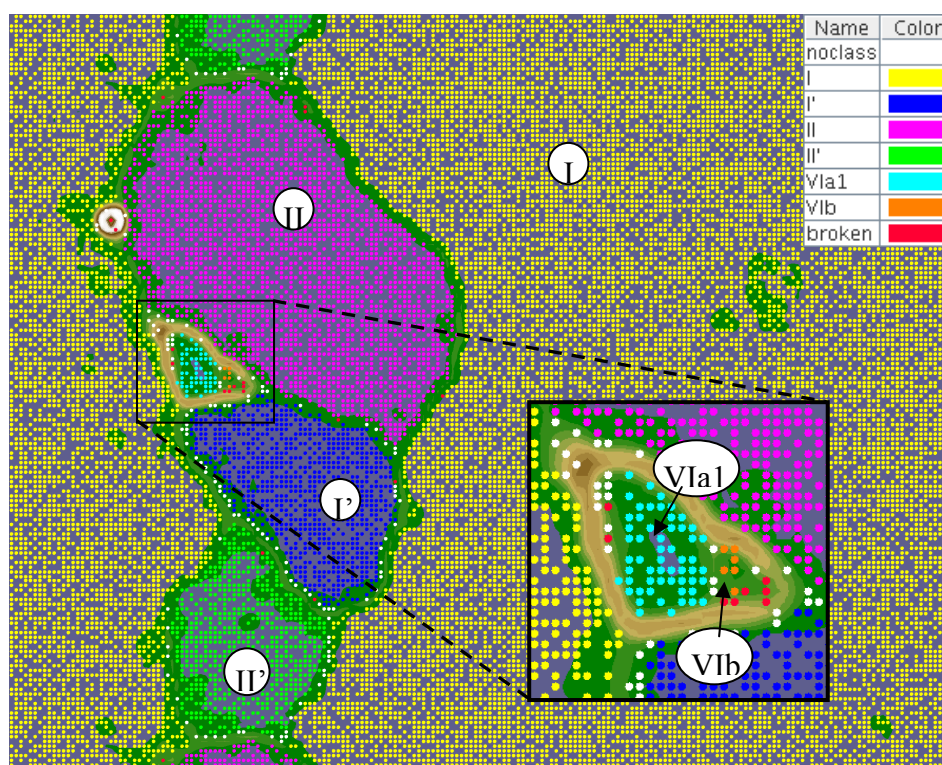


Figure 8.2: Trained ESOM map for normal β turns

8.1.3 Torsion angle analysis

During the following analysis the mean torsion angles and standard deviations for torsion angles among a specific turn-type are described. These values are calculated from all turn structures of each turn-type. In the case of the transformed ω angles, the mean torsion angles were re-transformed back to their origin value to represent the torsion angle around $\pm 180^\circ$ and 0° .

8.2 Two residue turns

8.2.1 δ -turn

As mentioned previously, δ -turns with a hydrogen bond between NH_i and CO_{i+1} were never classified in proteins before, but using the DSSP energy function 194 δ -turn structures with weak hydrogen bonding (between -0.5 and -0.7 kcal/mol) were identified.

Clustering these turn structures leads to 9 different clusters (Figure 8.3). Some of these clusters show similar torsion angle distributions (Table 8.1), which leads to four different conformations that cross the torsion angle boundary of $\pm 180^\circ$. Therefore, seven turn-types could be found within the δ -turn family. Comparison of the torsion angles with the theoretically described torsion angles for tripeptides by Nagarajaram et al. (Nagarajaram 1992) showed that only two of the described turn-types (type 3 and 4, see section 6.1 for details) could be roughly identified as similar to turn-type class IV which exhibit a cis peptide bond.

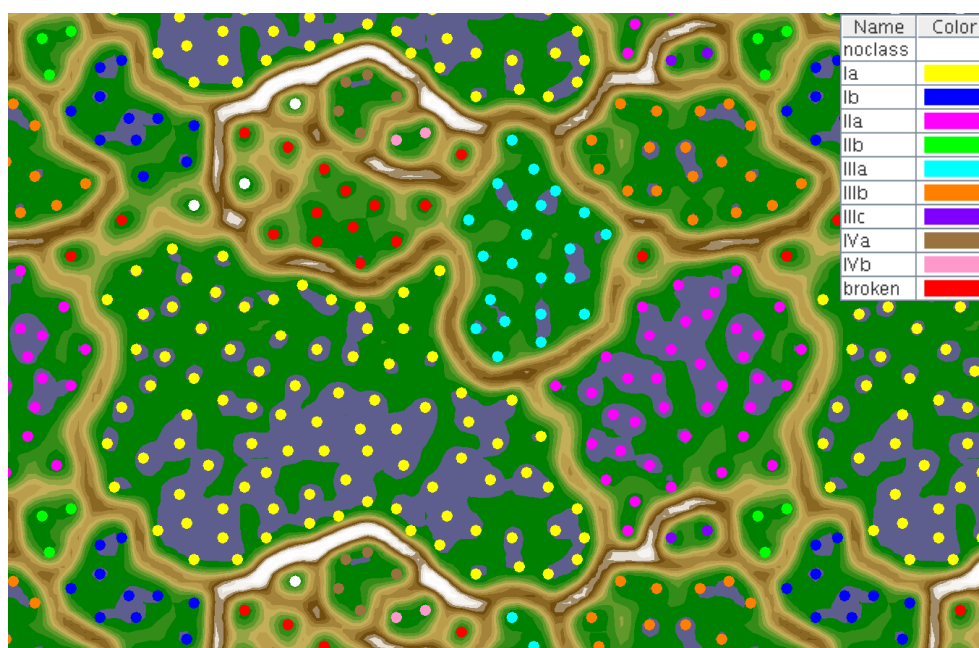


Figure 8.3: ESOM clustering for δ -turns

cluster	type	nr	ϕ_i		ψ_i		ω_i		ϕ_{i+1}		ψ_{i+1}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	3	-145.1	6.5	-57.9	136.2	-63.8	110.9	-33.2	83.4	-153.8	29.2
1	Ia	90	-124.0	21.0	-51.0	27.3	175.9	6.4	-126.7	34.2	155.1	15.5
2	Ib	11	-116.2	17.7	-49.8	18.2	174.4	7.4	-148.6	22.7	-169.5	12.6
3	IIa	38	127.4	21.6	-26.3	22.2	176.4	6.8	-87.0	29.4	148.5	24.9
4	IIb	3	104.8	10.0	-24.5	25.6	174.1	4.6	-105.2	8.7	-172.8	3.6
5	IIIa	22	-133.2	19.6	27.4	30.1	-177.5	5.3	124.2	32.7	162.9	15.1
6	IIIb	18	-131.8	21.6	42.2	35.5	-175.0	6.2	84.5	30.8	-160.8	15.8
7	IIIc	2	166.5	11.1	32.7	23.9	-176.5	0.1	127.5	30.0	-163.9	0.2
8	IVa	5	-148.4	13.8	98.2	12.4	-0.4	3.9	-74.0	12.6	159.7	15.3
9	IVb	2	131.7	23.4	115.5	7.9	-9.7	13.3	-57.9	9.0	173.1	5.1

Table 8.1: Mean torsion angle and standard deviation for δ -turn cluster

Evaluating the overall frequencies for δ -turns, Table 8.2 clearly shows that Gly is over represented with a high conformational parameter at both positions in addition to Ile at position 1. In contrast, Pro and Ala are under represented at position 1 and Met, Val and Lys at position 2.

aa	turn		position 1		position 2	
	f	P	f	P	f	P
ALA	22	0.75	< 5	0.34	17	1.16
ILE	23	1.01	> 18	1.59	5	0.44
LEU	28	0.80	13	0.74	15	0.86
MET	1	0.13	1	0.27	< 0	0.00
PHE	11	0.67	5	0.61	6	0.73
PRO	12	0.69	< 0	0.00	12	1.37
VAL	< 10	0.37	9	0.66	< 1	0.07
ARG	9	0.46	4	0.41	5	0.52
ASP	24	1.07	13	1.15	11	0.98
GLU	18	0.69	9	0.69	9	0.69
LYS	18	0.77	13	1.11	< 5	0.43
ASN	21	1.22	11	1.28	10	1.16
CYS	8	1.26	2	0.63	6	1.89
GLN	17	1.13	9	1.20	8	1.06
HIS	16	1.69	8	1.69	8	1.69
SER	21	0.89	10	0.85	11	0.93
THR	24	1.12	11	1.02	13	1.21
TRP	5	0.86	2	0.69	3	1.03
TYR	11	0.79	5	0.72	6	0.86
GLY	> 89	3.32	> 46	3.43	> 43	3.21
OTH	0	0.00	0	0.00	0	0.00
ALL	388	1.00	194	1.00	194	1.00

Table 8.2: Overall frequencies ('f') and conformational parameter ('P') for δ -turn types I – VIb (statistical over representation: '>', statistical under representation: '<', other amino acids: 'OTH')

A detailed analysis of the amino acid propensities (see appendix) reveals a miscellaneous turn-type I with a high occurrence of Ile at position 1 and Asn/Asp at position 2. In contrast, the frequency of Met, Pro and Val is low in this turn-type. Among the other turn-types, a dominance of one amino acid is observed, more precisely Gly at position 1 in turn-type II, at position 2 in turn-type III and Pro at position 2 in turn-type IV.

The Ramachandran plot of the δ -turn-types (Figure 8.4) clearly indicates the reason for the specific amino acid propensities. Mainly Asp and Asn can adopt conformations that are typical for turn-type Ib, in more detail ψ around 180° . The phi and psi angles for the position with a high occurrence of Gly of turn-type II ($\phi_i: \sim 120^\circ / \psi_i: \sim 25^\circ$) and III ($\phi_i: \sim 120^\circ / \psi_i: \sim \pm 180^\circ$) are lying in regions that are typical for Gly. Normally, only Gly shows these conformations because of the missing sidechain. Furthermore, the high frequency of Pro in turn-type IV is in agreement with the knowledge that mostly Pro exhibit a cis-peptide bond, which is characteristically for this turn-type.

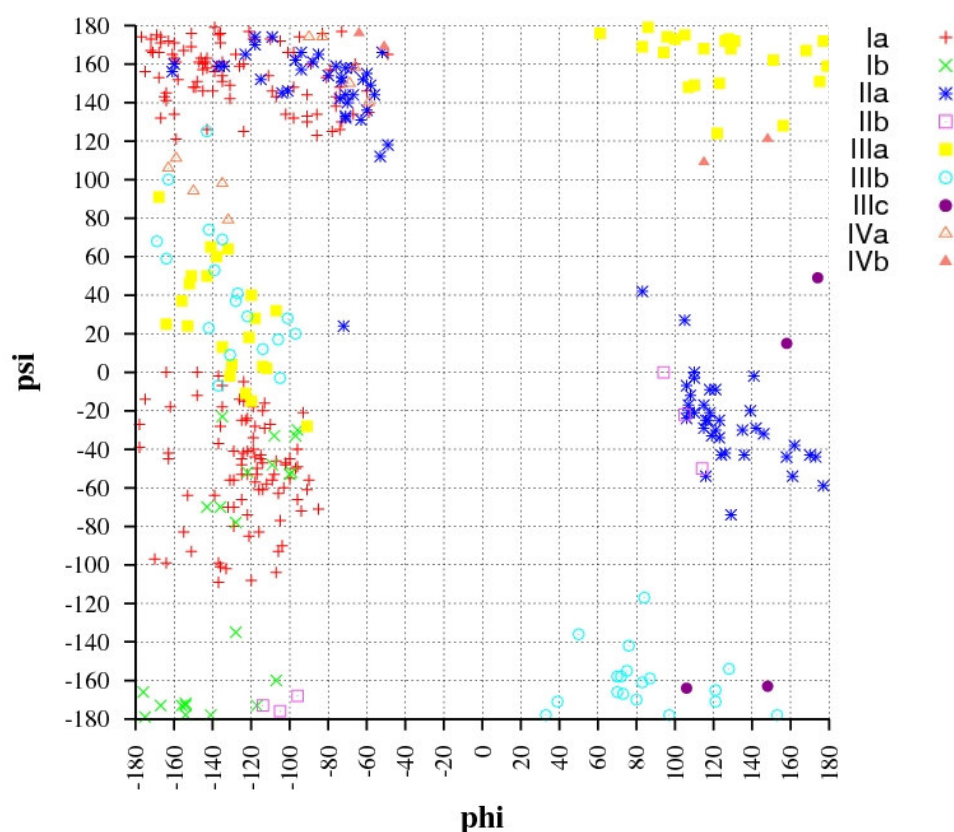


Figure 8.4: Ramachandran plot for δ turn-types Ia – IV
(both residue of a specific turn-type are shown in same color)

8.3 Three residues turns

8.3.1 γ -turns

γ -turns are the smallest group of described turns with a standard hydrogen bond between the main-chain CO_i and NH_{i+2} . The detected ~ 20000 γ -turn structures clustered into two already described turn-types, the normal and inverse γ -turn (Figure 8.5). Table 8.3 shows the mean torsion angles and Table 8.4 the minimum and maximum torsion angles within the described clusters. A large deviation from minimum/maximum torsion angles described by Guruprasad (Guruprasad 2000) (see section 6.3) is found.

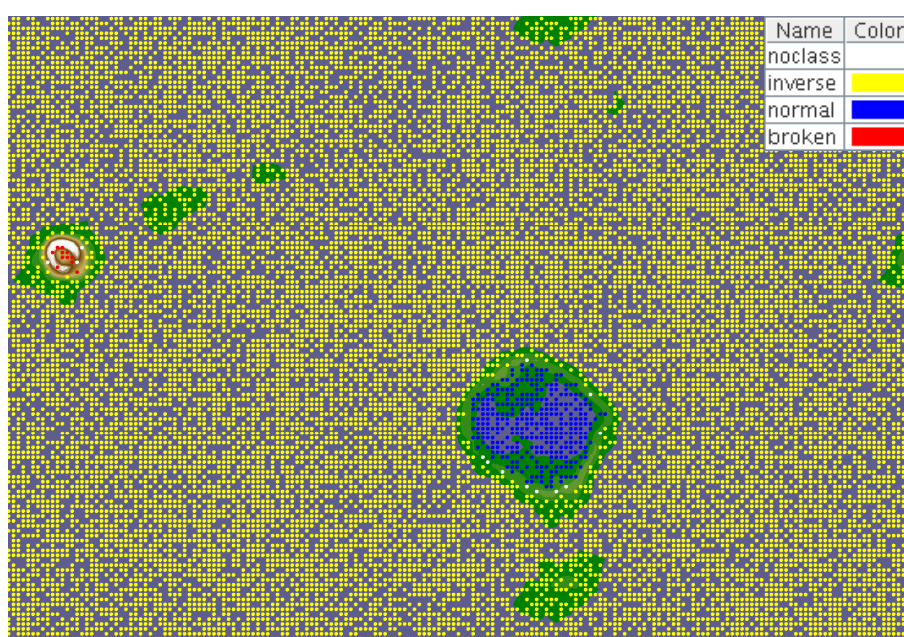


Figure 8.5: ESOM clustering for γ -turns

cluster	turn-type	nr	ω_i		ϕ_{i+1}		ψ_{i+1}		ω_{i+1}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	14	-167.6	35.7	-6.9	49.7	8.0	34.1	178.2	15.3
1	inverse	19713	179.0	4.6	-93.3	13.5	108.1	20.2	-179.8	4.7
2	normal	445	178.8	4.8	75.3	15.2	-65.8	28.7	-179.6	5.7

Table 8.3: Mean torsion angle and standard deviation for clustered γ -turns

cluster	turn-type	ω_i		ϕ_{i+1}		ψ_{i+1}		ω_{i+1}	
		min	max	min	max	min	max	min	max
1	inverse	116.5	-128.3	-141.2	11.4	-47.5	145.2	130.5	-137.4
2	normal	151.1	-165.0	15.8	128.5	-144.5	12.9	146.8	-144.6

Table 8.4: Minimum and maximum torsion angles for clustered γ -turns

The overall amino-acid propensity (Table 8.5) shows that Gly is highly under represented and Ile/Val and Asp/Asn are highly over represented at position 2. Additionally, Pro is highly under represented at position 3, which is not surprising, since Pro is unable to form a hydrogen bond. Asp shows a low occurrence at position 1, Ala, Pro and Ser at position 2 and Ala and Gly at position 3. In contrast, Val has a high occurrence at position 3.

aa	turn		position 1		position 2		position 3	
	f	P	f	P	f	P	f	P
ALA	< 3031	0.66	< 1088	0.71	< 836	0.55	< 1107	0.73
ILE	> 5820	1.65	> 1740	1.48	> 2348	1.99	> 1732	1.47
LEU	> 6414	1.18	> 1962	1.08	> 2152	1.18	> 2300	1.26
MET	1071	0.91	< 324	0.83	< 332	0.85	415	1.06
PHE	> 2934	1.15	> 1046	1.22	899	1.05	> 989	1.16
PRO	< 1413	0.52	< 825	0.91	< 588	0.65	< 0	0.00
VAL	> 7092	1.67	> 2210	1.56	> 2417	1.71	> 2465	1.74
ARG	< 2591	0.86	< 906	0.90	< 784	0.78	< 901	0.89
ASP	> 3767	1.07	< 781	0.67	> 1935	1.65	< 1051	0.90
GLU	< 2992	0.74	< 1029	0.76	< 969	0.72	< 994	0.74
LYS	< 3158	0.87	< 1120	0.92	< 972	0.80	< 1066	0.88
ASN	> 2924	1.09	< 816	0.91	> 1242	1.39	866	0.97
CYS	1066	1.08	315	0.96	360	1.09	> 391	1.19
GLN	< 1916	0.82	< 661	0.85	< 642	0.82	< 613	0.78
HIS	1535	1.04	527	1.07	503	1.02	505	1.03
SER	< 2580	0.70	< 857	0.70	< 661	0.54	< 1062	0.87
THR	> 3544	1.06	> 1183	1.06	< 912	0.82	> 1449	1.30
TRP	> 1065	1.17	328	1.08	329	1.09	> 408	1.35
TYR	> 2517	1.16	> 944	1.31	744	1.03	> 829	1.15
GLY	< 2893	0.69	1452	1.04	< 486	0.35	< 955	0.69
OTH	< 158	0.72	< 48	0.66	< 49	0.67	61	0.83
ALL	60481	1.00	20162	1.00	20160	1.00	20159	1.00

Table 8.5: Overall frequencies and conformational parameter for γ -turns
(statistical over representation: '>', statistical under representation: '<', other amino acids: 'OTH')

Considering the amino-acid distribution (see appendix) among the γ -turn-types, it appears that there is a high occurrence of Ile/Val at all three positions and Asp/Asn at position 2 within the inverse γ -turn-type and only a low occurrence of Ser, Gly, Ala at position 2. Ile and Val show similar sidechain features, they are both β -branched. Additionally, Asp and Asn have also a “branched” sidechain with similar sidechain features. A putative stabilizing feature of this branched sidechains could be proposed and the low occurrence of Ser/Gly/Ala with a short sidechain of Ser/Ala and a missing sidechain of Gly add weight to this hypothesis.

Within the normal γ -turn-type, Gly shows a high frequency with a very high occurrence at position 2 (see appendix). Asp and Ser show a high occurrence at position 3 and in contrast

Ile is under represented at position 1 and 2, Leu at position 2 and 3. Trp appears infrequently and Pro is never found at position two and three. In contrast to the inverse γ -turn the stabilizing effects of the described amino acids seem unneeded. The missing sidechain of Gly leaves enough conformational space.

The Ramachandran plot (Figure 8.6) shows, that both classes fall into clearly defined, separated regions with inverse torsion angles.

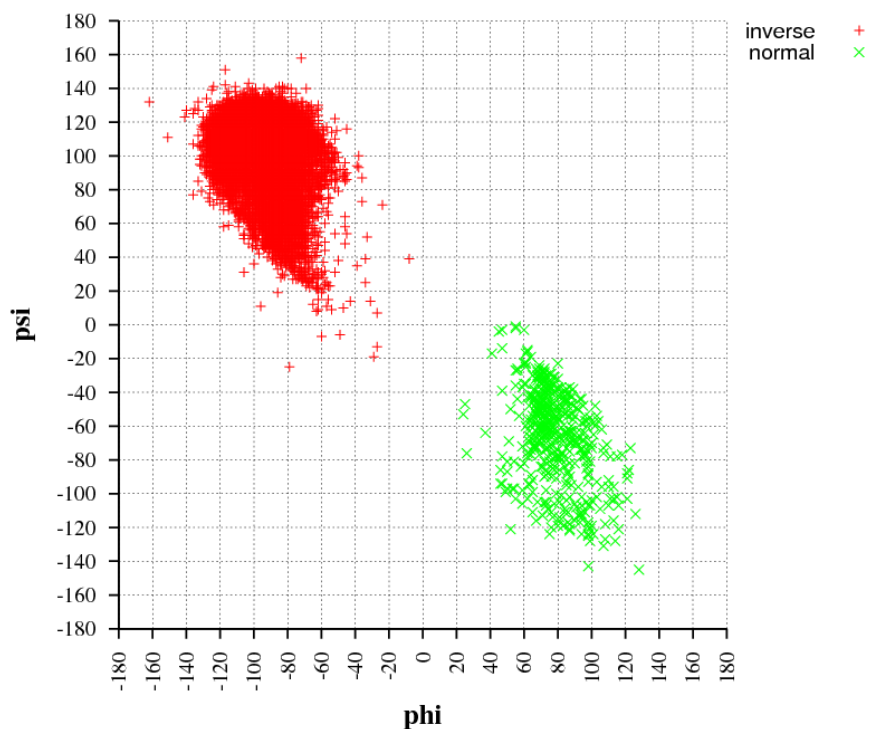


Figure 8.6: Ramachandran plot of the inner residue for normal and inverse γ -turns

8.3.2 ϵ -turns

In the dataset, 114 turn structures with a reverse hydrogen bond between NH_i and CO_{i+2} were identified in the data set. These structures cluster into five different turn-types (Figure 8.7); none of these classes include a cis-peptide bond (Table 8.6). In contrast to δ -turns, the reverse hydrogen bonds in these turns are slightly stronger (Figure 8.8). The characteristic for a normal turn (a hydrogen bond between CO_i and NH_{i+2}) does not occur within these turns. Thus completely new and previously not classified structures are observed that are independent from γ -turns.

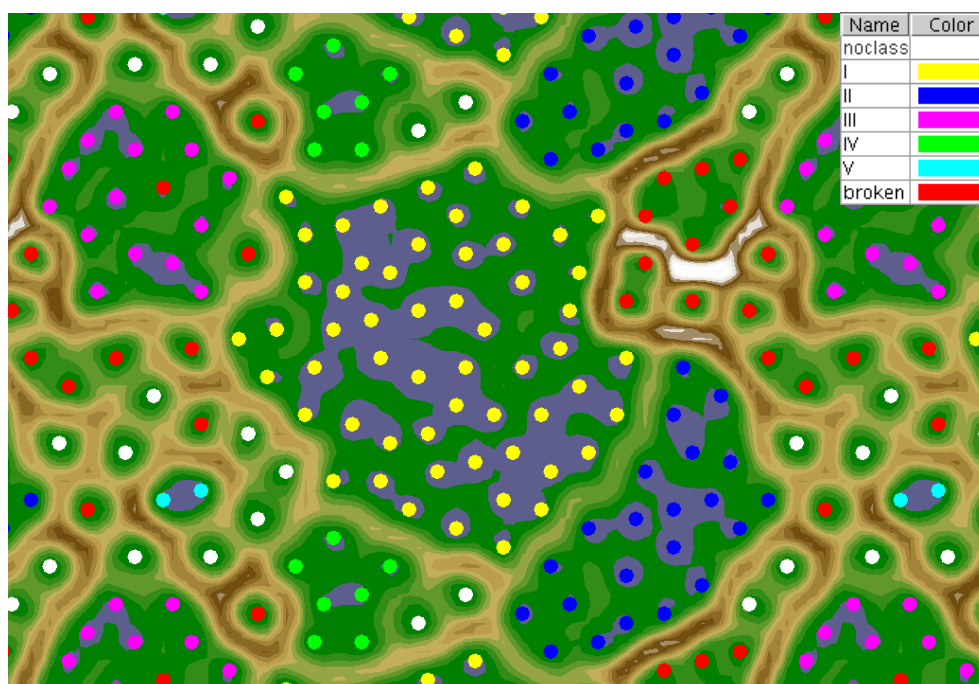


Figure 8.7: ESOM clustering for ϵ turns

type	nr	ϕ_i		ψ_i		ω_i		ϕ_{i+1}		ψ_{i+1}		ω_{i+1}		ϕ_{i+2}		ψ_{i+2}	
		\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
no	12	-31.1	111.5	-3.4	93.1	-149.6	75.1	-27.0	114.4	30.1	50.0	-176.2	16.1	5.4	113.7	-0.2	148.1
I	56	-127.2	24.3	-121.5	28.4	177.2	4.2	-73.4	17.6	-30.8	19.8	176.5	4.7	-118.9	22.4	155.1	17.8
II	22	107.4	28.2	-28.6	29.9	175.4	7.4	-99.1	18.1	-27.3	19.9	178.1	4.4	-124.1	25.6	148.3	19.2
III	15	-122.3	25.3	53.2	54.2	-178.8	5.0	95.9	28.9	-24.0	34.0	176.5	6.4	-141.4	26.7	157.9	9.7
IV	7	-142.1	28.7	-125.9	42.1	-179.8	3.0	-61.4	12.7	-38.8	18.2	175.8	6.3	-132.7	31.1	-170.2	7.5
V	2	-172.1	6.0	116.9	0.1	-178.1	0.3	48.3	4.6	46.5	4.5	-178.1	0.2	112.6	1.0	-159.3	7.3

Table 8.6: Mean torsion angle and standard deviation for ϵ -turn cluster

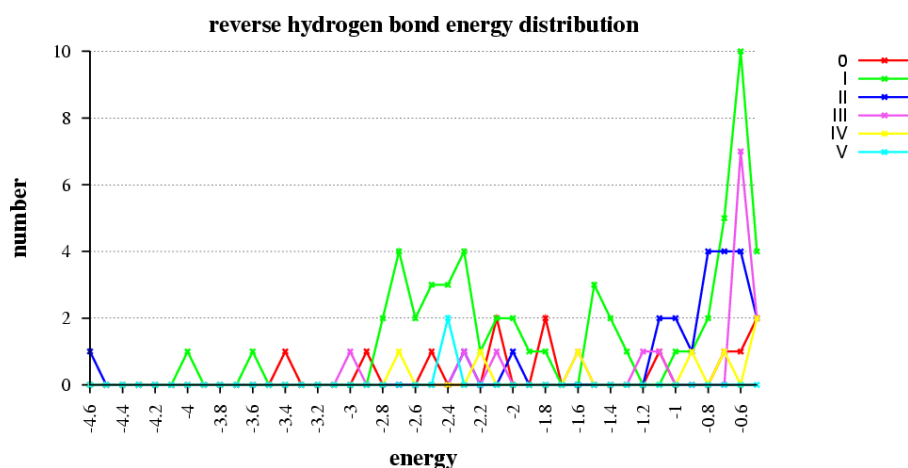


Figure 8.8: Reverse hydrogen bond energy distribution of ϵ -turn types

ϵ -turns show an over representation of Gly at position 1 and 2 (Table 8.7) which is mainly based on the high frequency at position 1 in turn-type I and II and position 2 in turn-type III (see appendix). This high frequency of Gly can be easily explained consulting the Ramachandran plot (Figure 8.9) since the residues at the described positions adopt conformations that are predominantly accessible only by Gly. In contrast, Pro and Val are under represented at position 1 and Phe at position 3. Interestingly, Pro often occurs at the first position following a ϵ -turn.

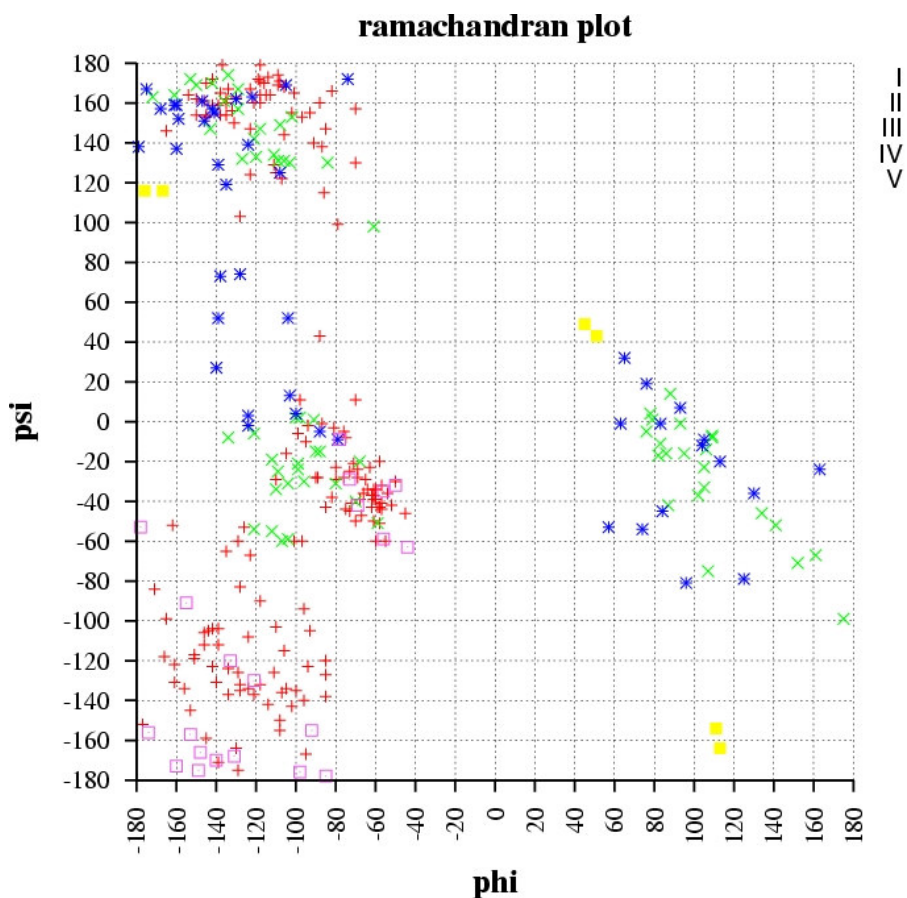


Figure 8.9: Ramachandran plot for ϵ turn-types

aa	turn		position 1		position 2		position 3		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	29	1.12	12	1.39	12	1.39	5	0.58	7	0.81	7	0.81
ILE	14	0.70	2	0.30	4	0.60	8	1.20	3	0.45	3	0.45
LEU	17	0.55	5	0.49	7	0.68	5	0.49	< 4	0.39	< 4	0.39
MET	3	0.45	2	0.90	1	0.45	0	0.00	2	0.90	1	0.45
PHE	5	0.35	1	0.21	4	0.83	< 0	0.00	8	1.66	4	0.83
PRO	6	0.39	< 0	0.00	3	0.58	3	0.58	6	1.17	> 19	3.70
VAL	14	0.58	< 1	0.13	5	0.63	8	1.00	8	1.00	6	0.75
ARG	9	0.53	2	0.35	3	0.53	4	0.70	5	0.88	4	0.70
ASP	24	1.21	6	0.91	8	1.21	10	1.51	6	0.91	11	1.66
GLU	31	1.35	8	1.05	11	1.44	12	1.57	< 2	0.26	4	0.52
LYS	18	0.87	2	0.29	7	1.02	9	1.31	10	1.46	9	1.31
ASN	18	1.18	8	1.58	8	1.58	2	0.40	6	1.18	4	0.79
CYS	3	0.54	0	0.00	1	0.54	2	1.07	1	0.54	3	1.61
GLN	8	0.60	1	0.23	1	0.23	6	1.36	4	0.91	3	0.68
HIS	13	1.56	4	1.44	5	1.80	4	1.44	6	2.16	6	2.16
SER	20	0.96	7	1.01	4	0.58	9	1.30	7	1.01	> 12	1.73
THR	24	1.27	5	0.79	11	1.74	8	1.27	10	1.58	7	1.11
TRP	8	1.56	4	2.33	1	0.58	3	1.75	2	1.17	2	1.17
TYR	12	0.98	4	0.98	4	0.98	4	0.98	> 8	1.96	1	0.25
GLY	> 66	2.80	> 40	5.08	> 14	1.78	12	1.53	9	1.14	3	0.38
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	2.42
ALL	342	1.00	114	1.00	114	1.00	114	1.00	114	1.00	114	1.00

Table 8.7: Overall frequencies and conformational parameter for ϵ turn types I – V (statistical overrepresentation: '>', statistical underrepresentation: '<', other amino acids: 'OTH')

8.4 Four residue turns

8.4.1 Normal β -turn

Within the dataset 28650 different turn structures containing a hydrogen bond between CO_i and NH_{i+3} could be determined and these structures were grouped in 6 different types (Figure 8.10).

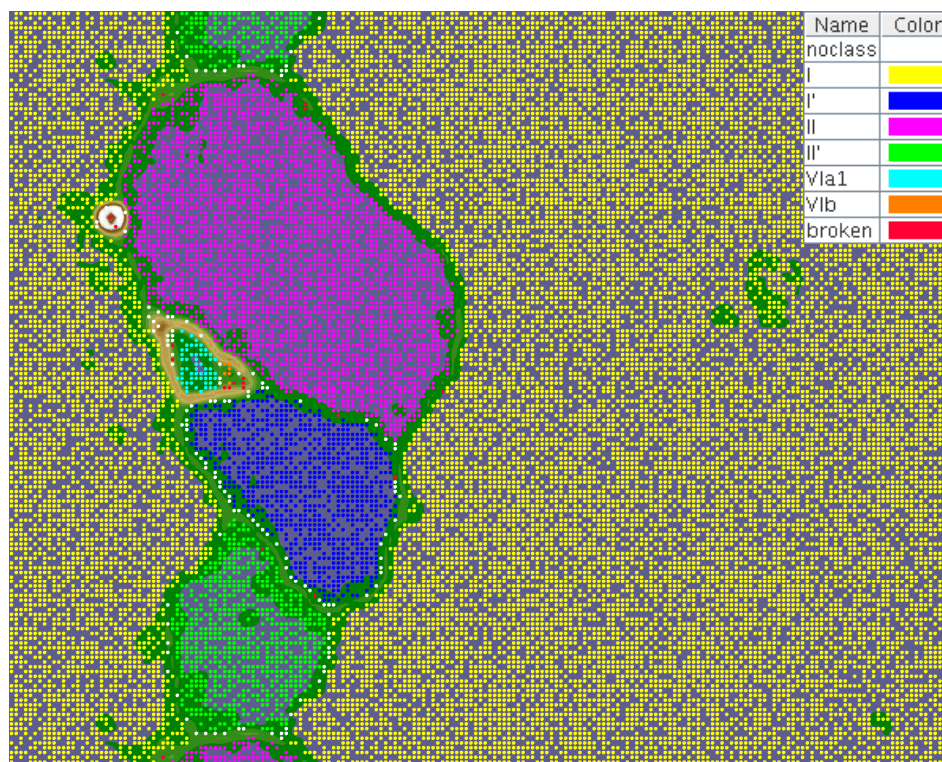


Figure 8.10: Trained ESOM map for normal β turns

Comparing these types and the corresponding mean torsion angle ranges (Table 8.8, Table 8.9) to the widely accepted classification from Hutchinson and Thornton (Hutchinson 1994) reveals good agreement with their classification showing slightly different mean torsion angles. Only turn-types VIa2 and VIII could not be identified within the normal β -turn family.

cluster	type	nr	ω_i		ϕ_{i+1}		ψ_{i+1}		ω_{i+1}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	153	-179.3	18.8	16.5	51.7	20.2	76.0	-162.5	53.9
1	I	23115	-179.0	3.7	-61.2	10.1	-25.7	15.2	-180.0	3.3
2	I'	1183	179.7	3.4	54.1	9.1	38.0	15.6	178.6	3.0
3	II	3160	179.5	4.2	-58.5	10.1	133.4	11.3	179.2	3.3
4	II'	934	-179.4	5.3	59.3	10.2	-128.4	14.0	-179.5	3.8
5	VIa1	93	178.6	5.4	-58.0	9.4	142.7	8.1	3.6	6.7
6	VIb	12	-179.0	4.5	-125.6	8.3	79.3	15.5	4.0	2.4

Table 8.8: Mean torsion angles ($\omega_i - \omega_{i+1}$) and standard deviation for normal β -turn types

cluster	type	nr	ϕ_{i+2}		ψ_{i+2}		ω_{i+2}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	153	16.5	78.0	3.0	41.0	-177.4	16.1
1	I	23115	-82.6	20.0	-6.3	19.1	-180.0	3.9
2	I'	1183	75.3	16.7	6.9	18.9	179.9	3.1
3	II	3160	79.9	19.0	4.7	21.4	179.5	3.9
4	II'	934	-85.5	20.7	-1.4	21.5	-180.0	4.2
5	VIa1	93	-90.8	10.9	15.7	15.2	179.7	4.2
6	VIb	12	-101.3	9.2	165.2	9.4	178.4	3.1

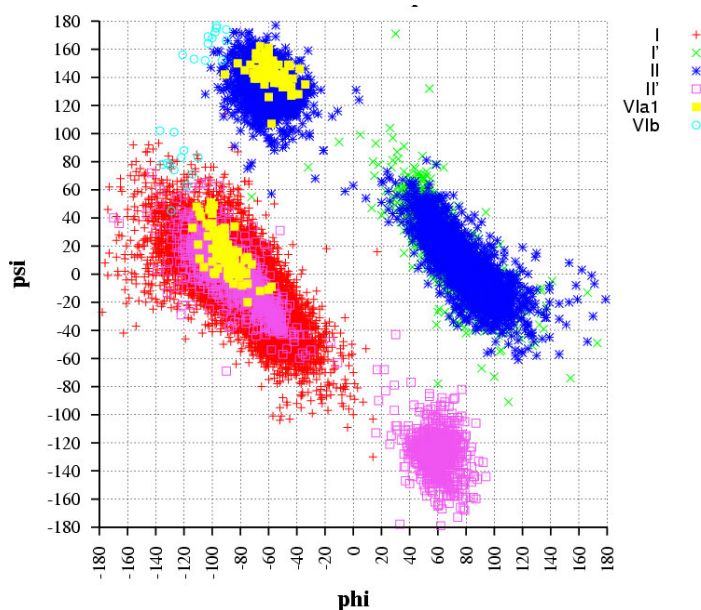
Table 8.9: Mean torsion angle ($\phi_{i+2} - \omega_{i+2}$) and standard deviation for normal β -turn types

Considering the nomenclature based on the region of the Ramachandran plot (see section 6.7.3), the comparison of the β -turn designation with the designation by Wilmot et al. (Table 8.10) revealed no difference, except the new turn-type VIb which was not mentioned previously.

cluster	type	designation	Wilmot et al.
1	I	$\alpha_R\alpha_R$	$\alpha\alpha$
2	I'	$\alpha_L\gamma$	$\gamma\gamma$
3	II	$\beta_P\gamma$	$\beta_P\gamma$
4	II'	$\epsilon\alpha_R$	$\epsilon\alpha$
5	VIa1	$\beta_P\alpha_R$	$\beta\alpha$
6	VIb	$\beta_E\beta_E$	--

Table 8.10: Designation of β -turn types according to Wilmot et al. and described β -turns

Additionally, the Ramachandran plot for normal β -turns shows clearly defined regions for all different turn types (Figure 8.11).

Figure 8.11: Ramachandran plot of inner residues for normal β turn-types

Interestingly, compared to turn-type I and II, the conformation I' and II' (inverse torsion angle ranges of turn-type I and II), could show an additional strong hydrogen bonding between NH_i and CO_{i+3} (Figure 8.12). This supports the theory by Wilmot and Thornton (Wilmot 1988) that, due to steric problems, type I' should be stabilized by an additional hydrogen bond (between NH_i and CO_{i+3}). The typical hydrogen bond between CO_i and NH_{i+3} could not fix this conformation alone.

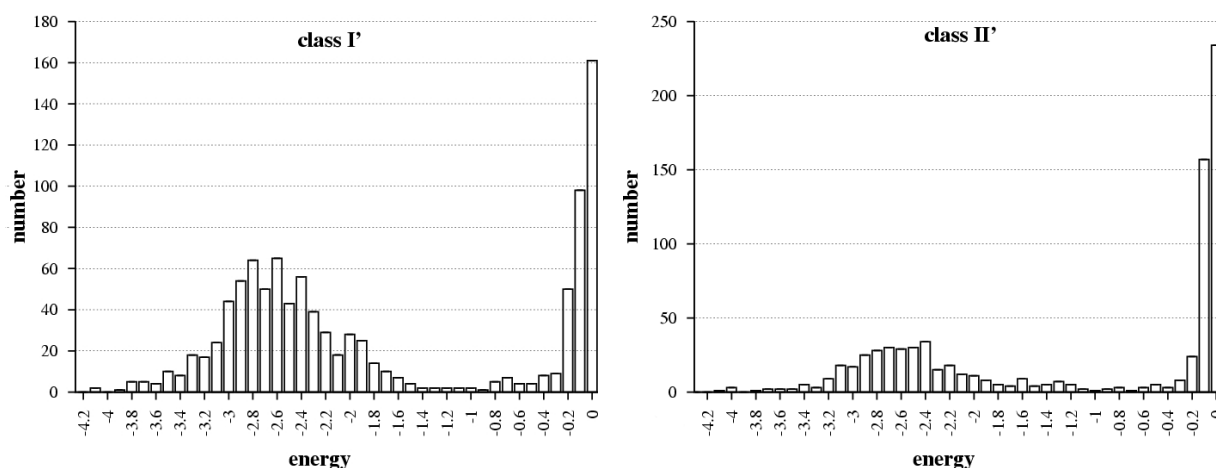


Figure 8.12: NH_i and CO_{i+3} hydrogen bond energy distribution for class I' and II' (DSSP energy shown, -3 kcal/mol represents a strong hydrogen bond)

The overall amino acid propensities (Table 8.11) reveals some interesting information; Pro shows an augmented high occurrence at position 2 and is over represented at position 1. The majority of these Pro residues occur in turn-type I and II, they are lacking a cis-peptide bond. Additionally, Gly is often observed at position 3 and 4 and in contrast, Ile, Val and Pro rarely present at position 3. Pro never occurs at position 4. This is expected as a Pro could never form a hydrogen bond from position 4 to position 1.

There is an extremely high occurrence of Gly at position 3 in turn-type I' and II, at position 2 in class II' and a high occurrence at position 4 in class I. A closer inspection of the Ramachandran plot of normal β -turns (Figure 8.11) clearly indicates the reason for the high occurrence of Gly. The conformation of the residues at position 3 in turn-type I' ($\phi_{i+2}, \psi_{i+2} = \sim 75.3^\circ, \sim 6.9^\circ$) and II ($\phi_{i+2}, \psi_{i+2} = \sim 79.9^\circ, \sim 4.7^\circ$) corresponds to a region that is populated primarily by conformations only accessible for Gly. Position 2 in turn-type II ($\phi_{i+1}, \psi_{i+1} = \sim 59.3^\circ, \sim -128.4^\circ$) represents a region that mostly Gly could occupy.

In contrast, Pro never occurs at position 2 and 3 in class I' and at position 2 in class II'. It shows only reduced occurrence at the other positions. This can be explained by the conformational constraints of the pyrrolidine ring. Since position 2 in turn-type I ($\phi_{i+1}, \psi_{i+1} = -61.2, -25.7^\circ$) and II ($\phi_{i+1}, \psi_{i+1} = -58.5^\circ, 133.4^\circ$) represents typical conformations of Pro it is

not surprising that Pro often occurs at position 2 in these turn-types. As expected, class VIa1 and VIb with a cis-peptide bond shows mostly Pro at position 3.

aa	turn		position 1		position 2		position 3		position 4	
	turn	Pt	pos 1	P1	pos 2	P2	pos 3	P3	pos 4	P4
ALA	8975	1.04	2080	0.96	> 2836	1.31	2088	0.97	< 1971	0.91
ILE	< 4234	0.63	< 1494	0.89	< 943	0.56	< 502	0.30	< 1295	0.77
LEU	< 9366	0.91	2622	1.01	< 2149	0.83	< 2032	0.79	2563	0.99
MET	< 2018	0.91	< 489	0.88	< 468	0.84	< 487	0.88	574	1.03
PHE	< 4241	0.87	1162	0.96	< 781	0.64	< 980	0.81	> 1318	1.09
PRO	> 5475	1.06	> 1856	1.44	> 3134	2.43	< 485	0.38	< 0	0.00
VAL	< 4525	0.56	< 1557	0.77	< 999	0.50	< 527	0.26	< 1442	0.72
ARG	5514	0.96	< 1311	0.92	1468	1.02	1368	0.96	1367	0.95
ASP	> 8674	1.30	> 2406	1.45	> 1865	1.12	> 2707	1.63	1696	1.02
GLU	7842	1.02	< 1696	0.89	> 2516	1.31	> 2091	1.09	< 1539	0.80
LYS	> 7364	1.07	1677	0.97	> 2123	1.23	1700	0.99	> 1864	1.08
ASN	> 6776	1.33	> 1534	1.21	1277	1.00	> 2403	1.89	> 1562	1.23
CYS	< 1679	0.90	478	1.02	< 345	0.74	< 338	0.72	> 518	1.11
GLN	> 4727	1.06	< 994	0.90	1127	1.01	> 1262	1.14	> 1344	1.21
HIS	> 3139	1.12	> 829	1.19	< 607	0.87	> 886	1.27	> 817	1.17
SER	> 8220	1.18	> 1917	1.10	> 2186	1.25	> 2276	1.31	> 1841	1.06
THR	< 4913	0.77	< 1381	0.87	< 967	0.61	< 1187	0.75	< 1378	0.87
TRP	1711	0.99	> 477	1.11	< 390	0.91	410	0.95	434	1.01
TYR	< 3722	0.91	977	0.95	< 689	0.67	< 924	0.90	> 1132	1.10
GLY	>11153	1.41	< 1621	0.82	< 1706	0.86	> 3919	1.98	> 3907	1.98
OTH	< 330	0.79	92	0.89	< 74	0.71	< 77	0.74	87	0.84
ALL	114598	1.00	28650	1.00	28650	1.00	28649	1.00	28649	1.00

Table 8.11: Overall frequencies and conformational parameter for normal β -turn types
(statistical over representation: '>', statistical under representation: '<')

8.4.2 Open β -turns

Open β -turns represents the largest single family of turn structures; 137101 structures were observed without a CO_i and NH_{i+3} hydrogen bond but a $\text{C}\alpha_i - \text{C}\alpha_{i+3}$ distance smaller than 10\AA . These structures cluster into 24 different groups (Figure 8.13), but after visual inspection 7 cluster were detected as not turn-like (nt-1 – nt-7) and additional 6 clusters appeared better described as a kink (VII1a-VII4b; named after Lewis et al. (Lewis 1973), see section 6.4). From the remaining clusters only turn-type I/I', VIa/VIb and VIIIa/VIIIb have related mean torsion angles (Table 8.12) compared to the β -turn types previously described by Hutchinson and Thornton (Hutchinson 1994) and turn-type I/I' and VIa/VIb were already identified within normal β -turns (see section 8.4.1). Turn type IX to XIII are newly discovered and the identification of these turn-types is mainly based on the increase of the $\text{C}\alpha$ - $\text{C}\alpha$ distance.

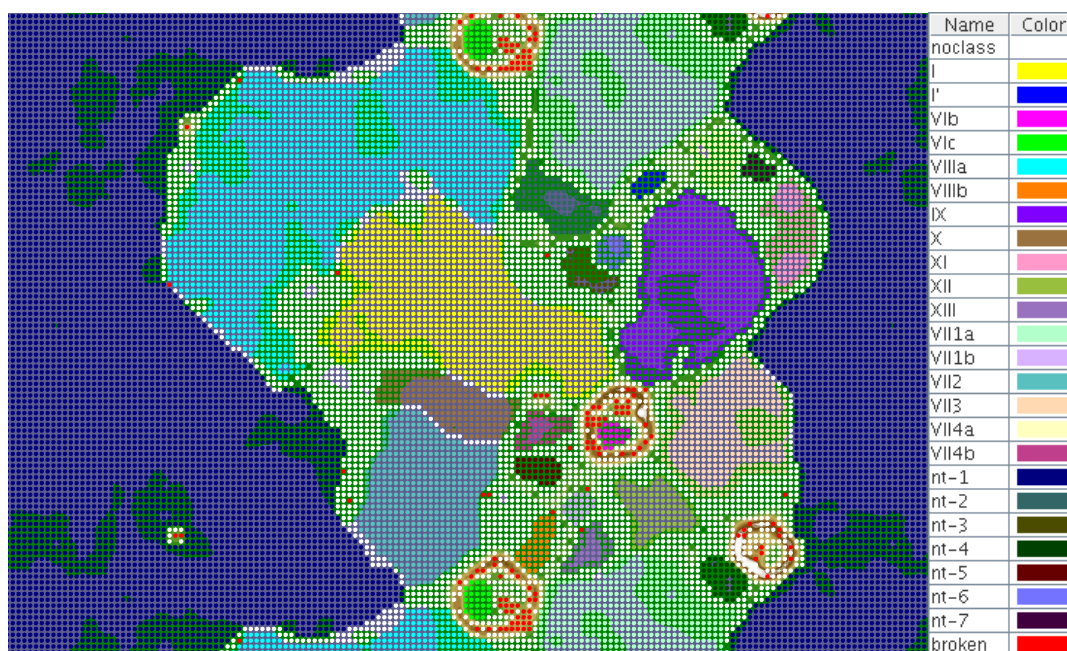


Figure 8.13: Trained ESOM map for open β -turns

cluster	type	nr	ω_i		ϕ_{i+1}		ψ_{i+1}		ω_{i+1}		ϕ_{i+2}		ψ_{i+2}		ω_{i+2}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	17665	-175.7	28.6	-63.4	79.1	29.3	93.3	-176.8	21.9	-65.8	77.9	23.8	96.1	-176.6	24.0
1	I	10966	-179.9	3.1	-79.1	17.7	-32.1	18.8	-178.3	3.7	-83.8	24.7	-25.3	20.1	-179.2	3.4
2	I'	160	178.7	4.7	74.8	20.7	38.5	34.8	177.6	3.6	66.2	15.3	37.7	21.3	179.5	3.1
3	VIb	321	179.2	4.1	-118.6	30.2	137.8	19.3	-0.9	3.6	-74.6	10.5	155.0	13.4	178.6	3.6
4	VIc	313	-0.4	4.2	-77.2	15.5	154.9	14.0	178.6	6.2	-90.6	27.6	138.5	22.0	178.5	5.7
5	VIIIa	18003	-179.6	3.9	-92.1	20.1	-16.0	23.3	-179.6	4.0	-110.5	34.5	134.7	27.2	178.9	4.2
6	VIIIb	447	-179.5	3.3	-95.3	20.0	-16.5	26.0	-180.0	2.9	-126.5	34.6	-166.2	14.4	179.8	2.7
7	IX	6390	-179.7	3.7	-92.1	17.9	-2.3	18.2	179.3	3.7	72.9	18.9	24.3	22.5	179.2	3.5
8	X	1941	-178.6	4.1	-107.7	14.0	8.5	16.8	-178.7	2.7	-58.1	10.3	-37.9	14.5	179.7	2.8
9	XI	1204	-179.9	4.5	-113.2	27.4	148.5	19.7	177.7	4.7	61.0	15.8	39.4	21.1	-179.4	4.8
10	XII	1291	179.6	2.5	69.9	16.8	22.6	21.6	-179.7	2.3	-75.0	17.9	-26.2	18.8	-179.8	2.0
11	XIII	423	179.1	3.1	74.0	18.3	15.7	24.6	-179.7	3.0	-122.4	12.6	-2.8	28.0	179.6	2.5
12	VIIIa	8415	179.1	3.7	75.6	19.4	16.8	25.7	179.3	3.7	-97.9	27.2	141.4	20.2	178.8	4.2
13	VIIIb	241	179.6	3.4	75.5	22.3	20.4	26.1	179.7	3.5	-110.1	25.7	-167.2	11.7	-179.6	4.1
14	VII2	5532	179.5	2.7	-87.7	24.2	116.7	27.5	-179.4	2.5	-92.5	21.9	-21.0	21.8	-180.0	2.1
15	VII3	3938	178.9	3.8	-109.5	35.5	146.8	18.6	179.2	4.1	-84.4	20.1	-19.8	23.4	-179.7	3.4
16	VII4a	629	-178.5	5.1	-93.4	22.3	-12.4	24.3	179.6	3.8	120.1	34.8	159.9	16.6	179.9	4.0
17	VII4b	492	-178.7	4.2	-92.6	23.1	-8.6	22.1	179.6	3.4	81.5	20.2	-155.5	17.3	-178.9	3.6
18	nt-1	55184	178.5	4.6	-102.8	28.9	134.1	21.4	178.6	4.8	-91.8	24.0	128.3	26.2	179.6	4.4
19	nt-2	1144	179.7	3.9	115.0	37.4	-160.3	13.9	-178.7	3.9	-100.5	28.0	136.9	18.9	179.2	4.4
20	nt-3	500	179.0	4.0	-139.9	25.4	-165.3	17.9	179.5	3.8	-101.8	23.2	141.9	17.0	178.7	3.5
21	nt-4	447	179.9	3.0	123.2	34.8	160.1	19.1	-180.0	3.0	-76.1	18.4	137.8	19.3	179.0	3.3
22	nt-5	383	179.2	4.6	-114.7	32.4	141.5	17.2	179.1	3.5	130.4	35.4	-160.2	14.4	-178.8	3.3
23	nt-6	278	179.8	4.1	-92.9	14.5	-160.1	19.3	-179.4	3.3	-73.5	12.4	138.5	19.3	177.7	3.8
24	nt-7	166	178.6	4.2	-97.5	29.2	139.8	21.4	179.8	4.4	106.1	24.2	160.2	16.6	-179.0	3.9

Table 8.12: Mean torsion angle and standard deviation of open β -turn types

Comparing Ramachandran plots of the open and normal β -turns leads to the conclusion that the region for normal β -turns is more limited and specific. Again this can be explained by the missing hydrogen bond and the extended $C\alpha$ - $C\alpha$ distance.

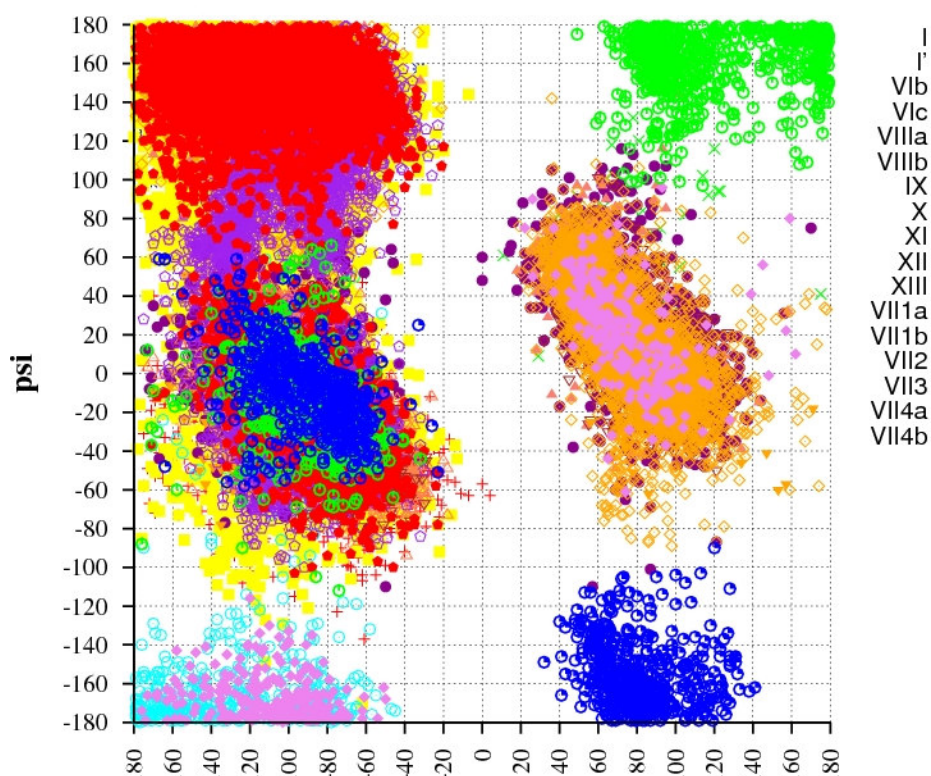


Figure 8.14: Ramachandran plot for open β -turn types

The amino-acid propensities display some similarities between the common classes of the normal and open β -turns, e.g. high occurrence of Gly at position 3 or a low occurrence of Val. In contrast to the high occurrence of Pro at position 2 in normal β -turn type I, Pro shows a low occurrence at position 2 in open β -turn type I. Again, the Ramachandran plot provides an explanation, the region ($\varphi_{i+1}, \psi_{i+1} = -79.1^\circ, -32.1^\circ$) that belongs to position 2 is remote from the typical Pro region. Interestingly, Pro shows also a low frequency at position 4, even though a hydrogen bond is not needed for this turn-type.

The open turn-type I' has a high frequency of Gly at position 3 as observed for the normal turn-type I' but, in contrast, Gly is also common at position 2. The region of the Ramachandran plot of position 2 ($\varphi_{i+1}, \psi_{i+1} = 74.8^\circ, 38.5^\circ$) is shifted towards the typical Gly region compared to the normal turn-type I', which explains the different amino acid propensities. Pro can be found at position 4 in open turn-type I' but not in normal turn-type I' at the same position, since Pro cannot form a hydrogen bond. As expected, both turn types containing a cis peptide bond show Pro almost exclusively at the position of the cis peptide bond (turn-type VIb: position 3, turn-type VIc: position 2).

Since the amino acid propensity compares the amino acid distribution within a turn to the overall amino acid distribution of the whole protein, the overall amino acid propensities for the open β -turn (Table 8.13) seems to have a smaller variance compared to the overall amino acid distribution among proteins as the normal β -turns. Only Gly at position 2 and 3, Asn at position 2 and Pro at position 4 show a significant over representation. Thus, the conclusion could be drawn, that normal β -turns represent a specific part of the polypeptide chain with specific amino-acid distributions whereas open β -turns represents more a conformation of a general sequence, since the local constraint of an hydrogen bond is missing. In general, the differentiation into normal and open β -turns seems reasonable, since the amino acid propensity differs.

aa	turn		position 1		position 2		position 3		position 4	
	f	P	f	P	f	P	f	P	f	P
ALA	<16323	0.89	> 4802	1.05	< 3705	0.81	< 4019	0.88	< 3797	0.83
ILE	<11129	0.79	< 2438	0.69	< 2368	0.67	< 3152	0.89	< 3171	0.89
LEU	<18382	0.84	< 4306	0.79	< 4690	0.86	< 5223	0.95	< 4163	0.76
MET	< 3897	0.83	< 954	0.81	< 895	0.76	< 1097	0.93	< 951	0.81
PHE	< 9108	0.89	< 2024	0.79	< 2163	0.84	2569	1.00	< 2352	0.91
PRO	>11658	1.07	> 3682	1.35	< 2055	0.75	< 1600	0.59	> 4321	1.58
VAL	<13506	0.79	< 2975	0.70	< 2897	0.68	< 3740	0.88	< 3894	0.91
ARG	11829	0.97	> 3232	1.07	< 2789	0.92	< 2884	0.95	< 2924	0.96
ASP	>16523	1.17	> 4349	1.23	> 4760	1.35	3618	1.03	> 3796	1.08
GLU	<15185	0.94	> 4759	1.17	< 3645	0.90	< 3330	0.82	< 3451	0.85
LYS	>15759	1.08	> 4258	1.16	3656	1.00	> 3894	1.07	> 3951	1.08
ASN	>13939	1.29	> 3367	1.25	> 4283	1.59	> 3301	1.22	> 2988	1.11
CYS	3839	0.97	< 891	0.90	974	0.98	1002	1.01	972	0.98
GLN	9271	0.99	2430	1.03	2312	0.98	2306	0.98	< 2223	0.95
HIS	6111	1.03	1428	0.97	> 1638	1.11	> 1572	1.06	1473	1.00
SER	>16407	1.11	> 4446	1.21	> 3912	1.06	3770	1.02	> 4279	1.16
THR	>14142	1.05	3336	0.99	3449	1.03	> 3547	1.06	> 3810	1.13
TRP	< 3106	0.85	< 670	0.73	< 701	0.77	860	0.94	875	0.96
TYR	< 7865	0.91	< 1759	0.81	< 1957	0.90	2186	1.01	< 1963	0.90
GLY	> 24172	1.44	> 4439	1.06	> 7702	1.84	> 6842	1.63	> 5189	1.24
OTH	< 654	0.74	< 161	0.73	< 153	0.69	< 186	0.84	< 154	0.70
ALL	242805	1.00	60706	1.00	60704	1.00	60698	1.00	60697	1.00

Table 8.13: Overall frequencies and conformational parameter for open β -turns

(statistical over representation: '>', statistical under representation: '<')

The $C\alpha$ - $C\alpha$ distance also shows an interesting trend, considering that the cut-off was set to 10 Å. Figure 8.15 shows the $C\alpha$ - $C\alpha$ distance distribution of turn-type VIIIa, which was already described by Hutchinson and Thornton (Hutchinson 1994). The distribution shows a maximum at 7.9 Å in a range between 4.8 Å and 9.6 Å. With a cut-off of 7.5 Å a large fraction of this turn-type would be ignored which also applies for other turn-types. Therefore, the chosen cut-off of 10 Å seems for open β -turns a better choice since it prevents the possibility of ignoring turn structures of similar conformation.

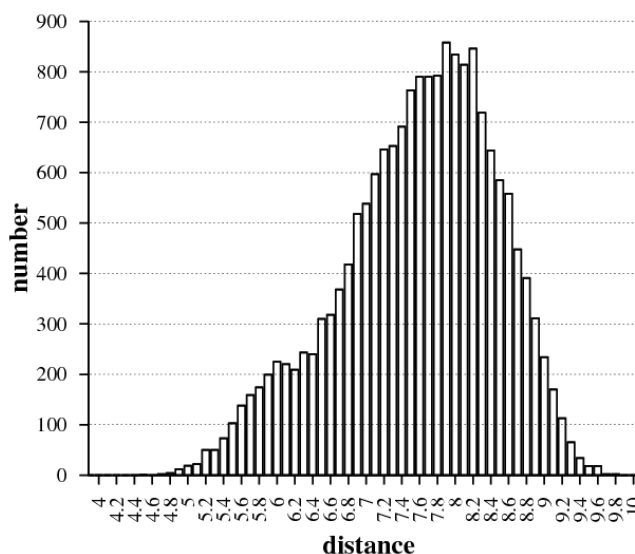


Figure 8.15: $C\alpha$ - $C\alpha$ distance distribution of turn-type VIIIa

8.4.3 Reverse β -turns

A total of 1957 β -turn structures with a reverse hydrogen bond between NH_i and CO_{i+3} were found. Of these structures, 13 different turn-types could be retrieved (Figure 8.16).

As mentioned, the normal β -turn types I' and II' also shows reverse hydrogen bonding. Based on the mean torsion angles of the normal β -turn-type I' and II' these types were clearly identified within this dataset and the corresponding clusters have therefore been named as type I' and II'. Comparing the reverse hydrogen bonding distribution from normal turn type I'/II' and the reverse turn type I'/II' (Figure 8.17) confirms, that the latter are subsets of the normal turn-types. They show also normal hydrogen bonding. Additionally, one cluster has similar mean torsion angles as normal β -turn type VIa1 and is also named type VIa1.

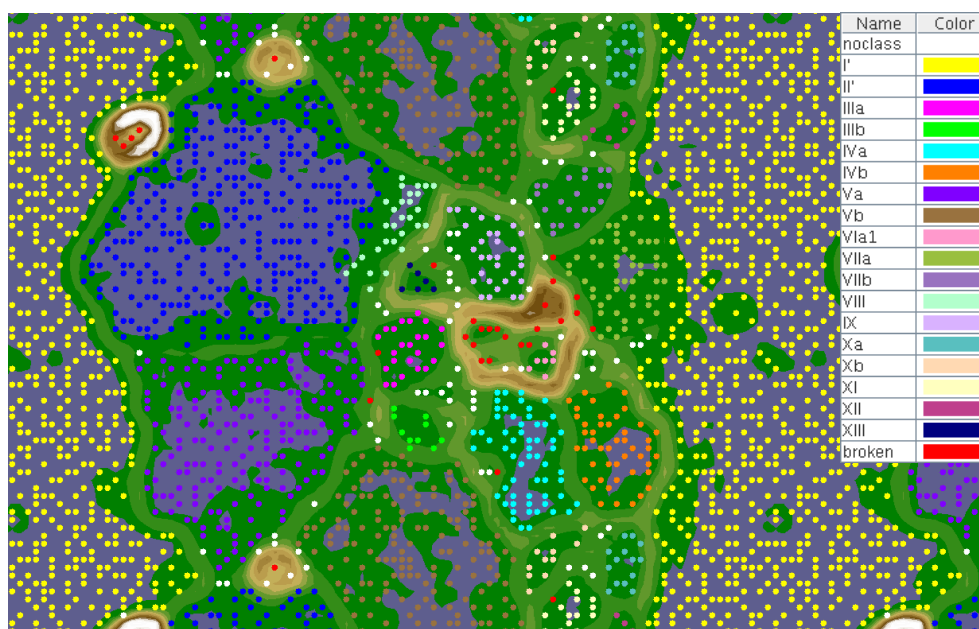


Figure 8.16: Trained ESOM map for reverse β turns

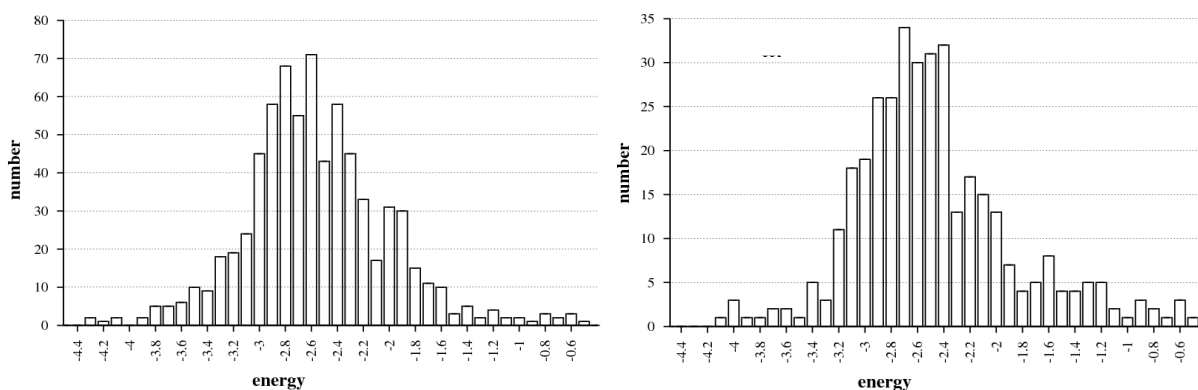


Figure 8.17: Reverse hydrogen bonding energy distribution for reverse β turn types I'/II' (DSSP energy shown, -3 kcal/mol represents a strong hydrogen bond)

cluster	type	nr	ϕ_i		ψ_i		ω_i		ϕ_{i+1}		ψ_{i+1}		ω_{i+1}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	69	-73.6	100.8	43.4	125.9	-179.8	6.7	-20.9	74.6	7.6	76.0	-72.5	37.3
1	I'	721	-133.8	17.3	120.5	15.7	-179.5	3.2	52.8	8.5	43.1	14.5	178.2	2.6
2	II'	360	-130.5	17.0	102.1	20.9	-176.4	4.6	65.2	11.3	-120.6	15.3	179.5	3.6
3	IIIa	28	-118.5	29.2	163.6	11.6	-179.9	1.8	-59.5	7.2	-24.4	12.2	179.9	2.4
4	IIIb	11	-113.5	23.2	-169.1	6.9	-179.2	4.2	-65.1	8.6	-23.8	8.4	-176.8	4.6
5	IVa	64	-133.2	21.0	-168.8	12.3	-179.1	3.3	-64.8	15.2	-25.9	16.5	-179.5	3.1
6	IVb	54	-122.7	21.6	170.9	7.2	179.2	2.9	-58.5	9.8	-27.5	14.4	-179.3	3.2
7	Va	197	-123.4	27.8	164.3	14.8	-179.9	3.0	-57.9	10.5	-31.6	17.5	-178.3	3.3
8	Vb	200	-126.6	24.3	-161.9	25.9	-179.3	4.2	-70.6	20.0	-33.4	24.8	-178.3	4.6
9	VIa1	7	-135.7	17.5	88.6	39.2	-176.0	6.8	-51.9	21.4	133.0	3.5	5.5	5.1
10	VIIa	54	-131.3	21.5	160.0	18.2	179.5	4.1	-56.6	15.0	119.6	12.7	179.3	3.3
11	VIIb	24	-121.9	24.3	-147.2	33.7	179.8	3.3	-83.6	22.7	99.1	16.6	-179.7	3.1
12	VIII	30	-106.6	15.1	9.0	12.6	179.9	5.7	103.6	17.6	-33.8	32.1	178.1	6.2
13	IX	35	91.3	12.1	-13.5	19.0	179.4	3.4	-109.8	18.0	-66.3	16.4	179.0	3.7
14	Xa	14	-126.9	17.9	171.2	8.1	179.3	2.8	-58.4	8.1	-29.1	15.7	-178.8	2.8
15	Xb	15	-115.5	19.7	-167.7	14.1	-178.5	4.5	-69.0	21.6	-29.7	21.4	-179.0	1.3
16	XI	18	-145.8	5.9	116.4	16.6	-178.9	2.5	56.6	9.0	-128.9	16.5	178.4	5.5
17	XII	10	-144.9	30.5	127.8	32.4	178.9	2.6	51.6	4.5	43.9	7.5	178.4	1.3
18	XIII	8	116.2	36.8	157.0	16.8	178.9	2.1	54.5	7.4	46.2	26.6	-179.4	3.4

cluster	type	nr	ϕ_{i+2}		ψ_{i+2}		ω_{i+2}		ϕ_{i+3}		ψ_{i+3}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	69	-35.4	99.4	8.0	48.2	-177.7	10.9	-27.5	135.0	41.6	138.0
1	I'	721	78.2	14.5	2.0	18.3	179.8	3.3	-113.3	19.0	139.5	14.9
2	II'	360	-96.5	16.0	3.7	17.2	-178.7	4.1	-106.3	20.2	139.5	18.2
3	IIIa	28	-98.6	15.4	8.4	25.9	179.3	1.6	142.7	25.7	167.4	13.2
4	IIIb	11	-109.7	22.4	10.9	21.7	178.7	2.7	141.9	24.2	173.1	8.2
5	IVa	64	-94.3	20.0	-0.4	22.3	179.8	3.0	127.7	24.3	-156.9	14.2
6	IVb	54	-95.5	18.1	2.5	20.6	-179.6	3.5	130.3	26.6	-165.1	12.8
7	Va	197	-100.6	21.5	-14.9	25.8	-179.1	3.2	-143.6	21.7	124.8	35.8
8	Vb	200	-111.7	19.8	-11.4	24.2	-177.7	3.9	-139.4	21.8	146.3	26.3
9	VIa1	7	-86.3	15.2	6.4	11.2	179.5	3.1	-98.3	16.9	131.6	27.7
10	VIIa	54	79.0	19.8	0.6	27.5	-179.3	2.5	-140.6	24.1	124.8	30.8
11	VIIb	24	71.7	15.9	10.1	23.6	-179.4	2.3	-138.1	25.5	146.4	24.9
12	VIII	30	-113.2	20.8	-31.9	20.3	-179.1	3.0	-130.8	21.4	149.9	19.4
13	IX	35	-95.0	19.3	-21.8	18.9	178.5	4.9	-137.0	23.8	148.1	19.9
14	Xa	14	-97.2	14.5	-32.0	25.4	-179.0	2.2	-161.8	17.0	-157.5	21.4
15	Xb	15	-110.8	22.4	-27.5	35.0	-179.2	1.8	-149.0	22.0	-169.3	8.1
16	XI	18	-93.5	17.7	7.7	16.6	-180.0	2.3	-130.1	16.0	-152.9	17.2
17	XII	10	70.6	10.9	10.5	18.9	-179.7	2.2	-137.5	24.4	-163.2	14.8
18	XIII	8	89.4	8.8	-9.4	15.7	179.1	2.8	-104.9	26.3	132.5	36.6

Table 8.14: Mean torsion angle and standard deviation for reverse β -turn types

The overall propensities of the reverse β -turns show that there is a high occurrence of Gly, Asp and Asn at position 2 and 3 and an under representation of Ala, Ile, Leu, Met, Phe, Pro and Val.

In contrast to the normal β -turn-types I'/II', Pro occurs very rarely at any position among these classes (only low occurrence at position 3 in turn-type II'). Additionally, turn-types I'/II' are dominated by the highly over representation of Gly, Asp and Asn. They seem to occur as pairs of neighbored residues: Gly at position 3 and Asp/Asn at position 2 among turn-type class I' and vice versa in turn-type class II'. Pro occurs with a cis-peptide bond in turn-type VIa1 and within other turn-types with a high occurrence of Pro at position 2 (IIIa, IVb and Va) without a cis-peptide bond.

The conformation of the reverse β -turns, especially the turn-types I'/II', are within specific region of the Ramachandran plot (Figure 8.18). Potentially, the reverse hydrogen bond and additionally interactions and specific features of Gly, Asp, and Asn are needed to stabilize this conformation. In contrast, Ala, Ile, Leu, Met, Phe, and Val lack these features, thus hydrophobic properties seems to be unfavoured and the reverse β -turns are appropriate to expose charged or polar amino-acid sidechains.

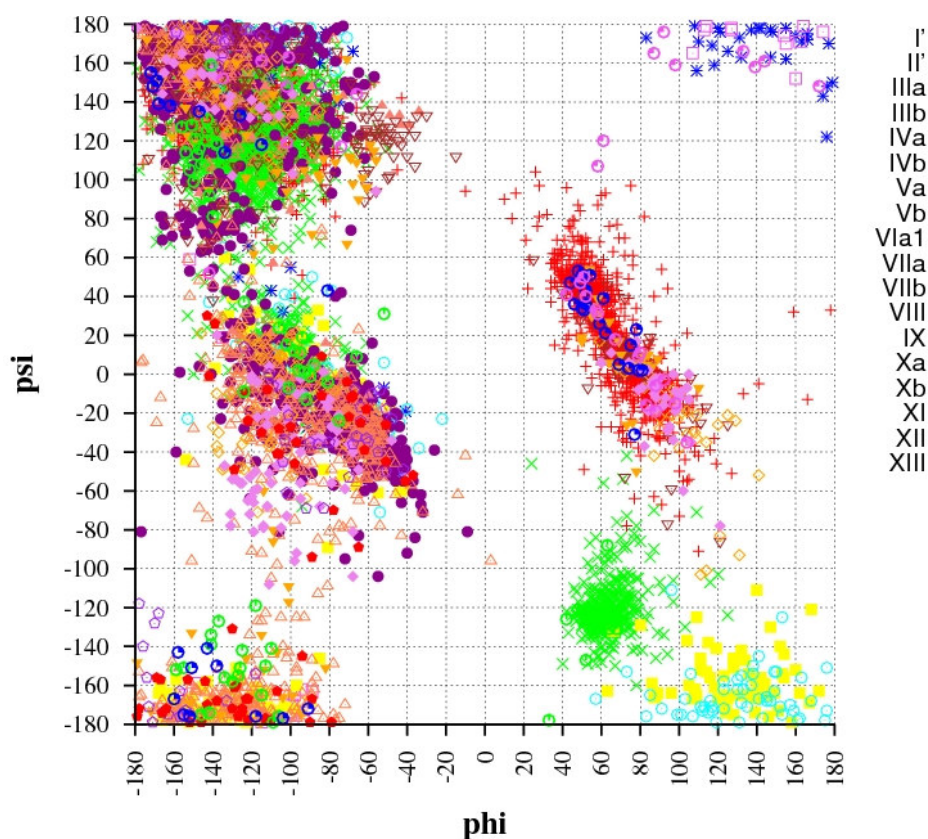


Figure 8.18: Ramachandran plot for reverse β -turn types

aa	turn		position 1		position 2		position 3		position 4	
	f	P	f	P	f	P	f	P	f	P
ALA	< 310	0.54	< 79	0.55	< 83	0.57	< 53	0.37	< 95	0.66
ILE	< 250	0.56	119	1.06	< 22	0.20	< 16	0.14	93	0.83
LEU	< 249	0.36	< 105	0.61	< 28	0.16	< 40	0.23	< 76	0.44
MET	< 76	0.51	< 19	0.51	< 14	0.38	< 10	0.27	33	0.89
PHE	< 209	0.64	92	1.13	< 27	0.33	< 30	0.37	< 60	0.74
PRO	< 140	0.40	< 0	0.00	> 122	1.41	< 18	0.21	< 0	0.00
VAL	< 394	0.73	> 181	1.34	< 36	0.27	< 27	0.20	150	1.11
ARG	325	0.85	85	0.89	< 60	0.63	< 67	0.70	113	1.18
ASP	> 734	1.65	> 147	1.32	> 289	2.59	> 242	2.17	< 56	0.50
GLU	547	1.07	109	0.85	> 150	1.17	136	1.06	> 152	1.18
LYS	> 580	1.25	111	0.96	120	1.04	104	0.90	> 245	2.12
ASN	> 680	1.99	> 109	1.28	> 284	3.33	> 189	2.22	98	1.15
CYS	< 62	0.49	26	0.83	< 8	0.26	< 8	0.26	< 20	0.64
GLN	303	1.02	74	1.00	63	0.85	60	0.81	> 106	1.43
HIS	195	1.04	58	1.24	35	0.75	37	0.79	> 65	1.39
SER	472	1.01	> 161	1.38	< 92	0.79	114	0.98	105	0.90
THR	407	0.96	> 173	1.63	< 31	0.29	88	0.83	115	1.08
TRP	< 63	0.55	< 16	0.55	< 14	0.49	< 13	0.45	20	0.69
TYR	246	0.90	> 106	1.54	< 22	0.32	< 34	0.50	84	1.22
GLY	> 1418	2.68	146	1.10	> 416	3.14	> 632	4.77	> 224	1.69
OTH	15	0.54	3	0.43	3	0.43	< 1	0.14	8	1.15
ALL	7676	1.00	1919	1.00	1919	1.00	1919	1.00	1919	1.00

Table 8.15: Overall frequencies and conformational parameters for reverse β -turns

(statistical over representation: '>', statistical under representation: '<')

8.5 Five residue turns

8.5.1 Normal α -turns

Most normal α -turns (91053 of 91726) reside in a single α helical class. The majority of these α -turns form α helices and show a sharp distribution around the mean torsion angles (Figure 8.20). The remaining turns cluster into seven different clusters and contain one class with cis-peptide bonds (Figure 8.19 and Table 8.16).



Figure 8.19: Trained ESOM map of normal α -turns

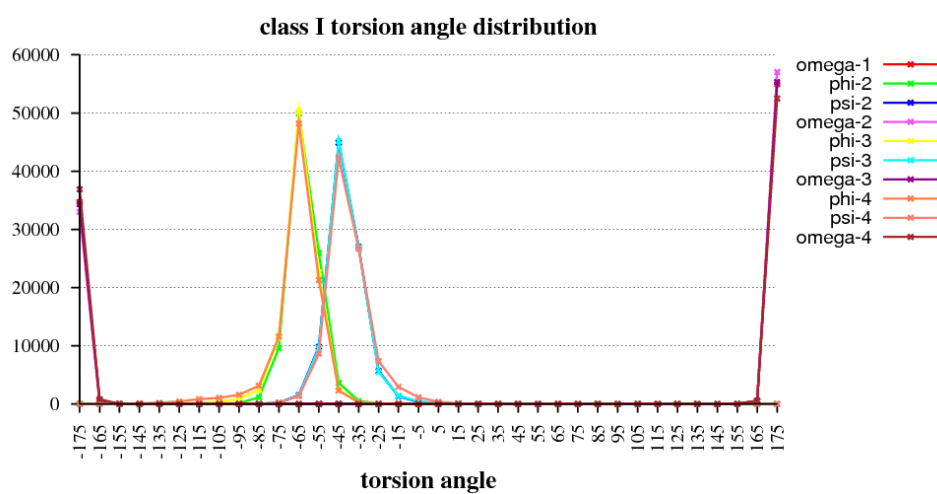
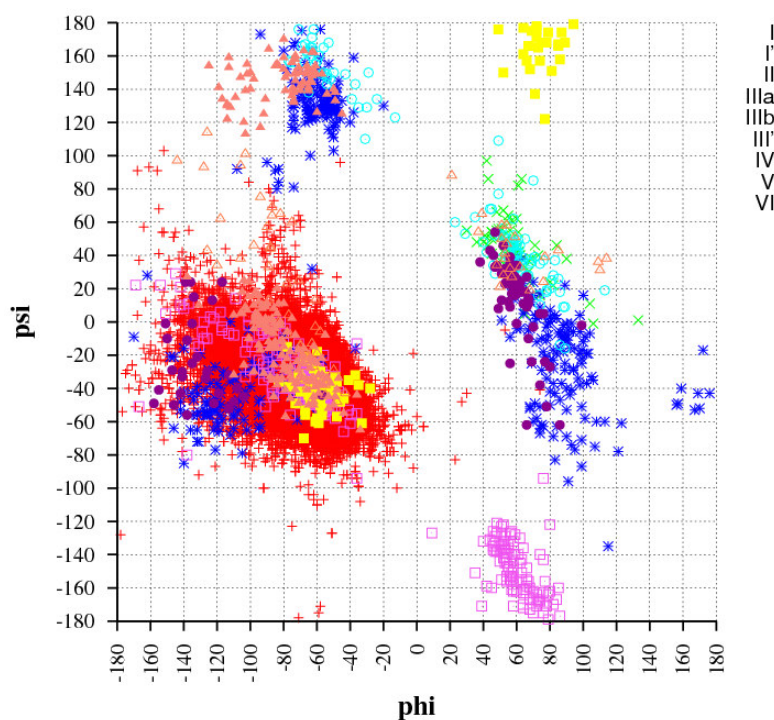


Figure 8.20: Torsion angle distribution for normal α -turn type I

cluster	type	nr	ω_i		ϕ_{i+1}		ψ_{i+1}		ω_{i+1}		ϕ_{i+2}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	164	-177.4	20.2	-42.3	46.0	-20.0	72.7	-168.1	45.3	-58.4	52.1
1	I	91053	179.5	2.9	-62.4	7.4	-41.5	9.1	179.3	2.6	-64.2	9.1
2	I'	24	179.7	1.7	52.3	9.2	52.2	14.1	179.2	2.7	57.7	10.2
3	II	136	-179.8	4.8	-60.8	11.7	131.8	18.8	-179.3	3.7	88.9	24.0
4	IIIa	114	176.7	5.6	60.2	11.9	-148.6	16.0	-177.4	4.1	-71.9	20.9
5	IIIb	26	-178.6	4.5	74.3	10.5	163.1	14.3	179.1	5.0	-55.6	12.1
6	III'	40	179.4	3.7	-55.8	15.0	153.4	14.6	179.2	3.2	65.3	15.0
7	IV	29	179.5	5.3	56.3	6.4	27.7	10.4	179.4	5.7	65.6	12.1
8	V	24	-178.1	3.9	-65.3	10.6	-17.4	17.3	179.1	3.1	-98.4	20.5
9	VI	75	178.0	4.6	-80.4	19.7	144.9	11.7	2.0	5.0	-90.4	9.2

cluster	type	nr	ψ_{i+2}		ω_{i+2}		ϕ_{i+3}		ψ_{i+3}		ω_{i+3}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	164	10.8	68.5	-168.6	37.3	-20.8	97.8	-14.4	46.3	179.6	5.0
1	I	91053	-41.4	9.4	179.5	2.9	-65.9	12.3	-40.0	10.7	179.7	3.0
2	I'	24	48.6	16.7	177.3	4.4	68.6	21.5	35.5	17.0	179.1	3.3
3	II	136	-23.7	30.5	179.9	4.8	-102.7	28.0	-42.4	17.2	-180.0	3.7
4	IIIa	114	-33.0	15.9	-179.7	4.1	-74.6	28.2	-33.2	19.5	179.2	3.2
5	IIIb	26	-36.7	8.6	-179.8	1.5	-61.0	10.0	-47.0	12.2	-179.2	3.1
6	III'	40	35.5	22.2	178.6	4.3	59.0	13.6	43.2	13.1	179.6	2.8
7	IV	29	-0.2	28.6	-175.4	4.8	-130.0	20.2	-25.0	25.7	-178.4	5.7
8	V	24	52.4	35.9	178.4	3.9	57.5	34.9	38.6	22.8	179.2	3.4
9	VI	75	-3.9	15.3	-176.7	4.9	-68.6	17.4	-32.3	15.9	-179.4	4.0

Table 8.16: Mean torsion angle and standard deviation for normal α -turnsFigure 8.21: Ramachandran plot for normal α -turn types

When comparing this classification (Table 8.17) with that of Natarja et al. (Nataraj 1995) partial agreement is observed. Natarja et al. did not describe turn-type V and VI. In contrast, Dasgupta et al. (Dasgupta 2004) described turn-types that highly differ from this classification; only turn-type II is similar to the current results. A possible explanation is the use of a distance criterion rather than the hydrogen bond.

cluster	type	nr	Nataraj et al.		Dasgupta et al	
			own	origin	own	origin
1	I	91053	$\alpha_R\alpha_R\alpha_R$	F1	AAA	(AAA)
2	I'	24	$\alpha_L\alpha_L\alpha_L$	F1'	rrr	
3	II	136	$E\alpha_L\alpha_R$	F2	PgA	PgA
4	IIIa	114	$E'\alpha_R\alpha_R$	g3	pAA	(pAA)
5	IIIb	26	$E'\alpha_R\alpha_R$		pAA	(pAA)
6	III'	40	$E\alpha_L\alpha_L$	g1	Prr	
7	IV	29	$\alpha_L\alpha_L\alpha_R$	g6	raA	
8	V	24	$\alpha_R P\alpha_L$		Adr	
9	VI	75	$E\alpha_R\alpha_R$		PAA	

Table 8.17: Designation after Dasgupta (Dasgupta 2004) and Nataraj (Nataraj 1995) and comparison to obtained turn-types

(empty boxes: not identical, brackets: similar designations, but different mean torsion angles)

The amino-acid propensities (see appendix) present only limited information. Pro and Gly are rarely found in turn-type I, especially at position 3, 4 and 5, but Gly often occurs at position 3 in turn-type II and position 2 in turn-type IIIa/IIIb. Pro occurs at position 2 in turn-type II without a cis peptide bond. This could be explained by the specific feature of Pro and the possibility to form conformations that are not allowed or not likely for the other amino acids. This is in contrast to turn-type VI where Pro is found predominantly at position 2. This turn-type requires a cis-peptide bond at this position.

The overall amino acid propensity (Table 8.18) is dominated by turn-type I and shows that Ala is over represented while Pro and Gly are under represented throughout the entire normal α -turn family. The low frequency of Pro is expected since α helices are formed mainly by α -helical turns and Pro cannot form the main-chain hydrogen bonds that are needed to build α -helices.

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	>49844	1.44	> 8983	1.30	>10076	1.46	>10147	1.47	>10268	1.48	>10370	1.50
ILE	>29401	1.10	5439	1.02	> 5786	1.08	> 5981	1.12	> 6152	1.15	> 6043	1.13
LEU	>55109	1.33	> 9735	1.18	>10724	1.30	>10932	1.32	>11737	1.42	>11981	1.45
MET	>11646	1.31	> 2057	1.16	> 2258	1.27	> 2325	1.31	> 2465	1.38	> 2541	1.43
PHE	19124	0.98	< 3604	0.93	3834	0.99	3768	0.97	3942	1.01	3976	1.02
PRO	< 5982	0.29	< 2423	0.59	< 2150	0.52	< 996	0.24	< 413	0.10	< 0	0.00
VAL	<29797	0.93	< 5755	0.89	< 6040	0.94	< 6031	0.94	< 6136	0.95	< 5835	0.91
ARG	>27653	1.21	> 4996	1.09	> 5348	1.17	> 5567	1.21	> 5812	1.27	> 5930	1.29
ASP	<23352	0.88	> 5941	1.12	< 4732	0.89	< 4807	0.90	< 4164	0.78	< 3708	0.70
GLU	>38772	1.26	> 7750	1.26	> 8306	1.35	> 8467	1.38	> 7566	1.23	> 6683	1.09
LYS	>30810	1.12	5479	0.99	> 5975	1.08	> 6325	1.15	> 6417	1.16	> 6614	1.20
ASN	<16238	0.80	< 3626	0.89	< 2874	0.71	< 3049	0.75	< 3114	0.76	< 3575	0.88
CYS	< 5951	0.79	< 1172	0.78	< 1097	0.73	< 1151	0.77	< 1250	0.83	< 1281	0.86
GLN	>23039	1.30	> 4417	1.24	> 4620	1.30	> 4742	1.33	> 4698	1.32	> 4562	1.28
HIS	< 9797	0.88	< 1950	0.87	< 1858	0.83	< 1940	0.87	< 1977	0.89	< 2072	0.93
SER	<21920	0.79	< 5249	0.94	< 4172	0.75	< 4060	0.73	< 4135	0.74	< 4304	0.77
THR	<20129	0.79	< 4766	0.94	< 3845	0.76	< 3771	0.74	< 4011	0.79	< 3736	0.74
TRP	6886	1.00	1380	1.00	> 1455	1.06	1373	1.00	1362	0.99	1316	0.96
TYR	<15183	0.93	< 2912	0.89	< 3037	0.93	< 2963	0.90	< 3107	0.95	< 3164	0.96
GLY	<16002	0.51	< 3734	0.59	< 3151	0.50	< 2936	0.46	< 2580	0.41	< 3601	0.57
OTH	1778	1.07	317	0.95	346	1.04	353	1.06	> 374	1.12	> 388	1.17
ALL	458413	1.00	91685	1.00	91684	1.00	91684	1.00	91680	1.00	91680	1.00

Table 8.18: Overall frequencies and conformational parameter for normal α -turns
(statistical over representation: '>', statistical under representation: '<')

8.5.2 Open α -turns

The 19607 turn structures with a $C\alpha_i - C\alpha_{i+4}$ distance smaller than 10\AA but without a hydrogen bond between CO_i and NH_{i+4} can be grouped into 22 different clusters. One turn-type (kink-1) agrees better with a kink than a turn and one turn-type (XVIII) contains a cis peptide bond (Figure 8.22, Table 8.19).

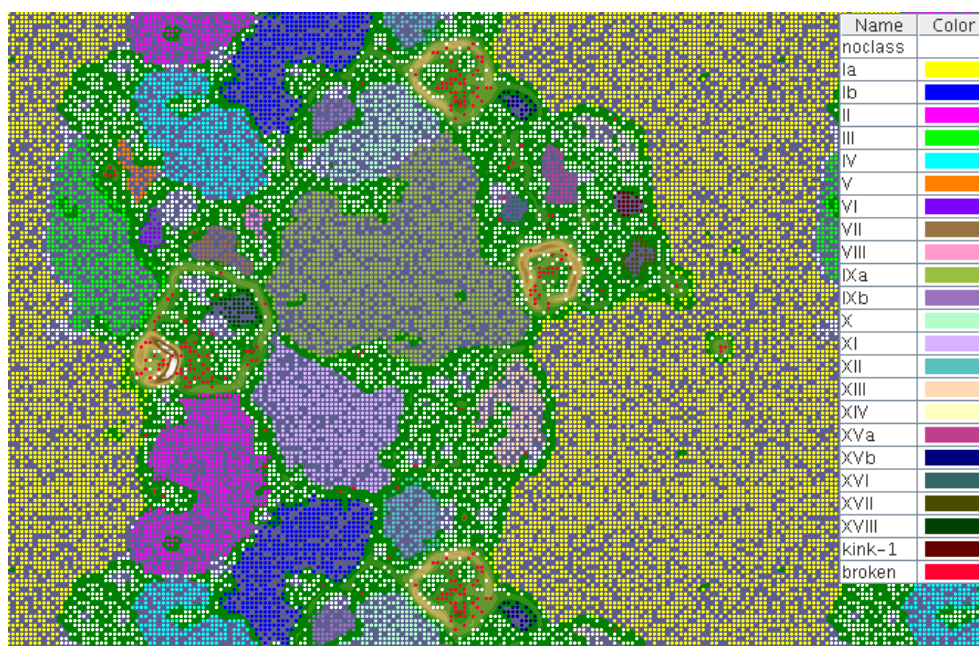


Figure 8.22: Trained ESOM map of open α turns

cluster	type	nr	ω_i		ϕ_{i+1}		ψ_{i+1}		ω_{i+1}		ϕ_{i+2}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	3308	-178.0	21.5	-55.1	89.0	29.9	113.4	-176.6	24.8	-51.3	83.5
1	Ia	8359	178.8	4.4	-95.8	33.0	144.7	24.2	-179.6	4.0	-59.0	9.7
2	Ib	671	179.1	3.9	-113.1	30.6	-170.1	10.5	-178.3	3.8	-63.9	12.7
3	II	837	-178.2	3.7	-91.7	22.2	-39.4	25.4	-177.0	4.2	-89.1	30.3
4	III	694	179.5	4.1	-102.2	36.6	149.4	16.6	178.7	4.1	-71.0	16.5
5	IV	602	179.7	3.4	-73.8	15.4	-34.2	16.1	-178.6	3.7	-87.3	19.3
6	V	74	-179.9	3.1	-90.0	21.0	-10.2	20.4	179.4	2.9	-122.9	30.6
7	VI	36	179.9	3.7	-109.8	15.5	-60.5	26.4	-179.8	3.0	-89.0	14.0
8	VII	117	-179.9	3.1	-90.1	17.4	-1.4	28.3	179.8	3.0	-110.5	31.8
9	VIII	44	-179.9	3.5	-92.0	21.4	-10.0	20.7	178.5	3.8	-85.4	19.7
10	IXa	2248	179.1	4.0	-95.9	30.8	141.3	23.0	178.8	4.0	-65.5	20.5
11	IXb	71	-179.7	3.4	-87.4	15.6	-169.2	10.6	-179.8	2.6	-68.6	18.5
12	X	406	-178.5	3.9	-95.2	19.3	-13.8	23.1	-179.3	3.6	-66.6	19.6
13	XI	932	-180.0	2.8	-66.2	13.7	-20.2	18.2	179.3	3.2	-93.4	18.0
14	XII	222	178.3	4.5	68.1	18.7	-142.7	20.8	-178.2	3.1	-64.4	15.3
15	XIII	292	179.7	4.2	109.2	28.8	152.3	22.7	-178.5	4.3	-61.2	11.6
16	XIV	88	-179.1	3.8	-112.5	25.8	147.9	19.5	177.1	4.6	56.4	9.7
17	XVa	94	178.9	3.2	-118.8	30.1	142.9	24.1	-180.0	2.8	90.7	22.3
18	XVb	40	179.7	3.4	-106.1	27.4	144.5	21.5	179.2	3.9	95.5	14.3

19	XVI	40	179.9	2.9	-63.7	11.8	146.6	13.8	178.1	3.8	56.4	12.2
20	XVII	68	-179.5	2.7	-95.1	20.3	-6.4	18.5	-179.9	2.4	90.9	17.9
21	XVIII	82	179.9	3.4	-122.9	21.8	144.8	19.5	179.0	3.1	-132.9	14.1
22	kink-1	39	-178.1	3.7	-71.8	12.8	-31.1	13.5	-178.2	3.8	56.8	6.9

cluster	type	nr	Ψ_{i+2}		ω_{i+2}		Φ_{i+3}		Ψ_{i+3}		ω_{i+3}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	3308	10.3	106.1	-166.1	48.9	-44.9	81.8	38.9	73.1	-179.4	9.9
1	Ia	8359	-31.6	14.0	-180.0	3.0	-73.4	17.8	-20.6	22.4	179.4	3.4
2	Ib	671	-31.6	16.1	-180.0	3.0	-75.0	20.5	-22.3	28.1	179.9	2.5
3	II	837	-49.1	25.2	-178.0	3.9	-85.3	27.0	-30.7	25.8	-179.4	2.8
4	III	694	-33.9	19.2	178.5	4.4	-116.8	22.0	96.7	30.9	-178.5	3.5
5	IV	602	-36.2	17.5	-178.9	3.7	-137.4	18.5	78.9	24.8	-177.3	3.8
6	V	74	-163.1	13.3	-180.0	2.7	-65.8	16.6	129.1	15.3	-179.8	1.9
7	VI	36	-27.3	17.1	179.8	4.6	-149.2	18.2	152.0	12.8	179.1	3.8
8	VII	117	162.7	10.2	178.2	3.5	-55.2	9.2	128.2	12.3	-178.4	3.0
9	VIII	44	160.5	10.5	178.4	3.8	-76.6	13.6	80.9	17.7	-179.2	2.3
10	IXa	2248	136.3	13.2	179.1	3.2	77.8	19.7	9.8	26.0	179.8	3.5
11	IXb	71	140.5	13.4	179.2	2.5	71.4	15.4	15.3	20.7	178.8	3.3
12	X	406	129.4	11.5	179.5	3.0	70.5	15.9	14.2	22.6	-179.8	2.8
13	XI	932	7.5	16.7	177.8	3.6	93.2	19.7	8.5	17.6	179.1	3.0
14	XII	222	-25.0	18.2	179.8	3.6	-79.4	20.3	-10.2	25.1	179.8	3.3
15	XIII	292	-33.8	13.9	-179.4	3.2	-72.1	18.5	-25.4	23.1	179.1	2.6
16	XIV	88	-129.3	13.2	-179.4	3.2	-82.3	19.7	5.3	27.5	179.1	3.5
17	XVa	94	-157.6	12.7	-179.4	2.8	-74.2	18.2	130.9	15.1	179.9	2.6
18	XVb	40	167.5	11.3	-179.7	3.1	-68.1	15.9	124.1	21.4	-179.1	2.6
19	XVI	40	31.5	16.9	179.2	2.5	73.4	11.5	12.0	12.2	179.4	2.8
20	XVII	68	-1.6	16.4	-179.8	2.6	-109.0	24.4	-13.2	29.0	179.7	2.7
21	XVIII	82	129.7	17.6	-1.0	3.8	-72.0	8.1	158.4	10.2	179.4	2.8
22	kink-1	39	-128.4	10.5	-179.9	2.8	-84.0	16.2	1.6	16.0	179.1	3.7

Table 8.19: Mean torsion angle and standard deviation for open α -turns

Not surprisingly, the designation of Dasgupta et al. (Dasgupta 2004), who also used a distance criterion, fits better than the hydrogen bonded based classes of Nataraj et al. (Nataraj 1995) (Table 8.20).

cluster	type	nr	own	Dasgupta et al.	own	Nataraj et al.
1	Ia	8359	PAA	PAA	EaRaR	
2	Ib	671	EAA		[E]aRaR	
3	II	837	AAA	AAA	aRaRaR	[aRaRaR] F1
4	III	694	PAE		EaR[E]	
5	IV	602	AAD	AAD	aRaR[E]	
6	V	74	AEP		aR[E][E]	
7	VI	36	AAB	[AAE]	aRaR[E]	
8	VII	117	APP		aRP[E]	
9	VIII	44	APD		aRP[E]	
10	IXa	2248	PPa	PPa	EPaL	
11	IXb	71	PPa		[E]PaL	
12	X	406	APa		aRPaL	
13	XI	932	AAa	AAa	aRaRaL	aRaRaL #
14	XII	222	pAA	pAA	EaRaR	EaRaR g3
15	XIII	292	eAA		[E']aRaR	
16	XIV	88	EpA	[BpA]	E[P']aR	
17	XVa	94	EpP		E[P'] [E]	
18	XVb	40	PpP		E[P'] [E]	
19	XVI	40	Pra	[Pra]	EaLaL	EaLaL g1
20	XVII	68	AaA		aRaLaR	
21	XVIII	82	EBP		EP[E]	
22	kink-1	39	ApA		aR[P']aR	

Table 8.20: Designation of open α -turns after Dasgupta et al. (Dasgupta 2004) and Nataraj et al. (Nataraj 1995) and comparison with obtained classes (empty boxes: not identical, brackets: similar designations, but different mean torsion angles)

The missing hydrogen bond and the larger C α -C α distance of 10Å leads to a Ramachandran plots (Figure 8.23) with a wide range of occupied regions especially if the Ramachandran plots of the normal and open α -turns are compared. The regions for possible torsion angles of amino acids in normal α -turns are much smaller and more specific.

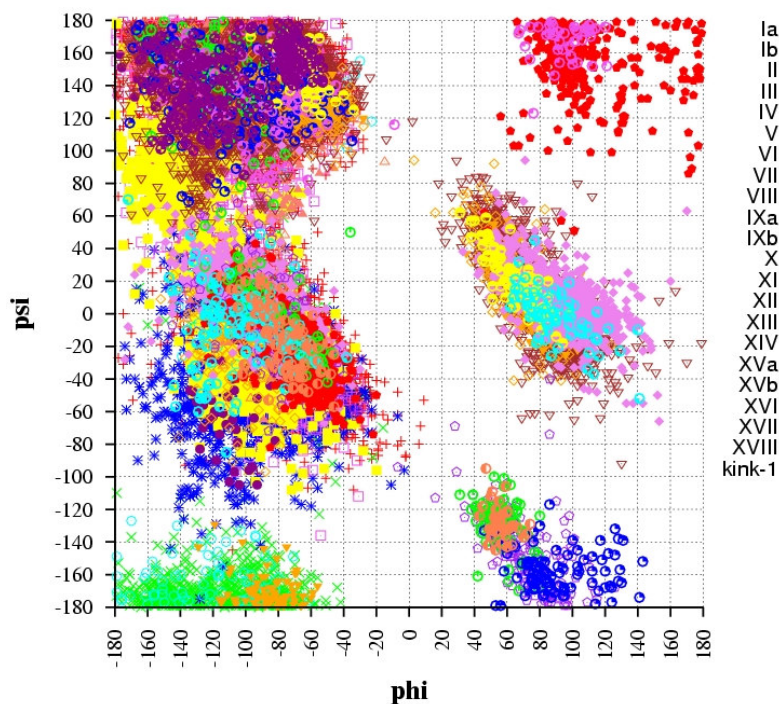


Figure 8.23: Ramachandran plot of open α turn-types

Analysing the overall amino acid propensities (Table 8.21) reveals that Ile, Leu, Met, Phe and Val show an under representation at position 2 and 4 which stays in contrast to the overall amino acid propensities of the normal α -turns. Additionally, Pro shows a high occurrence at position 3, Ser, Thr, Asp, Asn, Pro at position 2 and Gly, Asp, Asn at position 4.

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	aa	f	P	f	P	f
ALA	< 5598	0.93	< 1022	0.85	< 807	0.67	> 1361	1.13	< 1004	0.83	> 1404	1.16
ILE	< 3048	0.65	> 1222	1.31	< 377	0.40	< 622	0.67	< 306	0.33	< 521	0.56
LEU	< 5281	0.73	> 1894	1.31	< 693	0.48	< 1113	0.77	< 609	0.42	< 972	0.67
MET	< 1209	0.78	> 394	1.27	< 159	0.51	< 187	0.60	< 160	0.51	309	0.99
PHE	< 2756	0.81	> 810	1.19	< 351	0.52	< 480	0.71	< 456	0.67	659	0.97
PRO	> 5680	1.57	> 810	1.12	> 1291	1.79	> 2510	3.48	< 284	0.39	> 785	1.09
VAL	< 3902	0.69	> 1401	1.25	< 472	0.42	< 785	0.70	< 362	0.32	< 882	0.79
ARG	< 3300	0.82	< 740	0.92	< 659	0.82	752	0.94	< 610	0.76	< 539	0.67
ASP	> 6711	1.44	< 709	0.76	> 1736	1.87	> 1063	1.14	> 1794	1.93	> 1409	1.51
GLU	> 6079	1.13	< 722	0.67	< 808	0.75	> 1426	1.33	> 1673	1.56	> 1450	1.35
LYS	< 4346	0.90	< 867	0.90	< 854	0.89	> 1117	1.16	< 799	0.83	< 709	0.74
ASN	> 4270	1.20	< 563	0.79	> 1192	1.68	< 531	0.75	> 1299	1.83	685	0.96
CYS	1188	0.91	> 311	1.19	241	0.92	< 180	0.69	< 160	0.61	> 296	1.13
GLN	2906	0.94	< 471	0.76	< 458	0.74	< 554	0.89	< 554	0.89	> 869	1.40
HIS	1999	1.02	406	1.04	415	1.06	< 329	0.84	> 431	1.11	418	1.07
SER	> 6216	1.28	< 791	0.81	> 2190	2.25	979	1.00	> 1160	1.19	> 1096	1.13
THR	> 4757	1.07	865	0.98	> 1701	1.92	< 636	0.72	< 684	0.77	871	0.98
TRP	1075	0.89	268	1.11	< 135	0.56	< 197	0.82	241	1.00	234	0.97
TYR	< 2442	0.85	> 622	1.09	< 354	0.62	< 467	0.82	< 411	0.72	588	1.03
GLY	> 7113	1.29	1061	0.96	1094	0.99	< 694	0.63	> 2991	2.71	> 1273	1.15
OTH	< 209	0.72	68	1.17	< 30	0.52	< 34	0.59	< 29	0.50	48	0.83
ALL	80085	1.00	16017	1.00	16017	1.00	16017	1.00	16017	1.00	16017	1.00

Table 8.21: Overall frequencies and conformational parameter for open α -turns

(statistical over representation: '>', statistical under representation: '<')

A detailed analysis of the amino acid propensities reveals a higher occurrence of Asp, Asn, Ser, Thr at position 2 in turn-type Ia and Ib since these amino acids are the only ones that show conformations with a ϕ torsion angle around $\pm 180^\circ$. Interestingly, Pro has a higher occurrence at position 2 and 3 in turn-type Ia and, in contrast, Gly has a higher occurrence at position 2 in turn-type Ib. Since the ϕ torsion angle at this position is the only difference between these two turn-types, it should be the reason for the different amino acid propensities. A closer look (Figure 8.24) reveals that torsion angles at position 2 in turn-type Ib are lying in a region of the Ramachandran plot which only Gly can realize. Presumably, the conclusion could be drawn, that based on the available conformational space, Pro at position 2 induces a ϕ torsion angle smaller 180° whereas a Gly at this position reveals a torsion angle bigger than 180° .

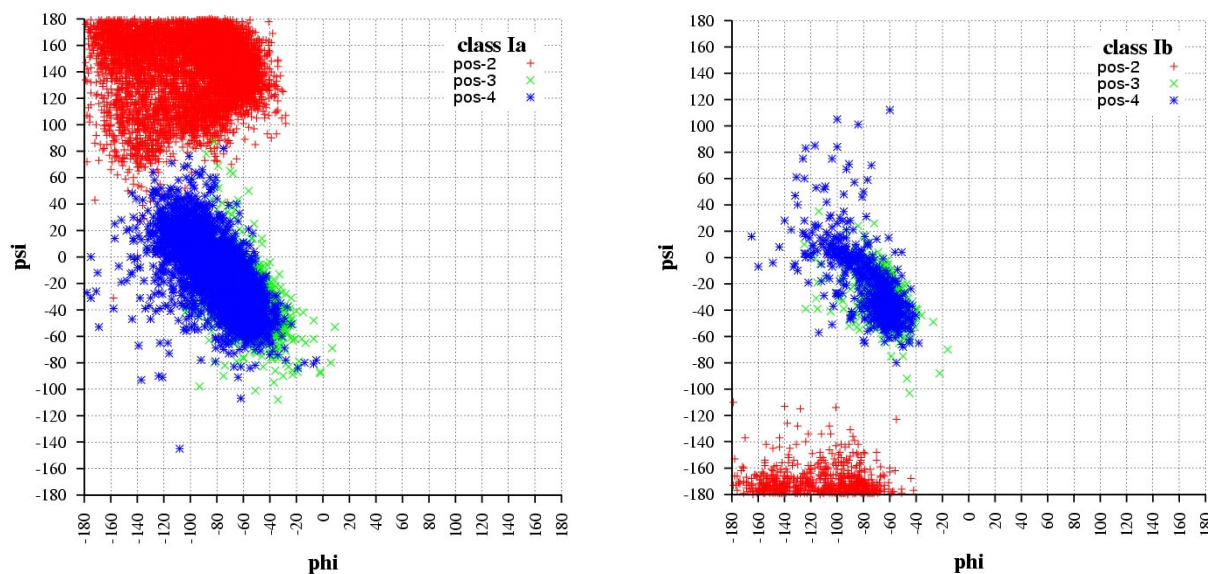


Figure 8.24: Ramachandran plot of open α turn-types Ia/Ib

Another interesting turn-type containing enough structures for analysis of the amino acid propensities is turn-type IXa/IXb. Both turn-types show a higher occurrence of Gly at position 2 but turn-type IXa has a higher occurrence of Pro at position 2 and 3. In contrast, Gly has a higher occurrence at position 2 in turn-type IXb. This is similar behaviour to turn-type Ia and Ib.

Furthermore, the increase of the $C\alpha$ - $C\alpha$ distance is notable. Like the open β -turns the increase of the $C\alpha$ - $C\alpha$ distance cut-off is reasonable. For example within the turn-type Ia a cut-off of 6.5Å, that Dasgupta et al. (Dasgupta 2004) used during their analysis, would ignore a large amount of turn structures of this type (Figure 8.25).

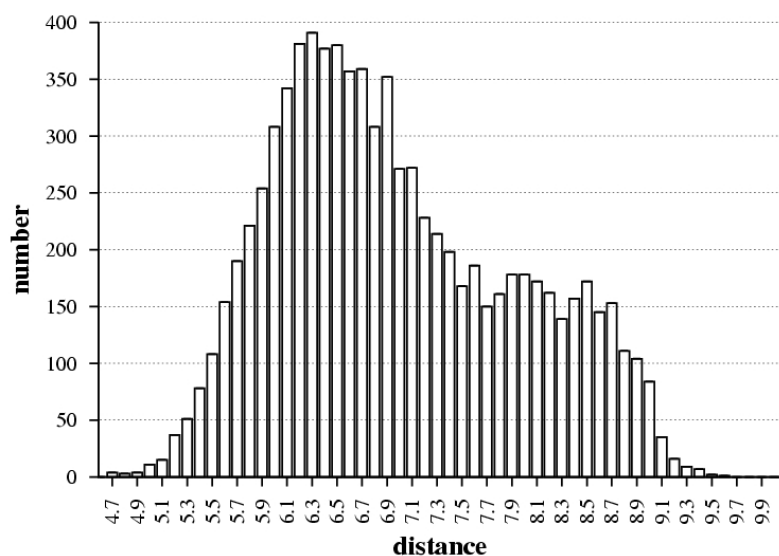


Figure 8.25: $C\alpha$ - $C\alpha$ distance distribution of open α turn-type Ia

8.5.3 Reverse α -turns

In comparison with the high frequency of normal α -turns, the occurrence of reverse α -turns with a hydrogen bond between NH_i and CO_{i+4} is fairly low (1340 structures). The resulting 21 clusters lead to 15 different turn-types (Figure 8.26); with one of these turn-types (turn-type XV) containing a *cis*-peptide bond (Table 8.21). The Ramachandran plot shows the distribution of the torsion angles within six regions (Figure 8.27), with the majority located in the region with a negative ψ angle.



Figure 8.26: Trained ESOM map for reverse α turns

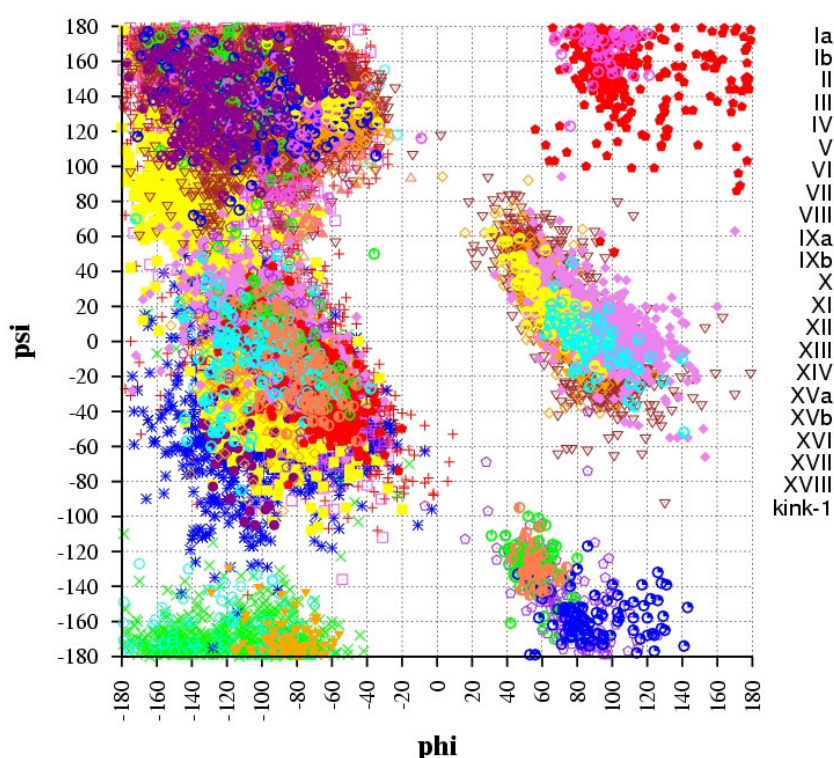


Figure 8.27: Ramachandran plot of reverse α turn-types

cluster	type	nr	ϕ_i		ψ_i		ω_i		ϕ_{i+1}		ψ_{i+1}		ω_{i+1}		ϕ_{i+2}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	0	149	-77.8	79.7	54.6	115.6	-177.1	23.4	-30.2	83.5	30.2	107.5	-169.8	40.1	-28.3	82.6
1	Ia	573	-87.4	17.3	165.9	10.1	-179.8	3.6	-59.9	9.5	-23.8	14.5	178.9	3.4	-87.4	14.5
2	Ib	234	-89.3	14.7	-172.7	7.4	-177.5	3.3	-65.9	9.9	-21.5	13.3	178.5	3.4	-92.6	12.7
3	Ic	13	-91.9	15.7	167.5	9.9	-179.9	2.8	-64.6	8.2	-25.9	18.5	179.2	2.3	-83.4	20.1
4	IIa	37	-100.6	18.3	-38.4	20.2	177.5	5.9	-151.4	20.0	-171.6	7.4	-179.4	4.8	-67.5	9.3
5	IIb	27	-100.3	17.3	-32.8	26.2	178.0	3.0	-152.1	22.7	171.5	7.4	-179.8	5.4	-58.4	8.9
6	IIIa	10	-102.2	19.3	-42.5	31.3	176.3	5.0	-154.5	28.3	-175.9	4.5	178.0	5.5	-55.3	22.6
7	IIIb	13	-85.7	18.3	-37.2	12.3	177.5	4.3	-156.4	15.2	164.4	8.6	-179.2	2.2	-52.6	8.7
8	IVa	31	-106.3	19.7	166.0	8.4	177.3	3.7	-56.8	13.8	145.6	16.5	179.5	3.8	81.1	21.5
9	IVb	21	-108.2	25.1	-153.9	18.7	-178.5	3.6	-67.9	27.4	140.7	23.0	177.6	4.3	58.8	10.7
10	V	40	-113.1	20.0	130.0	18.8	178.3	4.3	55.3	7.0	28.8	9.0	179.8	4.6	64.2	12.0
11	VIa	15	-127.3	16.6	107.4	13.4	-177.8	2.1	-48.6	8.2	132.2	10.0	179.2	2.0	76.5	14.5
12	VIb	19	-112.1	16.1	125.4	15.1	-178.4	4.5	-52.5	6.6	128.8	6.8	179.5	2.5	72.9	11.6
13	VII	49	-110.1	28.3	131.0	47.0	-179.0	4.2	-57.5	22.2	-32.4	23.5	-178.9	3.3	-96.9	20.8
14	VIII	5	-85.9	10.8	142.2	16.6	-179.6	1.7	-60.3	5.1	-30.9	4.5	-178.3	4.1	-86.5	10.0
15	IX	12	-109.3	11.2	2.9	9.3	-178.8	4.4	99.7	16.6	160.0	10.7	177.7	4.3	53.7	8.1
16	X	10	-87.6	12.0	1.4	27.8	-178.8	2.4	75.9	30.1	6.1	22.7	-179.4	2.9	-115.9	21.9
17	XI	16	-125.1	12.9	118.2	26.7	-177.8	2.9	60.8	13.5	-144.0	15.2	-179.1	3.5	-96.5	16.0
18	XII	9	-106.2	13.6	8.1	4.9	179.1	1.6	67.5	6.8	-5.3	22.9	-177.0	2.9	-118.3	19.0
19	XIII	6	106.8	19.8	-16.8	15.7	178.5	1.4	-63.4	14.5	128.9	10.6	-179.0	3.9	53.0	8.3
20	XIV	7	-94.7	13.4	-47.1	30.7	179.8	3.8	-162.1	23.6	145.6	14.1	178.9	2.0	46.7	6.3
21	XV	7	-144.2	21.9	86.4	15.8	-178.2	2.8	-68.8	6.1	146.4	9.1	0.1	1.0	-93.5	5.4

cluster	type	nr	ψ_{i+2}		ω_{i+2}		ϕ_{i+3}		ψ_{i+3}		ω_{i+3}		ϕ_{i+4}		ψ_{i+4}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	0	149	-10.0	76.4	-173.7	32.1	-47.5	96.8	-0.2	83.6	-178.4	15.4	-72.8	79.5	60.9	125.3
1	Ia	573	3.5	14.6	178.4	3.5	88.2	16.2	4.8	18.7	178.6	3.4	-86.1	19.3	145.4	14.3
2	Ib	234	4.5	11.4	179.6	3.1	80.6	13.6	12.5	14.8	179.0	3.2	-90.8	17.3	142.1	11.5
3	Ic	13	3.5	19.6	179.2	3.4	84.7	19.3	10.9	24.8	178.8	3.0	-109.7	26.6	-168.3	8.0
4	IIa	37	-21.7	12.7	-177.9	3.4	-116.3	17.0	-7.0	21.2	-175.8	5.4	-125.0	20.5	134.7	15.6
5	IIb	27	-24.5	12.1	-179.0	2.6	-112.3	15.4	4.8	17.1	-178.5	3.1	-132.4	13.9	121.4	20.6
6	IIIa	10	117.2	18.1	179.5	2.5	65.9	4.1	19.3	8.8	-179.5	3.2	-129.0	4.7	120.5	12.6
7	IIIb	13	123.7	7.4	179.6	1.7	78.6	17.2	6.3	16.2	-178.7	3.4	-129.8	11.7	114.9	17.7
8	IVa	31	14.5	21.0	178.5	5.3	71.5	19.0	30.3	25.0	179.2	3.1	-98.6	28.2	148.0	21.3
9	IVb	21	33.6	26.5	178.3	3.6	70.4	16.4	19.1	19.3	178.3	2.0	-86.0	17.4	147.2	18.9
10	V	40	3.4	26.8	-176.1	4.4	-129.3	18.3	-21.8	23.3	-177.8	5.0	-116.3	24.6	137.1	18.7
11	VIa	15	-6.2	23.4	-176.3	3.6	-125.3	14.8	-53.4	15.6	-178.1	3.1	-151.3	11.5	138.6	28.9
12	VIb	19	-1.2	23.3	-178.0	5.2	-119.7	23.9	-38.7	28.2	-179.2	4.5	122.1	36.1	-154.2	17.0
13	VII	49	-30.9	21.4	-176.0	6.0	-106.3	19.4	-10.5	20.9	-179.0	3.5	85.0	21.4	-142.2	26.9
14	VIII	5	-1.2	10.6	-179.2	2.3	117.4	15.4	149.7	8.8	-179.1	5.4	86.3	13.7	-170.9	5.7
15	IX	12	37.8	12.5	179.2	2.9	67.8	11.8	21.9	19.9	-180.0	5.8	-113.4	16.4	140.2	16.1
16	X	10	19.0	20.2	178.6	4.4	93.1	17.2	6.8	24.6	-178.0	4.2	-118.5	20.5	143.1	20.7
17	XI	16	-15.7	19.5	-175.7	3.7	-127.3	24.8	0.4	19.4	-178.8	4.7	-116.0	27.4	142.4	18.6
18	XII	9	-54.7	33.7	-179.1	2.1	-119.2	22.9	-23.5	28.2	179.6	6.6	-132.9	13.5	140.4	24.4
19	XIII	6	30.2	10.8	179.6	3.5	72.0	10.6	16.2	6.5	-178.2	3.2	-102.5	15.9	128.9	23.2
20	XIV	7	-123.2	7.2	179.6	2.0	-103.9	16.6	20.7	11.7	-178.5	2.6	-95.9	8.5	115.5	16.8
21	XV	7	6.0	7.9	-175.3	4.2	-78.3	5.8	-38.1	9.9	-177.9	4.6	-136.6	9.1	116.7	18.9

Table 8.22: Mean torsion angle and standard deviation for reverse α -turns

The overall amino acid propensities (Table 8.23) of the reverse α -turns exhibit a similar behaviour compared to the reverse β -turns. Ile, Leu, Met, Phe, Val, Cys and Trp are under represented at the position 2, 3 and 4 and in contrast Asp and Asn are highly over represented at position 1 and 3. Additionally, Gly and Pro are observed at position 4 and 2 respectively.

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	< 284	0.58	< 40	0.41	> 126	1.28	< 56	0.57	< 17	0.17	< 45	0.46
ILE	< 149	0.39	< 56	0.74	< 20	0.26	< 10	0.13	< 13	0.17	< 50	0.66
LEU	< 217	0.37	< 83	0.71	< 35	0.30	< 27	0.23	< 32	0.27	< 40	0.34
MET	< 55	0.44	16	0.63	< 8	0.32	< 6	0.24	< 11	0.44	< 14	0.55
PHE	< 135	0.49	< 31	0.56	< 22	0.40	< 26	0.47	< 24	0.44	< 32	0.58
PRO	< 207	0.71	< 0	0.00	> 162	2.76	< 38	0.65	< 4	0.07	< 3	0.05
VAL	< 189	0.41	< 62	0.68	< 25	0.27	< 15	0.16	< 13	0.14	74	0.81
ARG	282	0.87	< 47	0.72	68	1.04	< 42	0.65	< 28	0.43	> 97	1.49
ASP	> 838	2.21	> 290	3.83	> 105	1.39	> 330	4.36	< 59	0.78	< 54	0.71
GLU	396	0.91	< 33	0.38	> 127	1.46	> 109	1.25	< 29	0.33	98	1.12
LYS	407	1.04	< 48	0.61	> 104	1.33	76	0.97	< 48	0.61	> 131	1.67
ASN	> 597	2.06	> 137	2.37	> 75	1.30	> 194	3.35	> 91	1.57	> 100	1.73
CYS	< 54	0.51	24	1.13	< 4	0.19	< 8	0.38	< 10	0.47	< 8	0.38
GLN	226	0.90	< 36	0.71	47	0.93	44	0.87	< 25	0.50	> 74	1.47
HIS	161	1.01	31	0.98	33	1.04	31	0.98	26	0.82	40	1.26
SER	466	1.18	> 97	1.23	> 133	1.68	93	1.17	< 35	0.44	> 108	1.36
THR	392	1.09	> 125	1.73	56	0.78	< 49	0.68	< 26	0.36	> 136	1.89
TRP	< 48	0.49	13	0.66	< 8	0.41	< 6	0.31	< 7	0.36	14	0.72
TYR	< 155	0.67	< 32	0.69	< 25	0.54	< 30	0.64	< 26	0.56	42	0.90
GLY	> 1245	2.77	99	1.10	> 118	1.31	> 110	1.22	> 777	8.64	> 141	1.57
OTH	12	0.51	3	0.63	2	0.42	3	0.63	2	0.42	2	0.42
ALL	6515	1.00	1303	1.00	1303	1.00	1303	1.00	1303	1.00	1303	1.00

Table 8.23: Overall frequencies and conformational parameter for reverse α turns

(statistical over representation: '>', statistical under representation: '<')

In more detail, turn-type Ia has mostly Gly at position 4, a high occurrence of Asp and Asn at position 1 and 3, a high occurrence of Ser and Thr at position 1 and 5 and Pro at position 2. Turn-type Ib is also over represented by of Asp/Asn at position 1 and 3 and Gly at position 4. In contrast, there is an over representation of Glu at position 2/5 and Lys/Asn at position 5. For a detailed explanation see chapter 8.4.3, since reverse β -turns also show this over representation of Gly, Asp and Asn and an under representation of Ile, Leu, Met, Phe, Val, and the putative explanations are similar.

There is insufficient data for reliable statistical analysis amongst the other turn-types.

When the turn-types are designated after the region of their torsion angles on the Ramachandran plot (see section 6.7.3, using Dasgupta et al. naming scheme (Dasgupta 2004)), Table 8.24 clearly shows that with a torsion angle around $\pm 180^\circ$ the geometries are located in the same region of the Ramachandran plot (e.g. turn-types Ia/Ib/Ic). Although, they share similar conformations, the amino acid propensities differ. Here, the approach to cluster the turn conformation based on the region of the Ramachandran plot would fail. Additionally, comparing the conformations of the inner residues to open α -turns (Table 8.24) shows that some open α -turn types share a similar conformation to reverse α -turn types (Table 8.25).

cluster	class	designation	cluster	class	designation
1	Ia	PAAaP	12	VIb	EPaAe
2	Ib	PAAaP	13	VII	EAAAp
3	Ic	PAAaP	14	VIII	PAAep
4	IIa	ABAAB	15	IX	Aprae
5	IIb	ABAAB	16	X	AgAae
6	IIIa	ABPaB	17	XI	BpAAe
7	IIIb	ABPaB	18	XII	AaAAe
8	IVa	PPaP	19	XIII	aPrap
9	IVb	PPraP	20	XIV	ABpAp
10	V	EraAE	21	XV	DPAAe
11	VIa	BPaDB			

Table 8.24: Designation after Dasgupta et al.
(Dasgupta 2004)

open α -turns		reverse α -turns	
II	AAA	VII	EAAAp
XI	AAa	Ia/Ib/Ic	PAAaP

Table 8.25: Open α -turns within reverse α -turn types

8.6 Six residue turns

8.6.1 Normal π -turns

3994 normal π -turns (with a hydrogen bond between NH_i and CO_{i+5}) were identified within the dataset. These turns could be divided into 7 different turn-types (Figure 8.28, Table 8.26), with one turn-type containing a cis peptide bond (turn-type VII). Only five turn-types are similar to the eleven described subclasses by Rajashankar et al. (Rajashankar 1996), the other six described turn-types could not be identified within our clustering (Table 8.27, see section 1.1). Rajashankar et al. used a nomenclature based on the position of the amino acids in the Ramachandran plot; here a geometrical similarity measure is used for clustering. Consequently, a different classification is observed.

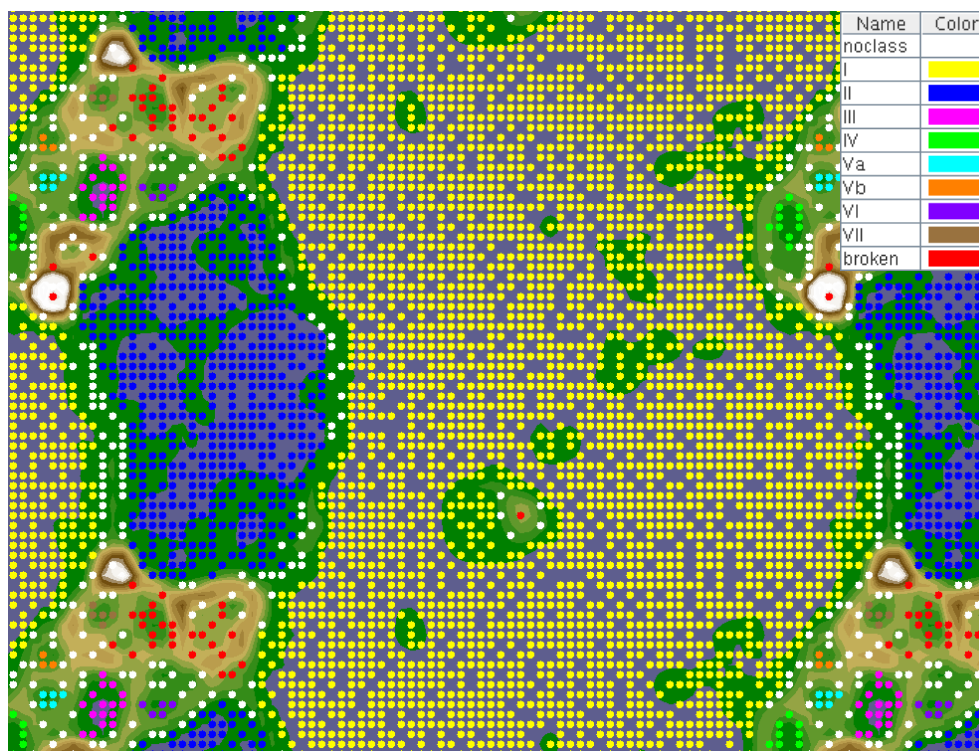


Figure 8.28: Trained ESOM map for normal π turns

cluster	type	nr	ω_i		Φ_{i+1}		Ψ_{i+1}		ω_{i+1}		Φ_{i+2}		Ψ_{i+2}		ω_{i+2}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	213	-177.6	13.6	-54.4	50.5	-12.3	74.3	-173.5	33.3	-57.5	61.7	-10.8	63.8	-169.0	40.7
1	I	2982	179.4	2.8	-64.7	7.8	-42.0	9.8	-179.9	2.7	-63.7	8.9	-31.4	12.8	179.7	3.1
2	II	689	-179.0	3.2	-67.9	10.8	-39.6	15.0	-177.7	3.7	-83.6	16.5	-37.2	20.0	-176.8	4.4
3	III	24	-180.0	2.6	51.8	6.9	55.5	15.7	179.3	1.6	56.4	11.5	42.0	14.4	179.1	2.3
4	IV	18	-179.4	4.5	-64.6	8.1	130.4	8.3	-179.4	2.8	77.0	17.6	0.2	27.4	177.8	3.9
5	Va	9	179.4	5.2	85.3	8.7	168.6	11.6	-176.7	3.6	-63.8	7.2	-24.2	3.8	-179.9	4.2
6	Vb	5	177.6	4.2	79.6	9.3	-171.5	5.0	-175.8	6.7	-63.1	7.4	-30.2	10.4	179.5	2.2
7	VI	8	-178.6	4.5	-80.2	8.9	-12.1	16.5	-175.1	6.1	-114.9	20.0	38.0	17.0	174.6	7.9
8	VII	7	-179.6	3.5	-97.5	12.6	-31.6	15.9	179.7	2.0	-136.5	8.4	104.1	14.8	-1.0	1.9

cluster	type	nr	Φ_{i+3}		Ψ_{i+3}		ω_{i+3}		Φ_{i+4}		Ψ_{i+4}		ω_{i+4}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	213	-70.8	69.5	-8.4	67.0	-171.5	34.1	-34.1	81.4	6.2	60.2	-178.3	14.5
1	I	2982	-91.6	14.7	3.1	16.9	179.2	3.6	71.8	23.1	28.7	20.0	179.4	3.5
2	II	689	-99.5	24.8	-45.9	27.9	-177.2	4.3	-88.1	29.5	-47.1	16.5	-178.0	3.9
3	III	24	74.0	14.4	11.6	12.5	-180.0	2.5	-84.9	13.8	-23.2	13.2	179.7	3.2
4	IV	18	-117.5	21.5	-83.0	27.3	178.4	3.4	-88.6	21.8	-22.9	19.6	179.2	4.7
5	Va	9	-98.8	12.1	-0.4	13.8	179.5	6.6	57.4	11.9	55.5	22.5	-177.2	6.2
6	Vb	5	-101.2	5.5	10.0	10.6	178.3	3.0	52.6	9.6	48.3	20.7	-179.6	1.3
7	VI	8	140.9	21.5	-32.1	17.2	-177.0	4.3	-76.5	28.6	-31.8	19.8	177.7	4.7
8	VII	7	-76.4	7.9	163.9	15.6	-177.6	4.5	-69.3	11.2	113.1	12.5	-179.2	2.2

Table 8.26: Mean torsion angles and standard deviation for normal π -turns

type	designation	rajashankar et al(Rajashankar 1996)	
		class	subclass
I	$\alpha\alpha\alpha\gamma\alpha$	α	$\alpha\alpha\alpha\gamma\alpha$
II	$\alpha\alpha\alpha\alpha\alpha$	α	$\alpha\alpha\alpha\alpha\alpha$
III	$\alpha\alpha\alpha\alpha\alpha$	α'	$\alpha\alpha\alpha\alpha\alpha$
IV	$\beta\alpha\alpha\alpha\alpha$	α	$\beta\alpha\alpha\alpha\alpha$
Va	$\varepsilon\alpha\alpha\gamma\alpha$		
Vb	$\varepsilon'\alpha\alpha\gamma\alpha$		
VI	$\alpha\alpha\gamma\alpha\alpha$	α	$\alpha\alpha\gamma\alpha\alpha$
VII	$\alpha\alpha\beta\beta\beta$		

Table 8.27: Designation of normal turn types based on position in Ramachandran plot and comparing to subclasses described by Rajashankar et al.(Rajashankar 1996)

The Ramachandran plot (Figure 8.29) shows presumably that steric constraints within normal π -turns are low. This leads to a broad variation of torsion angles within the different turn-types. Especially, turn-type I has several outlier in the upper left quadrant of the plot.

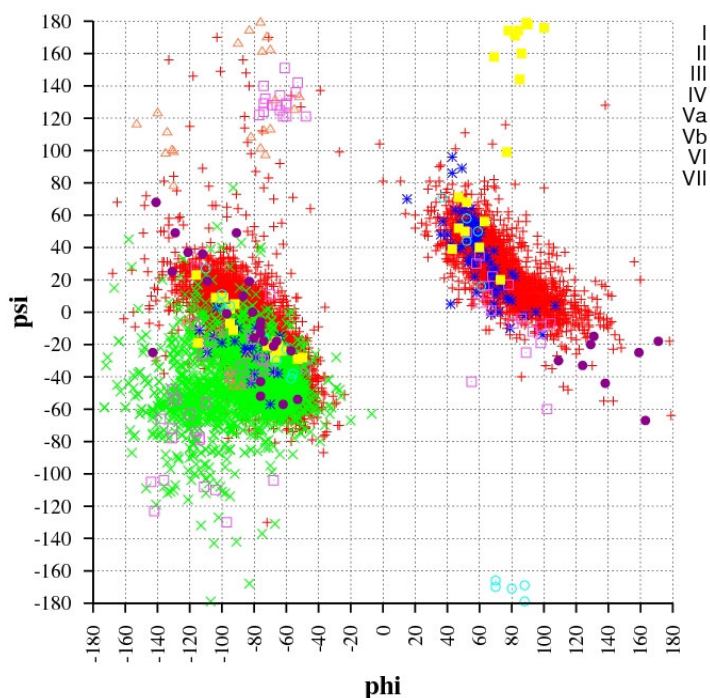


Figure 8.29: Ramachandran plot for normal π turns

The overall amino acid propensities for normal π -turns (Table 8.28) reveal that Gly is highly over represented at position 5. In contrast, Ala, Ile, Leu, Met, Phe, Pro, Val Thr, Trp, Tyr show only a low occurrence. Additionally, Asn is also over represented at this position. Interestingly, Pro is generally under represented in these turns; it has a low occurrence at all six positions.

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	P	f	P	f	P	f	P
ALA	> 2284	1.27	> 639	2.14	> 370	1.24	> 450	1.51	> 398	1.33	< 110	0.37	317	1.06
ILE	< 1174	0.85	< 202	0.87	> 269	1.17	< 114	0.49	< 128	0.55	< 67	0.29	> 394	1.71
LEU	> 2688	1.26	> 801	2.25	358	1.00	< 260	0.73	> 537	1.51	< 133	0.37	> 599	1.68
MET	517	1.12	> 153	1.99	77	1.00	70	0.91	> 102	1.33	< 24	0.31	91	1.19
PHE	861	0.86	> 212	1.26	< 122	0.73	< 80	0.48	< 140	0.84	< 56	0.33	> 251	1.50
PRO	< 220	0.21	< 52	0.29	< 89	0.50	< 37	0.21	< 25	0.14	< 17	0.10	< 0	0.00
VAL	< 1276	0.77	< 194	0.70	255	0.92	< 162	0.58	< 160	0.58	< 106	0.38	> 399	1.44
ARG	1336	1.13	< 138	0.70	> 322	1.63	> 316	1.60	> 250	1.26	< 145	0.73	< 165	0.83
ASP	< 1163	0.84	< 200	0.87	< 186	0.81	> 266	1.16	< 160	0.70	222	0.97	< 129	0.56
GLU	1580	1.00	< 110	0.42	> 362	1.37	> 537	2.03	267	1.01	< 145	0.55	< 159	0.60
LYS	> 1630	1.14	< 92	0.39	> 397	1.67	> 420	1.76	261	1.10	237	1.00	223	0.94
ASN	1203	1.14	< 124	0.71	< 127	0.72	200	1.14	> 311	1.77	> 334	1.90	< 107	0.61
CYS	363	0.94	> 128	1.98	< 41	0.63	< 29	0.45	52	0.80	< 39	0.60	74	1.15
GLN	976	1.06	< 106	0.69	> 201	1.31	> 241	1.57	> 194	1.27	< 128	0.84	< 106	0.69
HIS	605	1.05	80	0.83	> 117	1.21	81	0.84	> 165	1.71	89	0.92	< 73	0.76
SER	< 1089	0.76	< 203	0.84	< 188	0.78	259	1.08	219	0.91	< 96	0.40	< 124	0.52
THR	< 817	0.62	< 101	0.46	< 157	0.72	< 153	0.70	198	0.90	< 54	0.25	< 154	0.70
TRP	307	0.86	> 81	1.36	63	1.06	< 34	0.57	< 42	0.71	< 22	0.37	65	1.09
TYR	833	0.98	164	1.16	129	0.91	< 95	0.67	> 189	1.34	< 64	0.45	> 192	1.36
GLY	> 2708	1.65	< 148	0.54	< 109	0.40	< 141	0.52	< 138	0.51	> 1857	6.80	> 315	1.15
OTH	100	1.16	> 27	1.88	16	1.12	10	0.70	19	1.32	10	0.70	18	1.25
ALL	23730	1.00	3955	1.00	3955	1.00	3955	1.00	3955	1.00	3955	1.00	3955	1.00

Table 8.28: Overall frequencies and conformational parameter for normal π turns

(statistical over representation: '>', statistical under representation: '<')

In more detail, turn-type I shows a pronounced occurrence of Gly at Pos 5 and, in contrast, has a low occurrence of Gly at position 2, 3 and 4. The really high occurrence of Gly at position 5 could be explained by the conformation of the ϕ and ψ torsion angles of amino acids at this position. The region in the Ramachandran plot is only conformationally accessible by Gly, Asp and Asn. The latter two residues show normal and higher occurrences, respectively. In contrast Val, Ile, Thr do not show this conformation due to their C_{β} -branching, which would produce steric clashes. At position 1 there is an over representation of Ala, Leu, Met and Cys which are under represented at position 5. In contrast, in turn-type II Phe, Trp and Tyr are over represented at position 1, whereas Pro almost never occurs at this position.

8.6.2 Open π -turns

Figure 8.30 shows the 28 clusters obtained within the 21204 open π -turns. After visual inspection 6 classes were identified as non turn-like structures. These structures appear more like a hook. The designation after Efimov (Efimov 1993) shows that the first two residues of the hook-like structures are located within the β region of the Ramachandran plot (Figure 8.31), which indicates an extended protein chain and the $C\alpha$ - $C\alpha$ distances (Table 8.30) are at the upper limit of the chosen 10Å cut-off. Additionally, many structures do not belong to any cluster. This is not surprising, since the conformational space of parts of the protein chain with six residues in length is large. Furthermore, only turn-type I corresponds to a previously described geometry.

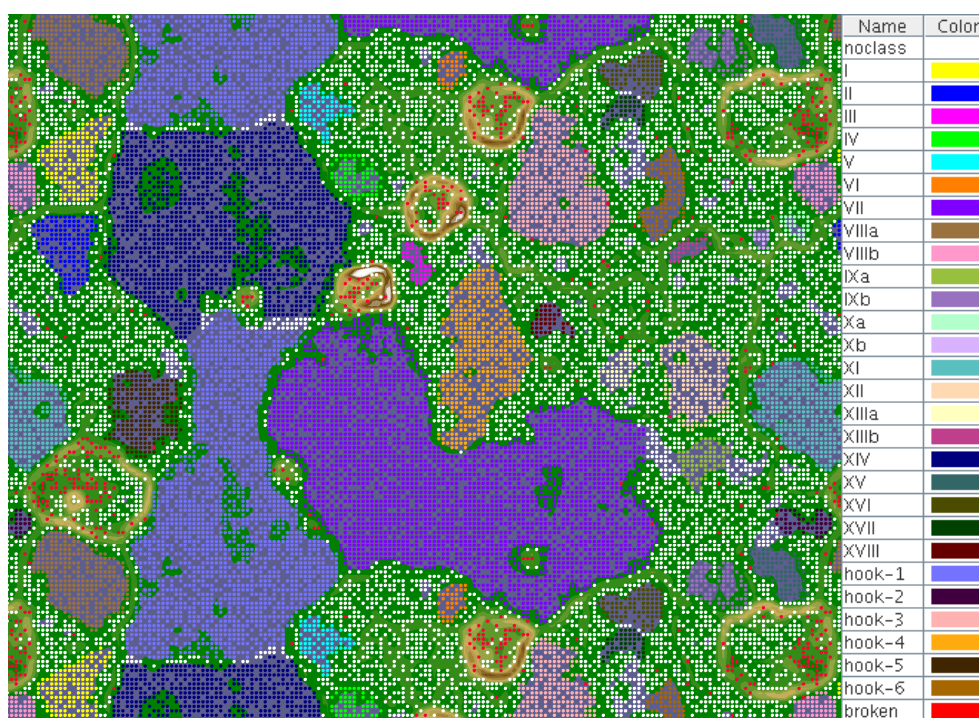
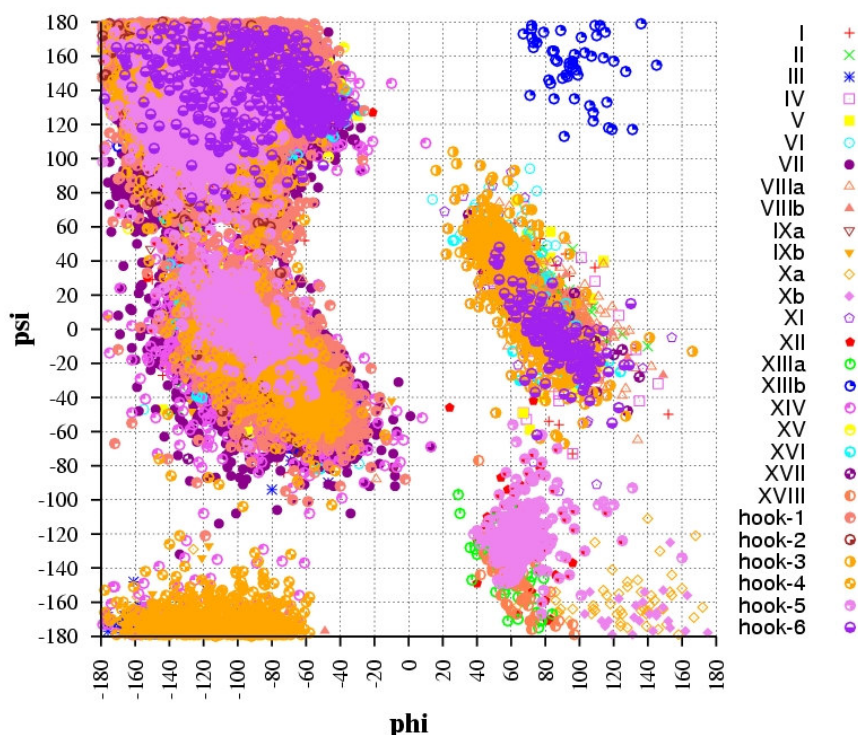


Figure 8.30: Trained ESOM map for open π turns

Figure 8.31: : Ramachandran plot of open π turns

cluster	class	nr	ω_i		ϕ_{i+1}		ψ_{i+1}		ω_{i+1}		ϕ_{i+2}		ψ_{i+2}		ω_{i+2}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset
0	0	6737	-179.2	15.1	-78.5	73.4	57.5	101.5	-179.8	12.6	-60.9	80.3	29.0	121.5	-175.6	28.5
1	I	253	-179.2	3.2	-65.5	12.9	-38.7	13.5	-178.2	3.0	-76.7	15.1	-32.6	16.4	-177.7	3.7
2	II	236	-179.4	2.6	-63.0	6.9	-30.5	10.3	179.7	2.4	-83.8	11.7	-3.5	12.1	179.5	3.4
3	III	59	-179.8	2.0	-77.7	17.6	-37.8	17.3	179.8	2.3	-156.2	13.1	-172.4	6.4	-179.1	2.3
4	IV	121	179.0	3.3	82.2	18.6	3.2	23.1	178.3	4.7	-87.6	26.5	148.0	18.8	-179.8	3.7
5	V	116	179.6	2.9	71.3	16.3	20.2	25.1	179.6	2.9	-114.4	29.3	138.9	26.0	-179.3	3.6
6	VI	64	179.0	3.4	53.0	12.3	51.4	15.2	-179.0	1.6	-55.4	8.8	-41.6	13.8	-179.8	1.9
7	VII	4107	178.4	4.4	-87.3	27.0	123.0	25.6	-178.0	3.9	-58.0	11.1	-37.7	13.9	-179.6	3.0
8	VIIIa	454	179.1	4.2	-88.2	16.9	165.7	9.8	-179.5	2.9	-61.1	10.5	-22.4	15.2	179.0	3.2
9	VIIIb	152	178.8	2.9	-90.3	14.8	-173.4	5.5	-177.5	3.1	-65.9	8.7	-20.7	11.7	178.6	3.1
10	IXa	104	178.6	2.2	-138.6	21.1	163.5	16.0	-179.8	2.1	-62.4	16.1	-27.6	16.3	-177.7	3.0
11	IXb	143	175.9	5.6	-128.9	21.9	-169.7	8.6	-178.6	3.6	-67.8	13.7	-27.3	16.6	-177.5	3.9
12	Xa	52	177.6	4.1	-131.6	19.5	-169.5	9.2	-178.3	4.5	-63.3	10.5	-24.7	16.6	-179.2	3.0
13	Xb	25	179.7	2.1	-118.0	22.6	169.6	8.2	179.8	2.1	-59.7	8.9	-28.9	10.4	-178.9	2.0
14	XI	675	178.5	3.9	-134.3	17.5	122.4	14.1	-179.5	3.2	53.0	7.5	41.8	13.6	178.2	2.7
15	XII	340	177.2	4.4	-129.5	19.4	105.8	21.2	-177.3	3.8	65.2	11.0	-122.2	15.5	179.9	2.6
16	XIIIa	82	179.6	3.7	-121.5	24.8	154.9	16.7	176.9	3.5	57.4	11.0	-139.0	15.5	-178.3	3.2
17	XIIIb	49	179.8	3.3	-109.7	24.6	136.2	28.4	-179.7	3.0	98.1	19.0	154.3	18.2	-179.8	3.8
18	XIV	2221	-179.4	4.5	-93.2	21.6	-16.2	32.3	-179.7	4.0	-99.5	36.1	143.8	20.5	-179.6	4.0
19	XV	134	-180.0	2.6	-94.4	15.8	-6.8	21.7	179.8	3.0	-72.6	14.7	153.1	13.8	178.9	3.6
20	XVI	136	-179.4	3.1	-92.0	20.4	-3.9	25.0	179.2	3.0	-71.9	14.2	140.5	16.0	179.4	3.8
21	XVII	69	-179.6	2.1	-83.8	15.8	-16.2	19.2	-179.8	2.5	-117.5	24.1	143.1	16.8	179.6	2.7
22	XVIII	69	-178.3	3.2	-88.6	16.2	-8.9	18.7	178.8	3.3	61.4	11.6	-143.8	17.7	-178.2	3.2
23	hook-1	2446	179.7	4.0	-101.4	28.5	129.8	28.6	177.8	4.8	-105.6	29.7	163.2	13.8	179.9	3.8
24	hook-2	71	179.6	2.7	-110.8	23.3	116.0	28.4	-179.3	3.0	-151.7	15.4	158.5	11.4	178.6	3.3

25	hook-3	672	179.1	3.7	-114.9	21.0	132.0	16.9	178.6	3.4	-134.6	17.7	123.4	16.5	-179.7	3.1
26	hook-4	639	179.6	2.9	-107.6	29.4	135.1	24.4	178.6	2.8	-110.0	27.0	-170.7	11.5	-178.5	3.2
27	hook-5	365	178.8	4.4	-111.4	26.7	134.3	16.9	177.7	3.9	-131.0	17.7	105.9	26.4	-177.3	4.2
28	hook-6	181	-179.3	2.8	-95.0	28.0	134.2	19.4	178.3	3.2	-105.1	28.0	160.8	11.4	178.7	3.5

cluster	class	nr	Φ_{i+3}		Ψ_{i+3}		Ω_{i+3}		Φ_{i+4}		Ψ_{i+4}		Ω_{i+4}	
			$\bar{\Phi}$	\pm	$\bar{\Psi}$	\pm	$\bar{\Omega}$	\pm	$\bar{\Phi}$	\pm	$\bar{\Psi}$	\pm	$\bar{\Omega}$	\pm
0	0	6737	-40.5	74.1	3.4	82.8	-174.3	31.9	-51.1	85.4	29.2	84.1	-179.8	8.7
1	I	253	-103.5	17.7	-7.8	17.8	-178.9	3.4	77.4	16.3	15.2	21.2	179.9	2.4
2	II	236	80.4	12.9	14.7	12.6	179.9	3.2	-93.6	21.3	-11.8	31.6	179.0	4.4
3	III	59	-68.8	13.0	-24.4	17.9	-180.0	2.5	-90.0	28.7	-23.1	26.1	-179.9	1.6
4	IV	121	-58.2	8.6	-38.7	11.3	-179.9	2.5	-67.1	12.9	-31.8	18.1	178.3	3.2
5	V	116	-63.4	9.9	-25.8	18.7	-180.0	2.7	-82.6	22.1	-14.8	24.2	179.5	2.1
6	VI	64	-62.9	9.6	-38.7	14.3	179.5	1.8	-77.2	23.9	-37.1	15.6	-179.8	1.4
7	VII	4107	-67.4	13.3	-37.5	15.8	179.9	3.3	-83.4	24.7	-26.1	28.2	-179.7	3.6
8	VIIIa	454	-87.6	14.0	4.2	13.8	178.3	3.1	88.5	15.6	5.2	19.1	179.0	2.9
9	VIIIb	152	-91.8	12.6	3.0	11.6	179.4	2.9	82.4	13.5	12.4	12.4	179.7	2.5
10	IXa	104	-108.4	18.3	-9.4	25.0	-179.4	3.6	-148.7	20.0	153.6	17.9	178.1	2.3
11	IXb	143	-115.4	14.5	-10.9	20.0	-177.4	3.8	-143.4	18.5	153.2	18.8	176.5	4.2
12	Xa	52	-98.1	17.3	2.9	19.7	179.4	3.4	127.4	25.1	-158.7	14.3	-178.4	3.6
13	Xb	25	-92.8	9.7	-3.4	12.3	-179.7	3.3	137.7	18.3	-166.2	11.8	-179.6	1.3
14	XI	675	78.5	14.7	2.1	17.7	179.9	3.6	-113.0	18.8	139.4	15.1	178.6	3.8
15	XII	340	-97.3	15.8	3.8	17.4	-179.1	3.6	-104.4	21.6	139.8	17.8	177.9	3.6
16	XIIIa	82	-65.0	13.1	-27.5	16.6	179.5	4.0	-64.5	10.5	-31.2	13.1	178.4	2.4
17	XIIIb	49	-59.3	10.4	-38.2	10.5	-179.7	2.3	-60.9	9.9	-40.3	15.5	-179.8	1.9
18	XIV	2221	-59.8	10.3	-32.9	14.9	-180.0	3.0	-72.1	18.3	-25.5	22.6	179.4	3.5
19	XV	134	-66.2	12.3	-31.1	16.8	178.6	3.5	-122.3	22.0	127.2	28.1	-179.4	2.8
20	XVI	136	-59.6	10.5	135.4	10.5	179.2	3.1	78.9	15.5	4.5	22.0	179.4	2.5
21	XVII	69	-58.9	10.1	134.6	13.5	179.4	2.0	76.8	21.9	11.2	23.2	179.9	1.5
22	XVIII	69	-66.3	13.2	-20.3	16.6	178.7	3.8	-73.8	18.0	-18.9	14.9	179.6	2.1
23	hook-1	2446	-57.5	10.0	-33.8	13.4	179.9	2.9	-76.6	19.0	-21.4	24.3	179.5	2.9
24	hook-2	71	-78.1	14.5	-26.0	18.3	179.0	2.6	-128.1	21.1	120.6	31.6	-179.8	2.1
25	hook-3	672	52.9	7.7	43.0	13.3	178.3	2.4	77.1	14.1	3.2	18.1	179.7	2.4
26	hook-4	639	-63.2	9.9	-30.3	15.6	-179.9	2.7	-86.7	22.3	-14.0	25.3	-180.0	2.0
27	hook-5	365	66.2	13.5	-121.8	14.6	179.4	3.8	-96.0	16.8	4.6	18.0	-179.3	2.9
28	hook-6	181	-55.3	8.4	131.5	9.9	179.8	2.3	88.0	16.9	-5.3	18.7	179.7	2.2

Table 8.29: Mean torsion angles and standard deviation for open π -turns

cluster	type	designation	min	max	median	mean	stdev
1	I	$\alpha_R \alpha_R \gamma \alpha_L$	3.8	6.4	4.9	4.9	0.4
2	II	$\alpha_R \gamma \alpha_L \gamma$	3.6	5.9	4.9	4.8	0.4
3	III	$\alpha_R \beta \alpha_R \alpha_R$	4.8	7.7	6.1	6.0	0.6
4	IV	$\alpha_L \beta \alpha_R \alpha_R$	4.6	9.3	6.9	7.1	1.1
5	V	$\alpha_L \beta \alpha_R \gamma$	4.5	9.1	7.1	7.0	1.3
6	VI	$\alpha_L \alpha_R \alpha_R \alpha_R$	3.8	8.8	8.2	8.1	0.7
7	VII	$\beta \alpha_R \alpha_R \alpha_R$	4.0	9.4	6.9	6.9	1.0
8	VIIIa	$\beta \alpha_R \gamma \alpha_L$	4.7	9.1	5.6	5.7	0.5
9	VIIIb	$\beta \alpha_R \gamma \alpha_L$	4.4	6.8	5.2	5.2	0.3
10	IXa	$\beta \alpha_R \gamma \beta$	3.9	9.4	4.8	5.5	1.5
11	IXb	$\beta \alpha_R \gamma \beta$	4.0	8.3	4.5	4.7	0.6
12	Xa	$\beta \alpha_R \gamma e$	3.8	5.2	4.3	4.3	0.3
13	Xb	$\beta \alpha_R \gamma e$	4.0	5.1	4.5	4.6	0.2
14	XI	$\beta \alpha_L \alpha_L \beta$	3.6	6.4	4.5	4.5	0.3
15	XII	$\beta e \gamma b$	3.4	8.8	4.5	4.6	0.6
16	XIIIa	$\beta e \alpha_R \alpha_R$	3.4	7.6	5.0	5.0	0.7
17	XIIIb	$\beta e \alpha_R \alpha_R$	3.9	9.4	6.0	6.1	1.1
18	XIV	$\gamma \beta \alpha_R \alpha_R$	3.6	10.0	6.3	6.4	1.0
19	XV	$\gamma \beta \alpha_R b$	4.5	9.1	7.5	7.3	1.1
20	XVI	$\gamma \beta \beta \alpha_L$	3.8	9.7	6.4	6.4	1.1
21	XVII	$\gamma \beta \beta \alpha_L$	4.2	7.8	6.1	6.2	0.9
22	XVIII	$\gamma e \alpha_R \alpha_R$	6.2	9.7	8.3	8.3	0.7
23	hook-1	$\beta \beta \alpha_R \alpha_R$	4.0	10.0	9.3	9.0	0.9
24	hook-2	$\beta \beta \alpha_R \beta$	6.0	10.0	9.1	8.8	0.9
25	hook-3	$\beta \beta \alpha_R \gamma$	6.9	10.0	8.3	8.3	0.3
26	hook-4	$\beta \beta \alpha_L \alpha_L$	4.7	10.0	8.8	8.7	0.8
27	hook-5	$\beta \beta e \gamma$	6.7	9.9	8.5	8.5	0.4
28	hook-6	$\beta \beta \beta \alpha_L$	7.0	10.0	9.5	9.3	0.5

Table 8.30: Designation after Efimov and Ca-Ca distance of open π -turn types

The overall amino acid propensities (Table 8.31) reveal an under representation of Ile, Leu, Met, Phe, Val at position 3, 4 and 5. Additionally, Cys at position 4 and Pro at position 5 are also under represented. An over representation is observed for Pro at position 4, Gly at position 4, 5 and Asp at position 3. The under representation of the hydrophobic amino acid could indicate that open π -turns often occur at the protein surface, where these residues are unfavourable.

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	P	f	P	f	P	f	P
ALA	< 7862	0.84	< 1403	0.89	< 1044	0.67	< 1152	0.73	< 1360	0.87	< 1289	0.82	1614	1.03
ILE	< 4886	0.67	> 1413	1.17	< 919	0.76	< 541	0.45	< 493	0.41	< 537	0.44	< 983	0.81
LEU	< 7690	0.68	1823	0.97	< 1484	0.79	< 953	0.51	< 842	0.45	< 1068	0.57	< 1520	0.81
MET	< 1930	0.80	432	1.07	371	0.92	< 219	0.54	< 197	0.49	< 270	0.67	441	1.09
PHE	< 4413	0.84	> 1100	1.25	825	0.94	< 485	0.55	< 465	0.53	< 632	0.72	906	1.03
PRO	> 6546	1.17	> 1010	1.08	986	1.05	> 1445	1.54	> 1994	2.13	< 460	0.49	< 651	0.70
VAL	< 6478	0.74	> 1802	1.24	< 1199	0.82	< 747	0.51	< 637	0.44	< 805	0.55	< 1288	0.88
ARG	< 5414	0.87	< 914	0.88	< 931	0.90	< 825	0.79	< 893	0.86	< 890	0.86	< 961	0.93
ASP	> 10400	1.44	< 867	0.72	> 2315	1.92	> 2071	1.72	> 2005	1.66	> 1862	1.54	> 1280	1.06
GLU	8676	1.04	< 1038	0.75	< 1028	0.74	< 1230	0.89	> 2094	1.51	> 1933	1.39	1353	0.97
LYS	7253	0.97	< 1146	0.92	< 1118	0.89	< 1108	0.89	> 1363	1.09	> 1386	1.11	< 1132	0.91
ASN	> 7406	1.34	< 777	0.84	> 1453	1.57	> 1465	1.59	> 1299	1.41	> 1451	1.57	961	1.04
CYS	1927	0.95	> 439	1.29	> 442	1.30	< 264	0.78	< 159	0.47	< 261	0.77	362	1.07
GLN	< 4482	0.93	< 680	0.85	< 653	0.81	< 645	0.80	< 730	0.91	814	1.01	> 960	1.19
HIS	3139	1.03	547	1.08	> 572	1.13	511	1.01	< 442	0.87	548	1.08	519	1.03
SER	> 9088	1.20	1270	1.01	> 1633	1.29	> 2178	1.73	> 1432	1.13	1317	1.04	1258	1.00
THR	> 7578	1.10	> 1238	1.08	> 1327	1.15	> 1867	1.62	< 892	0.78	1165	1.01	1089	0.95
TRP	< 1597	0.85	> 361	1.16	< 249	0.80	< 208	0.67	< 221	0.71	< 236	0.76	322	1.03
TYR	< 4049	0.91	> 995	1.34	< 678	0.91	< 524	0.71	< 506	0.68	< 595	0.80	751	1.01
GLY	> 13482	1.57	1442	1.00	1471	1.03	> 2296	1.60	> 2710	1.89	> 3208	2.24	> 2355	1.64
OTH	< 333	0.74	75	1.00	74	0.98	< 38	0.50	< 38	0.50	< 44	0.58	64	0.85
ALL	124629	1.00	20772	1.00	20772	1.00	20772	1.00	20772	1.00	20771	1.00	20770	1.00

Table 8.31: Overall frequencies and conformational parameter for open π turns

(statistical over representation: '>', statistical under representation: '<')

As turn-type I of normal and open π -turns show a similar torsion angle distribution, it is not surprising that open turn-type I also has a high occurrence of Gly at position 5 and a generally low occurrence of Pro. The occurrence of Gly at position 5 is so high, that only Asn shows a normal occurrence and the other amino acids more or less do not appear. In consequence, the conclusion can be drawn that, without the additionally constraints of a hydrogen bond, only Gly could show the conformation required for this turn-type. Similar properties can be found in turn-type II at position 4.

8.6.3 Reverse π -turns

Turns with a reverse hydrogen bond between NH_i and CO_{i+5} were detected only 954 times and these grouped into 13 clusters, whereas turn-type IIa/IIb/IIc share similar conformations that traversed the torsion angle boundary of $\pm 180^\circ$ (Figure 8.33, Table 8.32). Classifying these turn-types according to Efimov (Efimov 1993) based on the torsion angles in the region of the Ramachandran plot, leads to the finding, that the first and the last residue always appears as an extended part of the polypeptide chain, presumably as an β -hairpin turn to connect β -strands. This leads to the distinct β region within the Ramachandran plot (Figure 8.33).

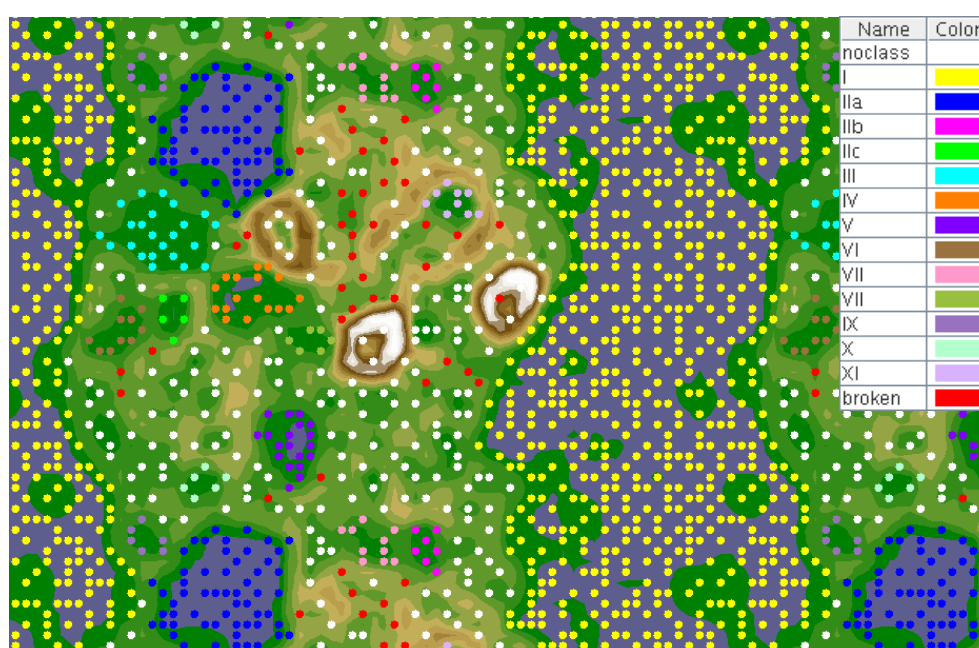


Figure 8.32: Trained ESOM map for reverse π turns

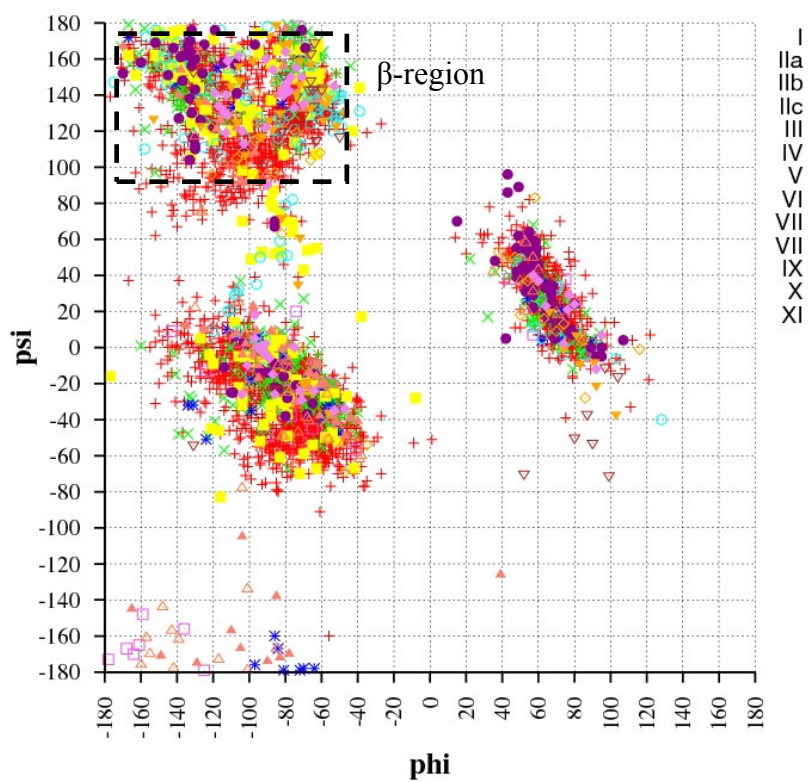


Figure 8.33: Ramachandran plot of reverse π turn-types

cluster	type	nr	φ_i		Ψ_i		ω_i		Φ_{i+1}		Ψ_{i+1}		Ω_{i+1}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	208	-69.0	77.1	54.4	107.1	-178.9	18.1	-49.5	81.1	23.6	106.5	-175.2	31.8
1	I	518	-93.8	24.0	113.0	17.1	-177.9	3.9	-63.5	11.3	-30.9	16.4	-179.3	3.3
2	IIa	65	-73.7	16.1	161.4	11.3	-178.6	3.0	-57.4	6.5	-32.3	10.5	179.6	2.6
3	IIb	7	-79.6	11.3	-174.5	7.4	-177.3	4.2	-57.4	7.2	-33.8	10.7	179.7	2.0
4	IIc	7	-71.8	7.5	160.2	14.2	-178.2	2.1	-58.1	9.0	-33.4	17.2	179.3	1.6
5	III	23	-84.0	27.9	137.9	15.8	-178.2	4.1	-76.3	16.7	-20.9	27.0	-177.7	3.2
6	IV	18	-79.9	17.8	135.3	16.0	-179.2	2.6	-55.3	7.6	132.3	8.4	179.2	3.1
7	V	19	-124.7	24.1	163.8	8.9	178.4	2.8	53.9	5.6	57.0	17.1	179.1	2.1
8	VI	11	-100.2	25.3	105.1	22.3	-177.6	4.6	-63.0	15.2	-34.6	15.6	179.1	3.9
9	VII	10	-110.2	29.3	-157.7	22.3	-178.7	6.9	-79.0	22.3	-15.9	18.6	178.6	4.4
10	VII	7	-82.4	14.9	120.2	21.4	-178.9	3.5	-61.8	5.2	140.4	16.6	179.6	4.5
11	IX	7	-87.3	11.7	166.6	8.4	178.1	8.2	-68.4	5.7	-4.6	37.4	175.5	6.7
12	X	6	-100.4	15.7	122.4	9.2	179.9	4.7	-61.2	17.6	-22.2	49.1	175.6	3.8
13	XI	9	-81.9	16.1	139.7	21.9	174.6	4.9	-108.7	12.1	134.3	13.8	4.3	7.9

cluster	type	nr	Φ_{i+2}		Ψ_{i+2}		ω_{i+2}		Φ_{i+3}		Ψ_{i+3}		Ω_{i+3}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	208	-53.1	76.3	19.9	89.9	-167.1	46.7	-54.5	79.5	11.2	97.7	-173.5	32.3
1	I	518	-74.3	16.7	-42.0	17.1	-178.6	3.4	-100.9	19.5	-9.1	14.3	-178.8	3.7
2	IIa	65	-82.6	10.3	-3.1	12.6	179.6	2.9	66.0	13.8	21.3	17.3	-178.7	3.9
3	IIb	7	-82.5	17.4	-8.7	18.0	-178.5	2.9	68.9	11.3	17.1	12.8	-177.7	4.5
4	IIc	7	-81.1	14.3	-3.3	16.5	179.8	1.0	65.8	14.9	24.5	19.4	-178.7	1.0
5	III	23	-97.2	28.6	-34.2	27.5	177.6	4.9	-119.6	27.9	152.4	25.2	-179.3	6.2
6	IV	18	73.9	20.1	13.3	22.8	179.6	3.3	-137.7	19.9	157.5	15.3	176.1	4.8
7	V	19	56.0	12.7	39.3	14.6	179.8	2.8	75.5	16.4	13.2	13.2	179.8	4.3
8	VI	11	-65.1	19.2	-49.5	17.1	-179.7	1.5	-95.8	17.4	0.8	15.0	-178.9	2.1
9	VII	10	-111.5	30.2	143.6	21.8	-179.8	3.9	-46.6	31.7	-37.7	33.8	-178.8	4.3
10	VII	7	82.6	16.5	-38.4	34.6	177.5	9.2	-94.0	19.5	-27.2	13.2	-175.7	6.1
11	IX	7	-100.1	29.9	129.4	13.2	-177.2	2.4	-59.1	8.4	133.3	7.7	179.9	2.4
12	X	6	-84.6	22.3	111.2	16.8	-175.6	8.3	82.8	18.7	2.4	16.2	-178.3	2.6
13	XI	9	-94.2	8.4	-2.0	12.2	-174.1	5.1	-74.2	16.8	-21.9	12.1	-179.6	4.2

cluster	type	nr	Φ_{i+4}		Ψ_{i+4}		ω_{i+4}		Φ_{i+5}		Ψ_{i+5}	
			\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
0	noclass	208	-48.0	78.8	32.4	77.4	-177.6	22.3	-70.5	80.2	90.0	99.9
1	I	518	65.9	14.3	28.1	21.5	179.5	3.9	-106.5	31.2	146.4	15.2
2	IIa	65	-99.9	19.8	-22.1	16.0	-179.1	3.7	-130.6	23.3	144.2	16.8
3	IIb	7	-114.6	15.2	-27.7	17.8	-178.7	2.7	-110.3	35.0	147.7	16.9
4	IIc	7	-98.5	26.4	-22.3	19.1	-178.8	1.8	-156.1	18.5	-166.0	10.5
5	III	23	-77.5	20.5	37.4	39.5	-178.6	5.1	-104.8	35.6	146.5	16.2
6	IV	18	-97.1	13.0	30.9	27.1	-179.0	2.8	-110.0	27.1	136.7	21.2
7	V	19	-85.1	15.2	-20.5	10.3	179.7	4.2	-126.3	19.4	127.3	27.9
8	VI	11	56.7	13.0	35.8	16.2	-179.4	1.7	-132.1	26.0	-164.2	14.5
9	VII	10	-90.3	14.7	-0.9	11.1	-178.8	1.7	-106.7	23.3	90.8	56.9
10	VII	7	71.5	15.9	20.8	21.4	-179.8	1.5	-94.6	29.3	146.7	14.6
11	IX	7	84.8	12.5	-7.3	19.0	178.7	1.9	-101.6	18.1	125.4	27.1

12	X	6	48.6	7.2	48.6	21.5	-176.1	5.2	-92.6	11.1	123.9	19.2
13	XI	9	75.5	13.2	19.2	17.1	178.1	3.4	-87.0	14.9	144.6	13.5

Table 8.32: Mean torsion angles and standard deviation for reverse π -turns

cluster	class	nr	designation	cluster	class	nr	designation
1	I	518	$\beta \alpha_R \alpha_R \gamma \alpha_L \beta$	8	VI	11	$\beta \alpha_R \alpha_R \gamma \beta \beta$
2	IIa	65	$\beta \alpha_R \gamma \alpha_L \alpha_R \beta$	9	VII	10	$\beta \alpha_R \beta \alpha_R \gamma \beta$
3	IIb	7	$\beta \alpha_R \gamma \alpha_L \alpha_R \beta$	10	VII	7	$\beta \beta \alpha_L \alpha_R \alpha_L \beta$
4	IIc	7	$\beta \alpha_R \gamma \alpha_L \alpha_R \beta$	11	IX	7	$\beta \gamma \beta \beta \gamma \beta$
5	III	23	$\beta \alpha_R \alpha_R \beta \gamma \beta$	12	X	6	$\beta \alpha_R \beta \alpha_L \alpha_L \beta$
6	IV	18	$\beta \beta \alpha_L \beta \gamma \beta$	13	XI	9	$\beta \beta \gamma \alpha_R \alpha_R \beta$
7	V	19	$\beta \alpha_L \alpha_L \alpha_L \alpha_R \beta$				

Table 8.33: Designation after Efimov(Efimov 1993) for reverse π -turns

Within the entire reverse π -turns a very high occurrence of Asp, Asn and Cys at position 1, Pro at position 2, Thr and Cys at position 4 and Gly at position 5 is observed. Additionally an over representation of His and Ser could be identified at position 2. In contrast, Ile, Leu, Phe, Val and Trp are generally under represented at all positions of the reverse π -turns and Cys at position 2, 3, 5 and 6. This is comparable to the finding described for the reverse α and β -turns.

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	P	f	P	f	P	f	P
ALA	< 299	0.72	< 37	0.54	75	1.09	65	0.94	< 37	0.54	< 34	0.49	< 51	0.74
ILE	< 141	0.44	< 28	0.52	< 29	0.54	< 21	0.39	< 9	0.17	< 9	0.17	45	0.84
LEU	< 244	0.49	< 39	0.47	< 57	0.69	< 26	0.32	< 34	0.41	< 28	0.34	< 60	0.73
MET	64	0.60	< 7	0.39	< 9	0.51	< 7	0.39	11	0.62	14	0.79	16	0.90
PHE	< 114	0.49	< 19	0.49	< 25	0.65	< 15	0.39	< 18	0.46	< 12	0.31	< 25	0.65
PRO	259	1.05	< 0	0.00	> 160	3.88	37	0.90	31	0.75	< 19	0.46	< 12	0.29
VAL	< 202	0.52	< 29	0.45	< 34	0.53	< 49	0.76	< 14	0.22	< 9	0.14	67	1.04
ARG	267	0.97	< 22	0.48	41	0.90	38	0.83	34	0.74	> 60	1.31	> 72	1.57
ASP	> 538	1.69	> 230	4.33	55	1.03	> 114	2.15	52	0.98	54	1.02	< 33	0.62
GLU	373	1.01	< 39	0.64	74	1.21	> 99	1.62	< 37	0.60	< 40	0.65	> 84	1.37
LYS	> 423	1.28	< 18	0.33	> 75	1.36	> 84	1.52	57	1.03	> 88	1.60	> 101	1.83
ASN	> 411	1.69	> 122	3.00	38	0.94	> 56	1.38	> 65	1.60	> 93	2.29	37	0.91
CYS	> 142	1.58	> 65	4.35	< 5	0.33	< 4	0.27	> 56	3.74	< 4	0.27	8	0.54
GLN	178	0.84	< 13	0.37	< 17	0.48	34	0.96	< 24	0.68	> 48	1.35	42	1.18
HIS	147	1.10	> 35	1.57	25	1.12	18	0.81	26	1.17	22	0.99	21	0.94
SER	> 455	1.36	> 93	1.67	61	1.10	> 84	1.51	> 107	1.92	51	0.92	59	1.06
THR	> 414	1.36	< 33	0.65	51	1.01	49	0.97	> 175	3.46	< 29	0.57	> 77	1.52
TRP	40	0.49	7	0.51	7	0.51	8	0.58	< 3	0.22	< 6	0.44	9	0.65
TYR	131	0.67	< 19	0.58	< 19	0.58	24	0.73	< 21	0.64	< 12	0.37	36	1.10
GLY	> 641	1.69	60	0.95	57	0.90	> 82	1.30	> 104	1.65	> 281	4.45	57	0.90
OTH	7	0.35	0	0.00	1	0.30	1	0.30	0	0.00	2	0.60	3	0.90
ALL	5490	1.00	915	1.00	915	1.00	915	1.00	915	1.00	915	1.00	915	1.00

Table 8.34: Overall frequencies and conformational parameter for reverse π turns (statistical over representation: '>', statistical under representation: '<')

8.7 Discussion

Several general conclusions could be drawn from the analysis of the amino acid propensities found in different turn families.

Pro is often observed as under represented at the first position of reverse turns or at the terminal position in normal turn since Pro lacks a polar NH hydrogen atom that could be involved in a hydrogen bond. Conversely, Pro shows an over representation at the last position in some open turn-types that do not need a hydrogen bond between the first and last residue. Interestingly, this is not the case in general which is presumably due to the fact that the C α -C α distance was increased and therefore a hydrogen bond is not possible. As expected, Pro shows also an over representation when a cis peptide bond occurs.

The main deviation in the amino acid propensities of certain amino acids in specific turn-types can be explained by the conformational requirements at this position (see also chapter 7.5). Gly populates region on the Ramachandran plot that are unique and if the turn-type requires these conformational preferences only given for Gly, this residue is usually highly over represented. Furthermore, Asn, Asp, Thr and Ser show also conformations with ψ around $\pm 180^\circ$, which leads to an over representation in turn-types with this conformation. In contrast, Val, Ile and Thr show a branch at C β and therefore, they never occur with conformations in the left-handed α -region.

Interestingly, the reverse turns show a generally high under representation of the hydrophobic residues Ala, Ile, Leu, Met, Phe, Pro, Val and a high over representation of Gly, Asp and Asn. There are two putative explanations to these observations. On the one hand, the reverse turns should be usually located at the protein surface, since hydrophobic residues are generally unfavoured. On the other hand, the specific conformations that Gly, Asp and Asn exhibit (conformations with ψ around $\pm 180^\circ$), are needed to form the reverse hydrogen bond.

Another aspect is the amino acid propensities of similar turn types where only one torsion angle traverses the torsion angle boundary of $\pm 180^\circ$ (e.g. open α -turn type Ia/Ib). It seems to be the case that based on the accessible conformational space Pro induces a ϕ torsion angle smaller 180° (Ia), whereas a Gly reveals a torsion angle bigger than -180° (Ib). In both cases, the corresponding amino acid shows a higher occurrence.

Finally, the classification into the three categories normal, open and reverse turns seems to be reasonable since the turn families show different amino acid propensities and different torsion angles of otherwise related conformations. Furthermore, the increase of the C α_i - C α_{i+n}

distance followed by a visual identification of clusters with turn conformations afterwards appears to be necessary. Comparing the C α -C α distance of turn-types in this study with already described turn-types, the chosen cut-off in former analysis appears artificial and would ignore similar conformations.

Another interesting aspect of this general clustering is showed in Figure 8.34. It compares the amount of secondary structure elements being present. It either gives the original strand and helix classification found in the original pdb file or as an expansion the β -strands and turns, based on the classification of this work. Table 8.35 shows the percentage of classified residues within the dataset used. All classifications are retrieved from Secbase using Reliscript, whereas the helix and strand classification are based on the PDB categorization.

The used turn information is based on all clustered turn or turn-like structures. This includes the hook and kink cluster, but exclude all “no class” and “broken” structures and the non-turn clusters. The overall percentage is less than the sum of the three groups of secondary structure elements, since turns can also occur within both, strands and helices. The latter are built up from multiple turn structures and, e.g., γ -turns can occur in β -strands. Considering the 96.1%, most of the protein chains seems to be fully classified now by helices, β -sheets and turns and the amount of random chain is minimized.

SSE	mean	stdev
turn	89.4 %	8.4%
strand	22.3%	16.5%
helix	42.4%	24.7%
overall	96.1%	4.0%

Table 8.35: Percentage of residues that are included in a secondary structure element within dataset

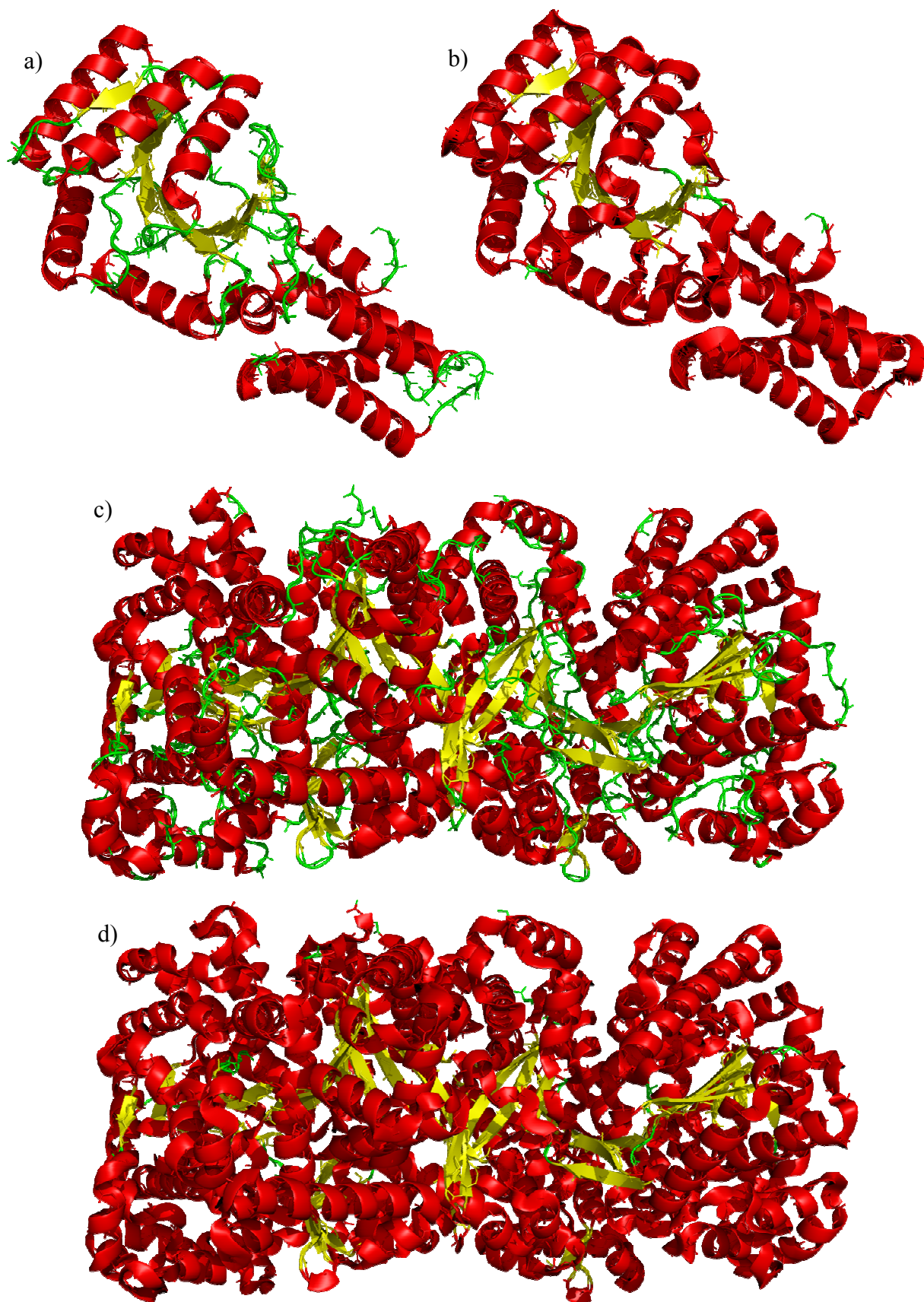


Figure 8.34: Example structures showing secondary structure elements (a, b: 2c03, c, d: 2c4m)
a) and c): original pdb β -strand (yellow) and helix (red) assignment
b) and d): original pdb β -strand assignment (yellow) and turn structures classified in this work (red)
green: random chain

8.8 Differences to known classification

As a summary, this section should point out the major differences of the methodology used in this work compared to former approaches and shows some selected examples to explain differences to already established classifications

8.8.1 Benefit of clustering using ESOMs

In contrast to previously applied approaches, clustering using ESOMs leads to clear data separation and the assignment to a specific cluster or turn type only depends on the overall distance similarity of all torsion angles within this cluster. It is not based on any predefined torsion angle ranges for each position independently from the whole turn conformation. Using independent mean torsion angles and specific tolerance for each position, as it was used in previous classification, represents more a comparison based on a single section of this turn. Therefore, the conformation as a whole is not regarded and in consequence quite different conformations can be considered, especially once an increasing number of torsion angles are included. Additionally, the allowed tolerance is adjusted arbitrarily. The applied nomenclature based on the region in the Ramachandran plot shows the same disadvantages.

Comparing the amino acid propensities of γ -turns to previously reported propensities (Guruprasad 2000) exhibit major differences with Ile, Met, Pro and Val showing quite different representations (see Table 8.36). Ile at position 2 and Val and position 1 and 2 are highly over represented in contrast to the results of Guruprasad et al. (Guruprasad 2000). The authors showed an under representation. In contrast, the occurrence of Pro at position 1 and 2 and Met reflects the opposite.

There are two reasons for this pronounced discrepancy: Firstly, the actual dataset is more than four times bigger than the dataset used by Guruprasad et al. Secondly, the definition of γ -turns differs. Guruprasad et al. used the program PROMOTIF (Hutchinson 1996) by Hutchinson and Thornton. In contrast to the present analysis, they defined the turn-type based on standard angles for normal (ϕ, ψ : 75° , -64°) and inverse (ϕ, ψ : -79° , 69°) turns and a maximum deviation of 40° . Figure 8.35 shows a Ramachandran plot where the colour indicates the number of occurrence. The blue rectangle encompasses the range of γ -turns used by Guruprasad et al. to calculate their amino acid propensities. It is clearly evident that inverse γ -turns, which are not included, fall in a range highly occupied by Val and Ile (see section

13.5). This explains the high occurrence of these amino acids in this analysis and the difference in number of occurrences to the previously study.

aa	P1		P2		P3	
	own	G	own	G	own	G
ALA	< 0.71	0.79	< 0.55	0.98	< 0.72	0.62
ILE	> 1.50	1.32	> 2.03	0.48	> 1.48	1.19
LEU	> 1.08	0.85	> 1.21	0.79	> 1.28	0.95
MET	< 0.83	1.11	< 0.86	1.40	1.06	0.64
PHE	> 1.23	1.12	> 1.07	0.86	> 1.17	1.04
PRO	< 0.90	1.68	< 0.66	2.32	< 0.00	0.00
VAL	> 1.58	0.86	> 1.73	0.35	> 1.76	1.50
ARG	< 0.90	1.05	< 0.78	0.70	< 0.89	1.20
ASP	< 0.66	0.73	> 1.67	2.87	< 0.88	1.09
GLU	< 0.76	0.96	< 0.72	0.75	< 0.73	0.55
LYS	< 0.92	0.83	< 0.80	0.79	< 0.87	0.77
ASN	< 0.90	0.98	> 1.40	2.61	0.96	1.50
CYS	0.96	1.14	> 1.12	0.81	> 1.20	1.47
GLN	< 0.84	0.81	< 0.83	0.84	< 0.78	1.10
HIS	1.07	0.97	1.02	1.23	1.02	1.07
SER	< 0.69	0.73	< 0.53	0.51	< 0.85	1.06
THR	> 1.07	0.76	< 0.82	0.35	> 1.30	1.61
TRP	1.09	1.08	1.10	1.31	> 1.36	1.55
TYR	> 1.31	1.16	1.04	0.77	> 1.16	1.19
GLY	1.03	1.50	< 0.20	0.37	< 0.68	0.75

Table 8.36: Comparison of inverse γ -turns amino acid propensities revealed in this analysis to those published by Guruprasad et al. (Guruprasad 2000)

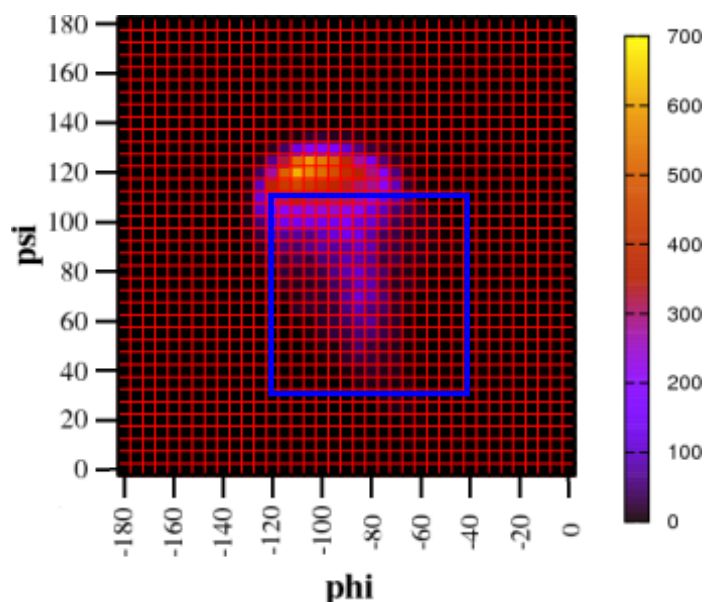


Figure 8.35: Torsion angle distribution of inverse γ turn-types (blue rectangle indicates torsion angle ranges used by Guruprasad et al.)

An alternative classification scheme is based on the region of the Ramachandran plot (see section 6.7.3), however, also this strategy has some limitations. Each position of a turn is classified by the region, the torsion angles are assigned to and turns with equivalent classification belong to the same turn type. Table 8.37 shows some reverse α -turn types.

turn-type	designation
Ia	PAAaP
Ib	PAAaP
Ic	PAAaP

Table 8.37: Reverse α -turn-types and designation after Dasgupta et al. (see section 8.5.3 for details)

Although these turn-types share similar conformations, the amino acid propensities reveal differences because of exceeding the torsion angle boundaries. In the present example, this classification scheme would fail to differentiate between the three different turn-types Ia/Ib/Ic.

Another important aspect is the advantage of ESOMs as a machine learning method with respect to clustering. On the one hand, the training is unsupervised and therefore no anticipated estimation the number of expected clusters is required. As a consequence, the number of detected clusters is not biased by any preconceived knowledge and the clustering starts from scratch. On the other hand, the trained ESOM with identified clusters could easily be used to classify turn-structures from novel proteins. This option is applied in Relibase to classify turns of all stored proteins. This information is made available in Secbase (see chapter 2). Turn types falling out of the current scheme will be classified as “no class” and in the case the number of unclassified turn types found in novel crystal structures increases beyond a certain level, a new clustering will be performed possible showing new turn types.

8.8.2 Increase of $C\alpha$ - $C\alpha$ distance for open turns

Instead of using a short cut-off to identify open turns, the maximum $C\alpha$ - $C\alpha$ distance was increased to 10 Å. Previously, the maximum distance was set to 6.5 Å for open α -turns (Dasgupta 2004) or 7.0 Å for open β -turns (Chou 2000). As mentioned already by Rose et al. (Rose 1985), a preselected distance cutoff is always a compromise between over inclusion and over exclusion of possible turn conformations. A visual inspection of possible turn structures would be better, since it is less parameter sensitive. Therefore, a cut-off of 10 Å was chosen to include every possible open turn conformation. After clustering, the grouped structures were visual inspected and only turn structures were chosen.

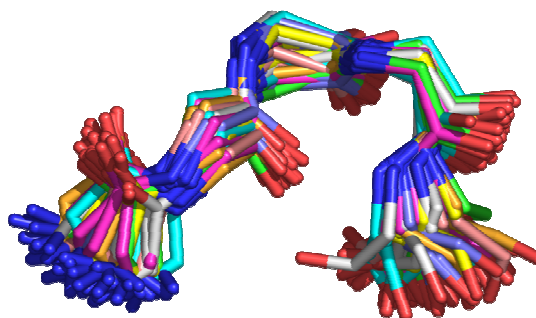


Figure 8.36: 100 randomly selected open α -turn type Ia

The C α -C α distance distribution of open β -turn type VIIIa (see Figure 8.15), which was already described by Hutchinson and Thornton (Hutchinson 1994) shows a maximum at 7.9 Å scattered over a range between 4.8 Å and 9.6 Å. The open α -turn type Ia (see Figure 8.25) shows a maximum at 6.3 Å in a range between 4.7 Å and 9.5 Å. Figure 8.36 shows 100 randomly selected open α -turns of turn-type Ia and this image clearly shows the similarity within this turn-type. Using a cut-off of 7.0 Å or 7.5 Å would ignore a large fraction of both described turn-types and with hindsight, this cut-off seems to be arbitrary. This also applies to other turn-types.

8.8.3 Differentiation between open and normal turns

Either a hydrogen bond or a C α -C α distance criterion was used during former analysis. Instead of including all hydrogen bonded turns that pass the distance criterion they were split in this analysis leading to normal and open turns. Although normal and open turn-types can share similar conformations (Table 8.38) they can show different amino acid propensities (Table 8.39). Therefore, a partitioning into different families, which was previously never anticipated in this way before, appears reasonable.

cluster	type	ω_i		ϕ_{i+1}		ψ_{i+1}		ω_{i+1}		ϕ_{i+2}		ψ_{i+2}		ω_{i+2}	
		\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm	\emptyset	\pm
normal	I	-179.0	3.7	-61.2	10.1	-25.7	15.2	-180.0	3.3	-82.6	20.0	-6.3	19.1	-180.0	3.9
open	I	-179.9	3.1	-79.1	17.7	-32.1	18.8	-178.3	3.7	-83.8	24.7	-25.3	20.1	-179.2	3.4

Table 8.38: Mean torsion angles for normal and open β -turn type I

aa	Position 1		Position 2		Position 3		Position 4	
	normal	open	normal	open	normal	open	normal	open
ALA	0.99	1.26	1.39	1.17	1.12	1.07	0.87	1.08
ILE	0.85	0.70	0.63	0.95	0.36	0.90	0.81	0.66
LEU	1.02	0.91	0.95	1.21	0.94	1.47	1.08	0.86
MET	0.87	0.99	0.94	1.14	0.99	1.16	0.99	1.03
PHE	0.89	0.85	0.67	0.95	0.91	1.18	1.13	0.89
PRO	1.46	0.92	2.36	0.41	0.32	0.08	0.00	0.72
VAL	0.69	0.54	0.52	0.83	0.31	0.78	0.66	0.66
ARG	0.88	1.26	1.07	1.17	1.08	1.05	0.94	0.92
ASP	1.65	1.24	1.04	1.05	1.68	0.86	1.02	0.84
GLU	0.89	1.42	1.36	1.49	1.24	1.00	0.74	0.70
LYS	0.92	1.29	1.22	1.39	1.12	1.14	0.99	1.05
ASN	1.30	1.10	0.79	0.93	1.82	1.17	1.29	1.37
CYS	1.00	0.70	0.78	0.78	0.79	1.12	1.07	1.05
GLN	0.86	1.19	1.07	1.27	1.31	1.08	1.17	1.14
HIS	1.21	0.87	0.88	1.03	1.42	1.20	1.18	1.04
SER	1.16	1.27	1.37	0.88	1.47	1.01	1.00	1.11
THR	0.85	0.87	0.65	0.86	0.87	1.35	0.79	0.70
TRP	1.11	0.84	0.99	1.02	1.08	1.08	1.02	0.68
TYR	0.87	0.74	0.66	0.87	1.01	1.26	1.11	0.80
GLY	0.80	0.65	0.49	0.38	0.60	0.34	2.16	2.39
OTH	0.86	1.08	0.75	0.88	0.85	1.06	0.85	0.93

Table 8.39: Amino acid propensities for normal and open β -turn type I (bold indicates big differences)

8.8.4 Introduce of reverse turns

The putative occurrence of a reverse hydrogen bonding between NH_i and CO_{i+n} was first described by Toniolo (Toniolo 1980) and later supported for δ -turns by a theoretical analysis from Nagarajaram et al. (Nagarajaram 1992). The present study provides the first general analysis of protein turn structures exhibiting a reverse hydrogen bond that reveals some interesting results. The three reverse turn families share a similar amino acid propensity with a high occurrence of Gly, Asp and Asn in all families and additional Cys and Thr among reverse π -turns. In contrast, the hydrophobic residues Ala, Ile, Leu, Met, Phe, and Val show a very low occurrence in these families.

Additionally, reverse turns can share similar turn conformation with turn-types of other turn families. For example, the reverse β -turn types I'/II' show also a normal hydrogen bonding and the mean torsion angles indicate, that these types are also part of the normal β -turn types I'/II'. This confirms the hypothesis by Wilmot and Thornton (Wilmot 1988), that due to steric hinderance, these turn-types should be stabilized by an additional hydrogen bond.

8.8.5 The peptide bond torsion angle

Chou and Blinn (Chou 1997) mentioned that for β -turn types VIa1, VIa2 and VIb the $i+2$ position should be occupied by a cis-proline. Otherwise, the $C\alpha$ - $C\alpha$ distance would be larger than the cut-off of 7 Å. Although, this is generally accepted, it is usually not mentioned that the other turn-types should adopt a trans peptide-bond. Of course, turn-types with similar ϕ and ψ angles, but deviating ω angles can exhibit different conformation. Therefore, the ω angle was included in our clustering from beginning and the final clustering reveals specific clusters for either cis or trans peptide bonds.

8.8.6 Summary of turn-types

Finally, Table 8.40 shows a summary of all turn-families and a comparison to former classifications. The Table reveals a partial overlap with known classifications, but in general, this analysis differs from these classifications.

family	structures	turn-types	overlap with former classifications	
reverse	2	194	9 [4]	-
normal	3	20162	2	2: different mean torsion angles / amino acid propensities (Guruprasad et al. (Guruprasad 2000))
reverse	3	114	5	-
normal	4	28650	6	5 of 8 (Hutchinson and Thornton (Hutchinson 1994))
open	4	137101	11 [10] + 6 [4] kink-types	6 of 8 (Hutchinson and Thornton (Hutchinson 1994))
reverse	4	1957	18 [13]	-
normal	5	91726	9 [8]	6 of 9 (Nataraj et al. (Nataraj 1995))
open	5	19607	21 [18] + 1 kink-type	6 of 15 (Dasgupta et al. (Dasgupta 2004))
reverse	5	1340	21 [15]	-
normal	6	3995	8 [7]	5 of 11 (Rajanshkar et al. (Rajashankar 1996))
open	6	21204	22 [18] + 6 hook-types	-
reverse	6	954	13 [11]	-

Table 8.40: Summary of final turn-types and overlap with former classification

([] indicates number of different conformations, where similar turn-types are combined that cross the angle boundary in one torsion angle)

8.9 Further results and discussion

8.9.1 Principal Component Analysis

Turn structures can contain up to 17 different torsion angles, leading to a very high dimensional data set. Principal component analysis was used to investigate whether the dimensionality could be reduced (Table 8.41).

Looking at the results for the normal γ -turn class (Figure 8.37) it is evident that the dimensionality of this data set could be reduced without losing too much information about the data variance. Using the first three principal components, 95.7% of the variance could be explained. Each of the components represents a combination of various different torsion angles.

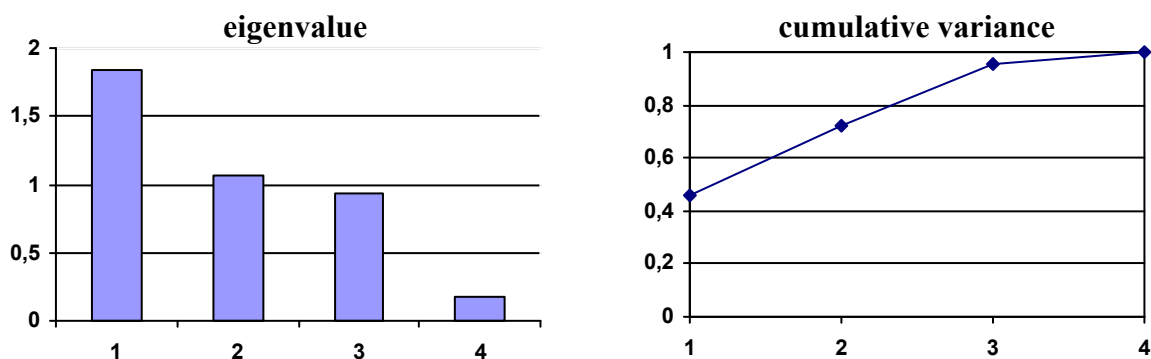


Figure 8.37: Eigenvalues and cumulative variance of normal γ -turn data

A similar result could be seen among the normal β -turn data (Figure 8.38), the last principal component could also be clearly identified and neglected without losing too much of information. The remaining first six principal components explain 99.1% of the variance. However, there is no reason for reducing the dimensionality of the dataset from seven to six with the help of a principal component analysis for clustering the data since only little gain in performance could be achieved.

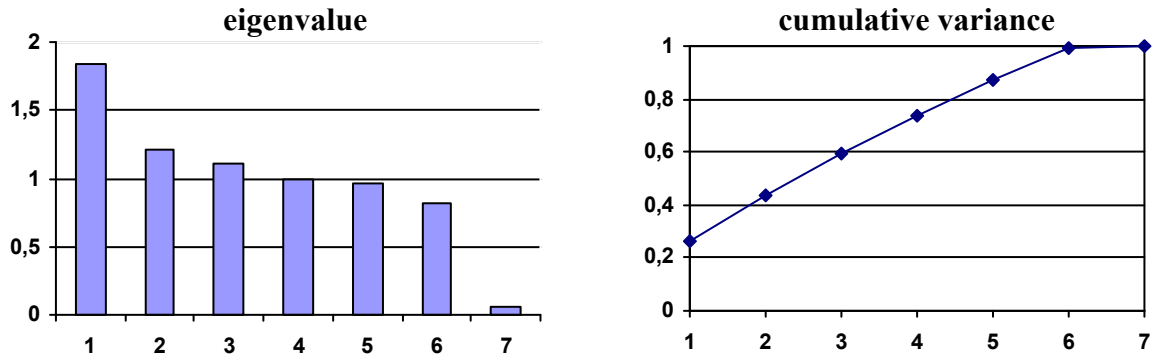


Figure 8.38: Eigenvalues and cumulative variance of normal β turn data

The other principal component analysis (Figure 8.39 - Figure 8.43) show analogous behaviour, reduction in dimensionality is hardly indicated. Thus, the principal component analysis clearly showed that only a minor or no correlations exists between the variables and there is no option to reduce the dimensionality of the data. Therefore, all torsion angles were used for clustering.

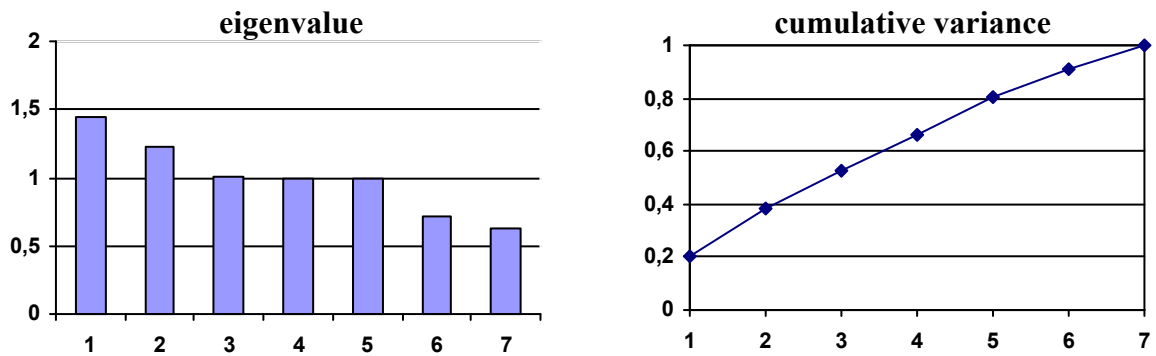


Figure 8.39: Eigenvalues and cumulative variance of open β turn data

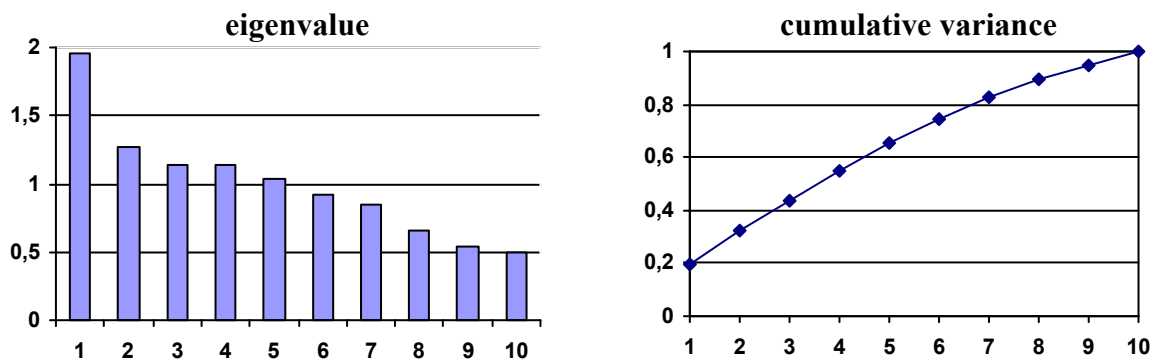


Figure 8.40: Eigenvalues and cumulative variance of normal α turn data

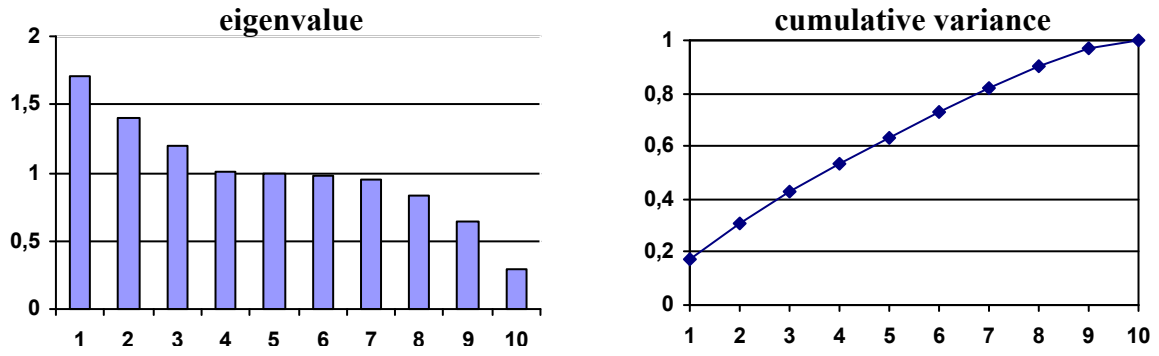


Figure 8.41: Eigenvalues and cumulative variance of normal α turn data

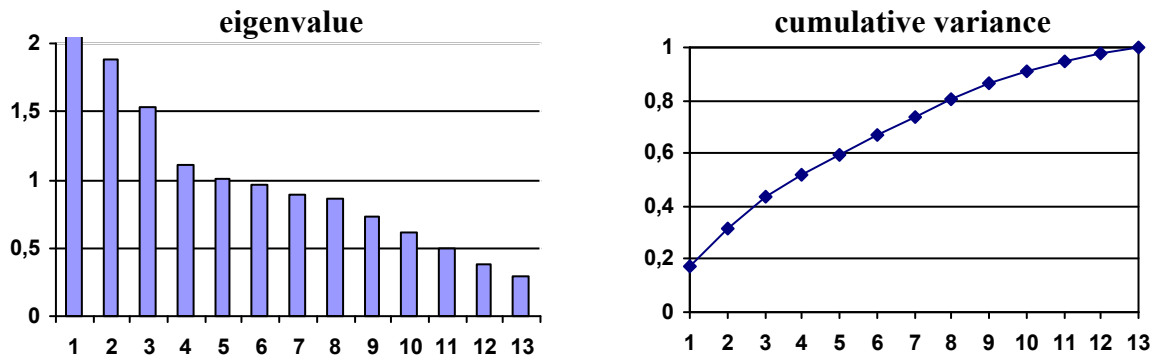


Figure 8.42: Eigenvalues and cumulative variance of normal π turn data

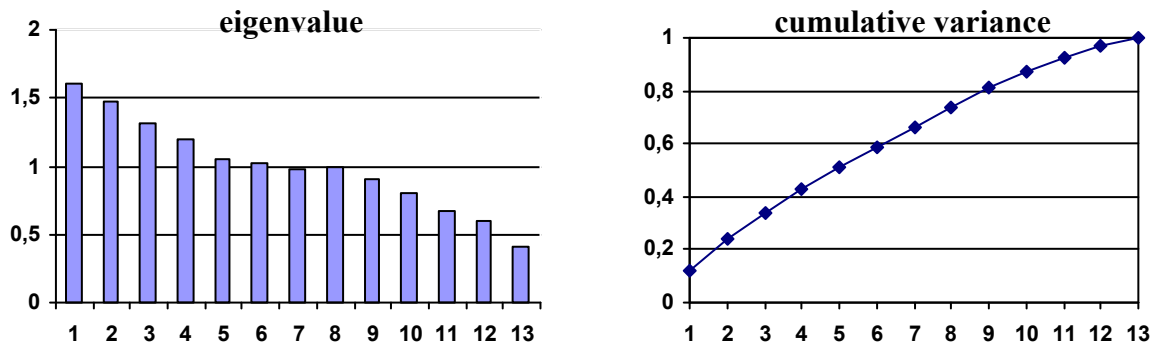


Figure 8.43: Eigenvalues and cumulative variance of open π turn data

γ -turn		1	2	3	4									
norm.	e.w.	1,835	1,064	0,930	0,171									
	c. v.	0,459	0,725	0,957	1,000									
β -turn		1	2	3	4	5	6	7						
norm.	e.w.	1,836	1,218	1,112	0,999	0,961	0,813	0,060						
	c. v.	0,262	0,436	0,595	0,738	0,875	0,991	1,000						
open	e.w.	1,443	1,219	1,002	0,989	0,999	0,716	0,631						
	c. v.	0,206	0,380	0,523	0,665	0,808	0,910	1,000						
α -turn		1	2	3	4	5	6	7	8	9	10			
norm.	e.w.	1,951	1,276	1,139	1,132	1,035	0,925	0,844	0,653	0,542	0,503			
	c. v.	0,195	0,323	0,437	0,550	0,653	0,746	0,830	0,896	0,950	1,000			
open	e.w.	1,706	1,399	1,194	1,014	0,994	0,979	0,947	0,830	0,646	0,293			
	c. v.	0,171	0,310	0,430	0,531	0,631	0,729	0,823	0,906	0,971	1,000			
π -turn		1	2	3	4	5	6	7	8	9	10	11	12	13
norm.	e.w.	2,220	1,879	1,539	1,106	1,005	0,967	0,897	0,861	0,734	0,617	0,497	0,383	0,295
	c. v.	0,171	0,315	0,434	0,519	0,596	0,670	0,739	0,806	0,862	0,910	0,948	0,977	1,000
open	e.w.	1,607	1,472	1,316	1,190	1,053	1,026	0,975	0,986	0,898	0,802	0,673	0,597	0,404
	c. v.	0,124	0,237	0,338	0,430	0,511	0,590	0,665	0,740	0,810	0,871	0,923	0,969	1,000

Table 8.41: Eigenvalues (e.v.) and cumulative variance (c.v.) for the open and normal turn data set

8.9.2 Prediction of turns in protein structure

The data compiled in this new and consistent turn classification, presented in this study, was already successfully applied for the turn prediction in given amino acid sequences (Meissner 2008). Since the number of known protein sequences is much higher than the number of known protein structures, successful and reliable structure prediction methods are desperately needed.

The first turn prediction methods were only based on the amino acid propensities (Chou 1974) but current methods use machine learning algorithms such as artificial neural networks or support vector machines (Cai 2003; Kaur 2003). We used support vector machines (SVM) for turn versus non-turn classification and probabilistic neural networks (PNN) for the prediction of certain normal β -turn-types (Meissner 2008). The SVM classifiers show a prediction accuracy of 80% and the PNN distinguish between the normal β -turn-types with 80% accuracy.

But, interestingly, the analysis also shows that the classification and predictive performance improve when open turns are excluded from hydrogen-bonded turns as putative geometries in the training and validation datasets. This conclusion is in agreement with the results in chapter 1, where comparable normal and open turn-types show different amino acid propensities despite having similar torsion angles. Additionally, the predictive performance for open turns is less distinct from non-turns than the hydrogen bonded turns. This indicates that the amino acid propensities shows lesser variation from random chain than the propensity of hydrogen bonded turns which makes it more difficult to distinguish between non-turns and open turns. Thus, this result supports the conclusion that it is reasonable to distinguish between open and normal turns and not combining them.

Finally, this analysis showed, that the new turn-type classification is distinct and well-defined and leads to a successful turn prediction without the help of additional information such as sequence alignment (Fuchs 2005) or longer sequences.

9 Turn Classification: Summary and Outlook

9.1 Summary and Outlook

Here, a comprehensive and consistent classification for all known and unknown turn families is presented. Based on current structural data, this classification is exhaustive and likely covers most of the possible conformations of turns. Furthermore, the designations of the different turn-types are in accordance with the already known evidence in literature. All turn-types from the different turn families were compared with already known types and named accordingly in case the torsion angles coincide similar ranges.

It could be shown that the separation of the different turn classes into the three turn categories (normal, open and reverse) is reasonable. The amino acid propensities and the different turn-types are dissimilar enough to be discriminated. In particular, the reverse turn conformations are completely different from the others, they seem to be appropriate to expose charged or polar amino-acid sidechains. Additionally, augmenting the $C\alpha$ - $C\alpha$ distance for the open turns to 10 Å with a subsequent visual inspection gives convincing results. Examples are detected that would otherwise be ignored.

A detailed analysis and discussion of the high amount of now available data about turn structures is beyond the scope of this contribution, some interesting conclusion could be drawn. For example, the hypothesis of Wilmot and Thornton (Wilmot 1988) that the normal β -turn-type I' must be stabilized by an additional hydrogen bond due to destabilizing steric conflicts could be confirmed. The β -turn-type I' and II' contains turn conformations with an additional reverse hydrogen bond. Additionally, most of the significant amino acid propensities could be explained by the conformational properties of the individual amino acids.

Furthermore, the predictive power of the new classification was already demonstrated by an enhanced assignment of turn types to given amino acid sequences (Meissner 2008). The subsequent study indicates that the differentiation between hydrogen-bonded and open turns increase the prediction quality for hydrogen bonded turns.

Hopefully, this consistent classification of all turn families will be supportive for turn prediction based on the amino-acid sequence and therefore also for protein fold and structure prediction. Furthermore, the classification could be used for the identification of similar

protein domains or structural motifs within different turn families and accordingly for the understanding of protein-ligand and protein-protein interactions. Additionally, it could reveal new information for the development of turn surrogates as bioisosteric replacements in peptides to develop new peptidomimetic drugs.

10 Summary, Zusammenfassung

10.1 Summary

The main focus of this work is the development of new methods for computer aided drug design. This development and the final tools are described together with studies to show possible applications and to prove the usefulness of these tools.

The first part of this work describes the integration of secondary structure element information together with geometric descriptions into a protein-ligand database (Relibase), which leads to the new modul Secbase. The python-based interface to Relibase (Reliscript) was used to add the information. Furthermore, the C++ core code of Relibase was extended to get access to this data and to add Secbase constraints to the substructure search. This leads to the opportunity to specify Secbase constraints within the substructure search accessible through the Relibase webinterface.

The motivation for the development of Secbase is guided through two main ideas: Firstly, Secbase should provide means to analyse protein-ligand interactions with respect to secondary structure elements and, secondly, should allow analysis and discovery of functional similarity within related folding patterns. This is based on the knowledge that the function of a protein is often based on the structure and the spatial structure is more conserved in evolution than amino acid sequence. In general, Secbase, in combination with Relibase, can be used for knowledge discovery about the influence of secondary structural elements on protein-ligand interactions and should be valuable for structure based drug design and molecular modelling.

Two major analyses were carried out using Reliscript and the Relibase Webinterface. The first analysis revealed some notable trends in hydrogen bonding geometry across the different secondary structural elements. The mean hydrogen bond length of accumulated hydrogen bonds in α -helices and parallel β -sheets decrease with increasing number of helix-turns and number of β -strands, respectively. The cooperative effect, which leads to a decrease of the mean hydrogen bond length, can be explained by a similar directionality of the peptide bond dipole vectors and the backbone hydrogen bonds in α -helices and parallel β -sheets. The second analysis describes a survey of water molecules next to the N-terminus of an α -helix

and shows their involvement in ligand binding. Furthermore, the kinked backbone shows interactions between two neighbored backbone amide groups and carboxylate or phosphate groups, respectively. In agreement with theoretical calculations described in the literature, this analysis suggest that the first/last turn of an α -helix is the main source for charge stabilising effects, mainly by providing hydrogen bonds. This is in contrast to the widely used explanation that the overall dipole of the helix has an influence.

The second part of this work deals with turns as an irregular secondary structure element with a hydrogen bond or a specific $C\alpha$ - $C\alpha$ distance between the first and the last residue. Because of the irregularity, the classification into subfamilies changed over the last decades with growing data from protein structures and is not completely adapted to the actual data basis, yet. Additionally, there was a lack of an overall classification for all turn families. Therefore, a uniform classification for all normal ($CO_i - NH_{i+n}$ hydrogen bond), open (a $C\alpha_i$ - $C\alpha_{i+n}$ distance up to 10 Å) and reverse ($NH_i - CO_{i+n}$ hydrogen bond) turn families is presented based on current structural data.

The emergent self-organizing maps (ESOM) were used to cluster all turn-conformations of a non-redundant protein chain dataset. In combination with β -sheet and helix classification on average 96% of the given protein chain is now successfully classified. The classification can be used for the identification of similar protein domains or structural motifs within different turn families and accordingly for the understanding of protein-ligand and protein-protein interactions. The created turn classification was used to classify the turn conformations within all protein structures. This information was also added to Secbase. Protein sequence-based turn prediction with high accuracy has already confirmed this new categorization based on machine learning methods as consistent and well-defined. Hopefully, this classification will also be supportive for protein fold and structure prediction.

10.2 Zusammenfassung

Die Hauptzielsetzung dieser Arbeit war die Methodenentwicklung im Bereich des computerunterstützten Wirkstoffdesigns. Diese Entwicklung und die finalen Werkzeuge und Programme werden zusammen mit durchgeführten Analysen beschrieben, welche die mögliche Anwendungen und die Nützlichkeit dieser Programme verdeutlichen.

Der erste Teil dieser Arbeit beschreibt die Integration der Information über Sekundärstrukturelemente und ihrer geometrischen Beschreibungen in die Protein-Ligand Datenbank Relibase. Reliscript, die Python-basierte Schnittstelle, wurde genutzt um diese Daten hinzuzufügen, dies führte zu dem Modul Secbase. Des Weiteren wurde der Relibase C++-Kern erweitert um Zugriff auf diese Informationen und zusätzliche Beschränkungen für die Substruktur-Suche zu erhalten. Diese können innerhalb des Relibase-Webinterfaces genutzt werden um die Substruktur-Suche auf spezielle Secbase-Informationen einzuschränken.

Die Motivation für die Secbase basiert auf zwei Ideen: Einerseits soll Secbase die Methoden zur Verfügung stellen, um Protein-Ligand Interaktionen unter der speziellen Berücksichtigungen der Sekundärstrukturelemente zu untersuchen. Andererseits soll ermöglicht werden, ähnliche Funktionsmechanismen innerhalb von verwandten Faltungsmustern zu entdecken und zu analysieren. Dies basiert auf dem Wissen, dass die Funktion eines Proteins auf seiner Struktur basiert und die räumliche Anordnung während der Evolution höher konserviert ist als die Aminosäuresequenz. Im Allgemeinen soll die Kombination aus Secbase und Relibase dazu genutzt werden, um Erkenntnisse über den Einfluss der Sekundärstrukturelemente auf die Protein-Ligand Interaktion zu erlangen. Ausserdem sollte Secbase von Nutzen sein für das computerunterstützte Wirkstoffdesign und das Molecular Modelling.

Zwei ausführliche Analysen wurden mit Hilfe von Reliscript und dem Relibase Webinterface durchgeführt. Die erste beschreibt interessante Trends der Wasserstoffbrücken-Geometrien innerhalb von Sekundärstrukturelementen. Die durchschnittliche Länge von hintereinanderliegenden H-Brücken in α -Helices und parallelen β -Faltblättern verringert sich mit zunehmender Anzahl an Helix-Schleifen bzw. β -Strängen. Der kooperative Effekt, der zur Abnahme der durchschnittlichen Wasserstoffbrücken-Länge führt, kann mit Hilfe von ähnlich ausgerichteten Peptidbindungs-Dipolen und Rückgrat-Wasserstoffbrücken in α -Helices und β -Faltblättern erklärt werden. Die zweite Analyse beschreibt eine Untersuchung von Wassermolekülen in der Nähe des N-terminalen Ende einer α -Helix und

deren Mitwirkung an der Ligandbindung. Des Weiteren zeigt das gewinkelte Kettenrückgrat eine Interaktion zwischen zwei benachbarten Amid-Gruppen und Carboxyl- bzw. Phosphatgruppen. In Übereinstimmung mit theoretischen Untersuchungen aus der Literatur zeigt diese Analyse, dass vor allem der erste bzw. letzte Turn innerhalb einer α -helix für mögliche Ladungsstabilisierung verantwortlich ist, hauptsächlich durch das Vorhandensein von möglichen Wasserstoffbrückenpartnern. Dies steht im Gegensatz zur allgemeinen Annahme, dass der gesamte Dipol der Helix einen Einfluss auf die Ladungsstabilisierung hat.

Der zweite Teil der Arbeit handelt von Turns als irreguläres Sekundärstrukturelement mit einer Wasserstoffbrücke oder einer speziellen $C\alpha$ - $C\alpha$ Distanz zwischen der ersten und der letzten Aminosäure. Aufgrund dieser Irregularität und der wachsenden Zahl an Proteinstrukturen, hat sich die Klassifizierung in unterschiedliche Unterfamilien über die letzten Jahrzehnte immer wieder verändert und ist noch nicht an die aktuellen Datenlage angepasst. Zusätzlich fehlte noch eine allgemeine Klassifizierung für alle Turn-Familien. Aus diesem Grund wird hier eine Klassifizierung für alle normalen ($CO_i - NH_{i+n}$ H-Brücke), offenen (max. $C\alpha_i - C\alpha_{i+n}$ Distanz von 10 Å) und reversen ($NH_i - CO_{i+n}$ H-Brücke) präsentiert, die auf der aktuellen Datenlage basiert.

Die emergenten selbst-organisierenden Merkmalskarten (engl. emergent self-organizing maps, ESOM) wurden benutzt, um alle Turn-Konformationen eines nicht-redundanten Protein-Ketten Datensatzes zu clustern. Zusammen mit der Information über β -Faltblätter und Helices sind im Durchschnitt nun 96% der Protein-Kette erfolgreich klassifiziert. Diese Klassifizierung kann für die Identifizierung von ähnlichen Protein Domänen oder Struktur-Motiven innerhalb unterschiedlicher Turn-Familien benutzt werden und somit für das Verständnis von Protein-Ligand- und Protein-Protein-Interaktionen. Die resultierende Turn-Klassifizierung wurde genutzt um alle Turn Konformationen innerhalb von Protein Strukturen zu klassifizieren. Diese Information wurde ebenfalls der Seabase hinzugefügt. Protein sequenz-basierte Vorhersage von Turns basierend auf Methoden des Maschinellen Lernens mit hoher Genauigkeit hat diese neue Klassifizierung bereits als konsistent und wohl-definiert bestätigt. Hoffentlich unterstützt diese Klassifizierung somit auch die Faltungs- und Strukturvorhersage von Proteinen.

11 Literature

- Aqvist, J., Luecke, H., Quioco, F. A., Warshel, A. (1991). "Dipoles localized at helix termini of proteins stabilize charges." Proc Natl Acad Sci U S A **88**: 2026 - 2030.
- Aurora, R., Rose, G. D. (1998). "Helix capping." Protein Sci **7**(1): 21-38.
- Baker, E. N., Hubbard, R. E. (1984). "Hydrogen bonding in globular proteins." Progress in Biophysics & Molecular Biology **44**(2): 97-179.
- Bergner, A., Günther, J., Hendlich, M., Klebe, G., Verdonk, M. (2001). "Use of Relibase for retrieving complex three-dimensional interaction patterns including crystallographic packing effects." Biopolymers **61**(2): 99-110.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. (2000). "The Protein Data Bank." Nucleic Acids Res **28**(1): 235-42.
- Block, P. (2002). Secbase - Ein Werkzeug zur Korrelation von sekundären Strukturmustern mit der Ligandenbindung. Pharmaceutical Institute. Halle-Wittenberg, Martin-Luther-Universität Halle-Wittenberg.
- Bour, P., Keiderling, T. A. (2004). "Structure, spectra and the effects of twisting of beta-sheet peptides. A density functional theory study." Journal of Molecular Structure - Theochem **675**(1-3): 95-105.
- Brakch, N., El Abida, B., Rholam, M. (2006). "Functional role of β -Turn in Polypeptide Structure and its use as Template to Design Therapeutic Agents." Central Nervous System Agents in Medicinal Chemistry **6**: 163-173.
- Branden, C., Tooze, J. (1999). Introduction to Protein Structure.
- Breinbauer, R., Vetter, I. R., Waldmann, H. (2002). "From protein domains to drug candidates-natural products as guiding principles in the design and synthesis of compound libraries." Angew Chem Int Ed Engl **41**(16): 2879-90.
- Cai, Y. D., Liu, X. J., Li, Y. X., Xu, X. B., Chou, K. C. (2003). "Prediction of beta-turns with learning machines." Peptides **24**(5): 665-9.
- Calabrese, J. C., Jordan, D. B., Boodhoo, A., Sariaslani, S., Vannelli, T. (2004). "Crystal structure of phenylalanine ammonia lyase: multiple helix dipoles implicated in catalysis." Biochemistry **43**(36): 11403-16.
- Chakrabarti, P., Pal, D. (1998). "Main-chain conformational features at different conformations of the side-chains in proteins." Protein Eng **11**(8): 631-47.
- Chakrabarti, P., Pal, D. (2001). "The interrelationships of side-chain and main-chain conformations in proteins." Prog Biophys Mol Biol **76**(1-2): 1-102.
- Chen, S., Chrusciel, R. A., Nakanishi, H., Raktabutr, A., Johnson, M. E., Sato, A., Weiner, D., Hoxie, J., Saragovi, H. U., Greene, M. I. and Kahn, M. (1992). "Design and synthesis of a CD4 β -turn mimetic that inhibits human immunodeficiency virus envelope glycoprotein gp120 binding and infection of human lymphocytes." Proc. Natl. Acad. Sci. USA **89**: 5872-5876.
- Chou, K. C. (2000). "Prediction of tight turns and their types in proteins." Anal Biochem **286**(1): 1-16.

- Chou, K. C., Blinn, J. R. (1997). "Classification and prediction of beta-turn types." J Protein Chem **16**(6): 575-95.
- Chou, P. Y., Fasman, G. D. (1974). "Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins." Biochemistry **13**(2): 211-22.
- Chou, P. Y., Fasman, G. D. (1977). "Beta-turns in proteins." J Mol Biol **115**(2): 135-75.
- Cochran, D. A., Penel, S., Doig, A. J. (2001). "Effect of the N1 residue on the stability of the alpha-helix for all 20 amino acids." Protein Sci **10**(3): 463-70.
- Cui, Q., Elstner, M., Kaxiras, E., Frauenheim, T., Karplus, M. (2001). "A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method." Journal of Physical Chemistry B **105**(2): 569-585.
- Dasgupta, B., Pal, L., Basu, G., Chakrabarti, P. (2004). "Expanded turn conformations: characterization and sequence-structure correspondence in alpha-turns with implications in helix folding." Proteins **55**(2): 305-15.
- de Brevern, A. G., Etchebest, C., Hazout, S. (2000). "Bayesian Probabilistic Approach for Predicting Backbone Structures in Terms of Protein Blocks." Proteins **41**: 271-287.
- DeLano, W. L. (2002). The PyMOL Molecular Graphics System. San Carlos, CA, USA, DeLano Scientific.
- Doig, A. J., Baldwin, R. L. (1995). "N- and C-capping preferences for all 20 amino acids in alpha-helical peptides." Protein Sci **4**(7): 1325-36.
- Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T., MacKinnon, R. (2002). "X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity." Nature **415**(6869): 287-94.
- Efimov, A. V. (1993). "Standard structures in proteins." Prog Biophys Mol Biol **60**(3): 201-39.
- Eisenberg, D. (2003). "The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins." Proc Natl Acad Sci U S A **100**(20): 11207-10.
- Elstner, M., Jalkanen, K. J., Knapp-Mohammady, M., Frauenheim, T., Suhai, S (2000). "DFT studies on helix formation in N-acetyl-(L-alanyl)(n)-N'-methylamide for n=1-20." Chemical Physics **256**(1): 15-27.
- Engel, D. E., DeGrado, W. F. (2004). "Amino acid propensities are position-dependent throughout the length of alpha-helices." J Mol Biol **337**(5): 1195-205.
- Forsyth, W. R., Antosiewicz, J. M., Robertson, A. D. (2002). "Empirical relationships between protein structure and carboxyl pKa values in proteins." Proteins **48**(2): 388-403.
- Fuchs, P. F., Alix, A. J. (2005). "High accuracy prediction of beta-turns and their types using propensities and multiple alignments." Proteins **59**(4): 828-39.
- Grishin, N. V. (2001). "Fold change in evolution of protein structures." J Struct Biol **134**(2-3): 167-85.
- Guruprasad, K., Rajkumar, S. (2000). "Beta-and gamma-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials." J Biosci **25**(2): 143-56.

- Hendlich, M., Bergner, A., Gunther, J., Klebe, G. (2003). "Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions." J Mol Biol **326**(2): 607-20.
- Hirsch, A. K., Fischer, F. R., Diederich, F. (2007). "Phosphate recognition in structural biology." Angew Chem Int Ed Engl **46**(3): 338-52.
- Ho, B. K., Curmi, P. M. G. (2002). "Twist and shear in beta-sheets and beta-ribbons." J Mol Biol **317**(2): 291-308.
- Hobohm, U., Sander, C. (1994). "Enlarged representative set of protein structures." Protein Sci **3**(3): 522-4.
- Hol, W. G. (1985). "The role of the alpha-helix dipole in protein function and structure." Prog Biophys Mol Biol **45**(3): 149-95.
- Hol, W. G., Halie, L. M., Sander, C. (1981). "Dipoles of the alpha-helix and beta-sheet: their role in protein folding." Nature **294**(5841): 532-6.
- Hovmoller, S., Zhou, T., Ohlson, T. (2002). "Conformations of amino acids in proteins." Acta Crystallogr D Biol Crystallogr **58**(Pt 5): 768-76.
- Hunter, C. G., Subramaniam, S. (2003). "Protein Fragment Clustering and Canonical Local Shapes." Proteins **50**: 580-588.
- Hutchinson, E. G., Thornton, J. M. (1994). "A revised set of potentials for beta-turn formation in proteins." Protein Sci **3**(12): 2207-16.
- Hutchinson, E. G., Thornton, J. M. (1996). "PROMOTIF--a program to identify and analyze structural motifs in proteins." Protein Sci **5**(2): 212-20.
- Improta, R., Barone, V., Kudin, K. N., Scuseria, G. E. (2001). "Structure and conformational behavior of biopolymers by density functional calculations employing periodic boundary conditions. I. The case of polyglycine, polyalanine, and poly-alpha-aminoisobutyric acid in vacuo." JOURNAL OF THE AMERICAN CHEMICAL SOCIETY **123**(14): 3311-3322.
- Ireta, J., Neugebauer, J., Scheffler, M., Rojo, A., Galvan, M. (2003). "Density functional theory study of the cooperativity of hydrogen bonds in finite and infinite alpha-helices." JOURNAL OF PHYSICAL CHEMISTRY B **107**(6): 1432-1437.
- Jaravine, V. A., Alexandrescu, A. T., Grzesiek, S. (2001). "Observation of the closing of individual hydrogen bonds during TFE-induced helix formation in a peptide." PROTEIN SCIENCE **10**(5): 943-950.
- Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T., MacKinnon, R. (2003). "X-ray structure of a voltage-dependent K⁺ channel." Nature **423**(6935): 33-41.
- Kabsch, W., Sander, C. (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features." Biopolymers **22**(12): 2577-637.
- Karpen, M. E., de Haseth, P. L., Neet, K. E. (1992). "Differences in the amino acid distributions of 3(10)-helices and alpha-helices." Protein Sci **1**(10): 1333-42.
- Kaur, H., Raghava, G. P. (2003). "Prediction of beta-turns in proteins from multiple alignment using neural network." Protein Sci **12**(3): 627-34.
- Kee, K. S., Jois, S.D. (2003). "Design of β -turn Based Therapeutic Agents." Current Pharmaceutical Design **9**: 1209-1224.

- Kobko, N., Dannenberg, J. J. (2003). "Cooperativity in amide hydrogen bonding chains. Relation between energy, position, and H-bond chain length in peptide and protein folding models." Journal of Physical Chemistry A **107**(48): 10389-10395.
- Koch, M. A., Waldmann, H. (2005). "Protein structure similarity clustering and natural product structure as guiding principles in drug discovery." Drug Discov Today **10**(7): 471-83.
- Koch, O. (2004). Ein Werkzeug zur Korrelation von sekundären Strukturmustern mit der Ligandenbindung: Erweiterung der Seabase. Pharmaceutical Institute. Halle-Wittenberg, Martin-Luther-Universität Halle-Wittenberg. **Diploma**.
- Koch, O., Bocola, M., Klebe, G. (2005). "Cooperative effects in hydrogen-bonding of protein secondary structure elements: a systematic analysis of crystal data using Seabase." Proteins **61**(2): 310-7.
- Koehl, P., Levitt, M. (1999). "Structure-based conformational preferences of amino acids." Proc Natl Acad Sci U S A **96**(22): 12524-9.
- Kohonen, T. (1997). Self-Organizing Maps, Springer.
- Koonin, E. V., Wolf, Y. I., Karev, G. P. (2002). "The structure of the protein universe and genome evolution." Nature **420**(6912): 218-23.
- Kupas, K., Ultsch, A., Klebe, G. (2004). "Comparison of substructural epitopes in enzyme active sites using self-organizing maps." J Comput Aided Mol Des **18**(11): 697-708.
- Lario, P. I., Vrielink, A. (2003). "Atomic resolution density maps reveal secondary structure dependent differences in electronic distribution." J Am Chem Soc **125**(42): 12787-94.
- Lewis, P. N., Momany, F. A., Scheraga, H. A. (1973). "Chain reversals in proteins." Biochim Biophys Acta **303**(2): 211-29.
- Matthews, B. W. (1972). "The γ -turn, Evidence for a New Folded Conformation in Proteins." Macromolecules **5**(6): 818 - 819.
- McDonald, I. K., Thornton, J. M. (1994). "Satisfying hydrogen bonding potential in proteins." J Mol Biol **238**(5): 777-93.
- Meissner, M., Koch, O., Klebe, G., Schneider, G. (2008). "Prediction of Turns in Protein Structure by Kernel-Based Machine Learning Methods." Proteins **accepted**.
- Milner-White, E. J. (1990). "Situations of gamma-turns in proteins. Their relation to alpha-helices, beta-sheets and ligand binding sites." J Mol Biol **216**(2): 386-97.
- Milner-White, E. J., Ross, B. M., Ismail, R., Belhadj-Mostefa, K., Poet, R. (1988). "One type of gamma-turn, rather than the other gives rise to chain-reversal in proteins." J Mol Biol **204**: 777 - 782.
- Miranda, J. J. (2003). "Position-dependent interactions between cysteine residues and the helix dipole." Protein Sci **12**(1): 73-81.
- Motta, A., Reches, M., Pappalardo, L., Andreotti, G., Gazit, E. (2005). "The preferred conformation of the tripeptide Ala-Phe-Ala in water is an inverse gamma-turn: implications for protein folding and drug design." Biochemistry **44**(43): 14170-8.
- Nagarajaram, H. A., Paul, P. K., Ramanarayanan, K., Soman, K. V., Ramakrishnan, C. (1992). "Conformational studies on beta-bend containing a cis peptide unit." Int J Pept Protein Res **40**(5): 383-94.

- Nataraj, D. V., Srinivasan, N., Sowdhamini, R., Ramakrishnan, C. (1995). "Alpha-Turns in Protein Structures." Current Science **69**(5): 434-447.
- Némethy, G., Printz, M. P. (1972). "The γ -turn, a possible folded conformation of the polypeptide chain. Comparison with the β -turn." Macromolecules **5**(6): 755 - 758.
- Nemethy, G., Scheraga, H. A. (1980). "Stereochemical requirements for the existence of hydrogen bonds in beta-bends." Biochem Biophys Res Commun **95**(1): 320-7.
- Pal, L., Basu, G., Chakrabarti, P. (2002). "Variants of 3(10)-helices in proteins." Proteins **48**(3): 571-9.
- Panasik, N., Jr., Fleming, P. J., Rose, G. D. (2005). "Hydrogen-bonded turns in proteins: the case for a recount." Protein Sci **14**(11): 2910-4.
- Pauling, L., Corey, R. B. (1951). "The pleated sheet, a new layer configuration of polypeptide chains." Proc Natl Acad Sci U S A **37**(5): 251-6.
- Pauling, L., Corey, R. B., Branson, H. R. (1951). "The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain." Proc Natl Acad Sci U S A **37**(4): 205-11.
- Pavone, V., Gaeta, G., Lombardi, A., Natri, F., Maglio, O., Isernia, C., Saviano, M. (1996). "Discovering protein secondary structures: Classification and description of isolated alpha-turns." Biopolymers **38**(6): 705-721.
- Penel, S., Hughes, E., Doig, A. J. (1999). "Side-chain structures in the first turn of the alpha-helix." J Mol Biol **287**(1): 127-43.
- Porter, M. A., Hall, J. R., Locke, J. C., Jensen, J. H., Molina, P. A. (2006). "Hydrogen bonding is the prime determinant of carboxyl pKa values at the N-termini of alpha-helices." Proteins **63**(3): 621-35.
- Rajashankar, K. R., Ramakumar, S. (1996). "pi-Turns in proteins and peptides: Classification, conformation, occurrence, hydration and sequence." Protein Science **5**(5): 932-946.
- Ramachandran, G. N., Sasisekharan, V. (1968). "Conformation of polypeptides and proteins." Adv Protein Chem **23**: 283-438.
- Ramakrishnan, C., Nataraj, D. V. (1998). "Energy minimization studies on alpha-turns." Journal of Peptide Science **4**(4): 239-252.
- Richardson, J. S. (1981). "The anatomy and taxonomy of protein structure." Adv Protein Chem **34**: 167-339.
- Richardson, J. S., Richardson, D. C. (1988). "Amino acid preferences for specific locations at the ends of alpha helices." Science **240**(4859): 1648-52.
- Rick, S. W., Stuart, S. J. (2002). "Potentials and algorithms for incorporating polarizability in computer simulations." REVIEWS IN COMPUTATIONAL CHEMISTRY, VOL 18: 89-146.
- Rose, G. D., Gierasch, L. M., Smith, J. A. (1985). "Turns in peptides and proteins." Adv Protein Chem **37**: 1-109.
- Sander, O., Sommer, I, Lengauer, T. (2006). "Local protein structure prediction using discriminative models." BMC Bioinformatics **7**(14): <http://www.biomedcentral.com/1471-2105/7/14>.
- Saraste, M., Sibbald, P. R., Wittinghofer, A. (1990). "The P-loop--a common motif in ATP- and GTP-binding proteins." Trends Biochem Sci **15**(11): 430-4.

- Schellmann, C. (1980). The α L conformation at the ends of helices, Amsterdam: Elsevier.
- Schuchhardt, J., Schneider, G., Reichelt, J., Schomburg, D., Wrede, P. (1996). "Local structural motifs of protein backbones are classified by self-organizing neural networks." Protein Eng **9**(10): 833-42.
- Srinivasan, R., Rose, G. D. (1999). "A physical basis for protein secondary structure." Proc Natl Acad Sci U S A **96**(25): 14258-63.
- Stahl, M., Taroni, C., Schneider, G. (2000). "Mapping of protein surface cavities and prediction of enzyme class by a self-organizing neural network." Protein Eng **13**(2): 83-8.
- Thomas, A., Benhabiles, N., Meurisse, R., Ngwabije, R., Brasseur, R (2001). "Pex, analytical tools for PDB files. II. H-Pex: Noncanonical H-bonds in alpha-helices." Proteins - Structure, Function and Genetics **43**(1): 37-44.
- Tidor, B. (1994). "Helix-Capping Interaction in λ Cro Protein: A Free Energy Simulation Analysis." Proteins **19**: 310 - 323.
- Toniolo, C. (1980). "Intramolecularly hydrogen-bonded peptide conformations." CRC Crit Rev Biochem **9**(1): 1-44.
- Topol, I. A., Burt, S. K., Derety, E., Tang, T. H., Perczel, A., Rashin, A., Csizmadia, I. G. (2001). "alpha- and 3(10)-helix interconversion: a quantum-chemical study on polyalanine systems in the gas phase and in aqueous solvent." J Am Chem Soc **123**(25): 6054-60.
- Tyndall, J. D., Pfeiffer, B., Abbenante, G., Fairlie, D. P. (2005). "Over one hundred peptide-activated G protein-coupled receptors recognize ligands with turn structure." Chem Rev **105**(3): 793-826.
- Ultsch, A. (1999). Data Mining and Knowledge Discovery with Emergent Self-Organizing Feature Maps for Multivariate Time Series. Kohonen Maps. E. Oja, Kaski, S.: 33-46.
- Ultsch, A. (2003). Maps for the Visualization of high-dimensional Data Spaces. Proceedings of Workshop on Self-Organizing Maps (WSOM 2003), Kyushu, Japan.
- Venkatachalam, C. M. (1968). "Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units." Biopolymers **6**(10): 1425-36.
- Wilmot, C. M., Thornton, J. M. (1988). "Analysis and prediction of the different types of beta-turn in proteins." J Mol Biol **203**(1): 221-32.
- Wilmot, C. M., Thornton, J. M. (1990). "Beta-turns and their distortions: a proposed new nomenclature." Protein Eng **3**(6): 479-93.
- Wu, Y. D., Zhao, Y. L. (2001). "A theoretical study on the origin of cooperativity in the formation of 3(10)- and alpha-helices." J Am Chem Soc **123**(22): 5313-9.
- Yan, A., Gasteiger, J. (2003). "Prediction of aqueous solubility of organic compounds based on a 3D structure representation." J Chem Inf Comput Sci **43**(2): 429-34.

12 Acknowledgements

I would first like to acknowledge all the people who made this work possible and I am grateful to all of them. Especially, I want to thank the following people:

- *Prof. Dr. G. Klebe* for his supervision, the interesting topic and the possibility to investigate your own ideas. It was a great experience to work in his group.
- *Prof. Dr. G. Schneider* and his Ph.D student *Michael Meissner* for the fruitful cooperation concerning the turn prediction and *Prof. Dr. G. Schneider* for being the 2nd Examiner
- My girlfriend, *Elke Andress* for her support, especially during the stressful time at the final stage of this thesis
- *Dr. Marco Bocola* for the fruitful cooperation on hydrogen bonding in secondary structure elements
- *Angela Scholz* for taking care of the whole group and keeping the group running
- All the system-administrators for keeping the group network running, especially *Dr. Hans Velec*, who introduced me into Windows-Administration and *Gerd Neudert* who kept my computer running after I left.
- *Dr. Jason Cole* and *Dr. Greg Shields* for the supervision during my research stay at the Cambridge Crystallographic Data Centre
- The Cambridge Crystallographic Data Centre for financial support
- *Dr. Andreas Heine* and *Dr. Jason Cole* for carefully reading of this manuscript and helpful suggestion
- The whole group for this nice working atmosphere

and special thanks to my parents *Rolf* and *Gisela Koch*

13 Appendix

13.1 two residue turns

13.1.1 δ -turns

cluster 1: turn-type Ia

aa	turn		position 1		position 2		before		after	
	f	P	f	P	f	P	f	P	f	P
ALA	13	0.96	2	0.29	11	1.62	7	1.03	11	1.62
ILE	13	1.24	> 10	1.90	3	0.57	6	1.14	7	1.33
LEU	16	0.99	8	0.99	8	0.99	7	0.86	5	0.62
MET	1	0.29	1	0.57	0	0.00	0	0.00	0	0.00
PHE	7	0.92	2	0.52	5	1.31	4	1.05	3	0.79
PRO	1	0.12	< 0	0.00	1	0.25	7	1.73	7	1.73
VAL	6	0.48	5	0.79	< 1	0.16	8	1.27	8	1.27
ARG	6	0.67	2	0.44	4	0.89	6	1.33	3	0.67
ASP	8	0.77	6	1.15	2	0.38	3	0.57	3	0.57
GLU	8	0.66	5	0.83	3	0.50	4	0.66	9	1.49
LYS	9	0.83	6	1.11	3	0.55	3	0.55	2	0.37
ASN	12	1.50	5	1.25	7	1.75	3	0.75	< 0	0.00
CYS	4	1.36	1	0.68	3	2.04	2	1.36	2	1.36
GLN	13	1.86	6	1.72	7	2.01	3	0.86	2	0.57
HIS	9	2.05	5	2.28	4	1.82	2	0.91	4	1.82
SER	13	1.19	5	0.91	8	1.46	5	0.91	9	1.65
THR	17	1.71	9	1.81	8	1.61	5	1.00	1	0.20
TRP	5	1.85	2	1.48	3	2.22	1	0.74	1	0.74
TYR	9	1.40	5	1.55	4	1.24	5	1.55	1	0.31
GLY	10	0.81	5	0.81	5	0.81	8	1.29	> 12	1.93
OTH	0	0.00	0	0.00	0	0.00	1	3.06	0	0.00
ALL	180	1.00	90	1.00	90	1.00	90	1.00	90	1.00

cluster 2: turn-type Ib

aa	turn		position 1		position 2		before		after	
	f	P	f	P	f	aa	f	P	f	P
ALA	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ILE	3	2.34	> 3	4.67	0	0.00	0	0.00	0	0.00
LEU	3	1.51	> 3	3.03	0	0.00	0	0.00	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	1	1.07	1	2.15	0	0.00	1	2.15	1	2.15
PRO	0	0.00	0	0.00	0	0.00	1	2.02	0	0.00
VAL	2	1.30	2	2.59	0	0.00	1	1.30	0	0.00
ARG	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASP	> 6	4.70	0	0.00	> 6	9.39	> 5	7.83	> 3	4.70
GLU	0	0.00	0	0.00	0	0.00	0	0.00	> 3	4.08

LYS	1	0.75	1	1.51	0	0.00	1	1.51	0	0.00
ASN	2	2.05	0	0.00	> 2	4.09	0	0.00	1	2.05
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	0	0.00	0	0.00	0	0.00	1	2.35	0	0.00
HIS	1	1.87	1	3.73	0	0.00	0	0.00	0	0.00
SER	1	0.75	0	0.00	1	1.50	0	0.00	> 3	4.49
THR	1	0.82	0	0.00	1	1.64	1	1.64	0	0.00
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLY	1	0.66	0	0.00	1	1.32	0	0.00	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	22	1.00	11	1.00	11	1.00	11	1.00	11	1.00

cluster 3: turn-type IIa

aa	turn		position 1		position 2		before		after	
	f	P	f	P	f	f	P	f	P	f
ALA	5	0.87	0	0.00	5	1.74	2	0.70	2	0.70
ILE	3	0.68	1	0.45	2	0.90	1	0.45	1	0.45
LEU	6	0.88	0	0.00	6	1.75	1	0.29	3	0.88
MET	0	0.00	0	0.00	0	0.00	0	0.00	1	1.36
PHE	1	0.31	0	0.00	1	0.62	1	0.62	0	0.00
PRO	2	0.58	0	0.00	2	1.17	> 7	4.09	> 5	2.92
VAL	1	0.19	1	0.38	0	0.00	2	0.75	4	1.50
ARG	1	0.26	0	0.00	1	0.53	2	1.05	2	1.05
ASP	5	1.13	3	1.36	2	0.91	2	0.91	4	1.81
GLU	3	0.59	0	0.00	3	1.18	2	0.79	3	1.18
LYS	3	0.66	1	0.44	2	0.87	3	1.31	2	0.87
ASN	1	0.30	0	0.00	1	0.59	1	0.59	1	0.59
CYS	3	2.42	0	0.00	> 3	4.83	2	3.22	2	3.22
GLN	1	0.34	0	0.00	1	0.68	1	0.68	2	1.36
HIS	3	1.62	0	0.00	> 3	3.24	0	0.00	0	0.00
SER	2	0.43	1	0.43	1	0.43	3	1.30	2	0.87
THR	4	0.95	0	0.00	4	1.90	2	0.95	2	0.95
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	1	0.37	0	0.00	1	0.74	3	2.21	0	0.00
GLY	> 31	5.91	> 31	11.82	0	0.00	3	1.14	1	0.38
OTH	0	0.00	0	0.00	0	0.00	0	0.00	> 1	7.25
ALL	76	1.00	38	1.00	38	1.00	38	1.00	38	1.00

cluster 4: turn-type class IIb

aa	turn		position 1		position 2		before		after	
	f	P	f	P	f	f	P	f	P	f
ALA	1	2.21	0	0.00	1	4.41	0	0.00	0	0.00
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LEU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PRO	0	0.00	0	0.00	0	0.00	> 1	7.39	0	0.00
VAL	0	0.00	0	0.00	0	0.00	1	4.75	0	0.00

ARG	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASP	1	2.87	0	0.00	> 1	5.74	0	0.00	> 1	5.74
GLU	1	2.49	0	0.00	1	4.98	0	0.00	0	0.00
LYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASN	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
HIS	0	0.00	0	0.00	0	0.00	> 1	13.68	0	0.00
SER	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
THR	0	0.00	0	0.00	0	0.00	0	0.00	> 1	6.02
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLY	> 3	7.24	> 3	14.48	0	0.00	0	0.00	1	4.83
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	6	1.00	3	1.00	3	1.00	3	1.00	3	1.00

cluster 5: turn-type IIIa

aa	turn		position 1		position 2		before		after	
	f	P	f	P	f	f	P	f	P	f
ALA	1	0.30	1	0.60	0	0.00	3	1.81	0	0.00
ILE	2	0.78	2	1.56	0	0.00	1	0.78	2	1.56
LEU	3	0.76	2	1.01	1	0.50	2	1.01	2	1.01
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	2	1.07	2	2.15	0	0.00	1	1.07	0	0.00
PRO	0	0.00	0	0.00	0	0.00	1	1.01	> 5	5.04
VAL	1	0.32	1	0.65	0	0.00	1	0.65	1	0.65
ARG	0	0.00	0	0.00	0	0.00	0	0.00	3	2.73
ASP	0	0.00	0	0.00	0	0.00	2	1.57	1	0.78
GLU	2	0.68	2	1.36	0	0.00	1	0.68	1	0.68
LYS	3	1.13	3	2.26	0	0.00	1	0.75	0	0.00
ASN	1	0.51	1	1.02	0	0.00	1	1.02	1	1.02
CYS	1	1.39	1	2.78	0	0.00	0	0.00	0	0.00
GLN	1	0.59	1	1.17	0	0.00	2	2.35	0	0.00
HIS	0	0.00	0	0.00	0	0.00	1	1.87	0	0.00
SER	2	0.75	2	1.50	0	0.00	1	0.75	1	0.75
THR	1	0.41	1	0.82	0	0.00	1	0.82	0	0.00
TRP	0	0.00	0	0.00	0	0.00	0	0.00	1	3.02
TYR	1	0.64	0	0.00	1	1.27	2	2.54	1	1.27
GLY	> 23	7.57	3	1.98	> 20	13.17	1	0.66	3	1.98
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	44	1.00	22	1.00	22	1.00	22	1.00	22	1.00

cluster 6: turn-type IIIb

aa	turn		position 1		position 2		before		after	
	f	P	f	P	f	f	P	f	P	f
ALA	2	0.74	2	1.47	0	0.00	1	0.74	1	0.74
ILE	2	0.95	2	1.90	0	0.00	2	1.90	1	0.95
LEU	0	0.00	0	0.00	0	0.00	2	1.23	2	1.23
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00

PHE	0	0.00	0	0.00	0	0.00	1	1.31	0	0.00
PRO	0	0.00	0	0.00	0	0.00	0	0.00	1	1.23
VAL	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ARG	1	0.56	1	1.11	0	0.00	1	1.11	2	2.22
ASP	3	1.44	> 3	2.87	0	0.00	1	0.96	1	0.96
GLU	2	0.83	1	0.83	1	0.83	1	0.83	2	1.66
LYS	1	0.46	1	0.92	0	0.00	0	0.00	0	0.00
ASN	3	1.88	> 3	3.75	0	0.00	0	0.00	0	0.00
CYS	0	0.00	0	0.00	0	0.00	1	3.40	0	0.00
GLN	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
HIS	3	3.42	> 2	4.56	1	2.28	> 2	4.56	1	2.28
SER	2	0.91	1	0.91	1	0.91	> 4	3.66	2	1.83
THR	1	0.50	1	1.00	0	0.00	0	0.00	0	0.00
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	0	0.00	0	0.00	0	0.00	1	1.55	> 3	4.66
GLY	> 16	6.44	1	0.81	> 15	12.07	1	0.81	2	1.61
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	36	1.00	18	1.00	18	1.00	18	1.00	18	1.00

cluster 7: turn-type IIIc

aa	turn		position 1		position 2		before		after	
	f	P	f	P	f	f	P	f	P	f
ALA	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LEU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PRO	0	0.00	0	0.00	0	0.00	> 1	11.09	0	0.00
VAL	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ARG	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLU	> 2	7.47	> 1	7.47	> 1	7.47	0	0.00	0	0.00
LYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASN	1	5.63	> 1	11.25	0	0.00	0	0.00	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
HIS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
SER	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
THR	0	0.00	0	0.00	0	0.00	> 1	9.03	> 1	9.03
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	0	0.00	0	0.00	0	0.00	0	0.00	> 1	13.97
GLY	1	3.62	0	0.00	> 1	7.24	0	0.00	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	4	1.00	2	1.00	2	1.00	2	1.00	2	1.00

cluster 8: turn-type IVa

aa	turn		position 1		position 2		before		after	
	f	P	f	P	f	f	P	f	P	f
ALA	0	0.00	0	0.00	0	0.00	0	0.00	> 2	5.30
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LEU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PRO	> 5	11.09	0	0.00	> 5	22.18	0	0.00	0	0.00
VAL	0	0.00	0	0.00	0	0.00	1	2.85	0	0.00
ARG	1	2.00	1	4.00	0	0.00	0	0.00	0	0.00
ASP	1	1.72	1	3.44	0	0.00	0	0.00	0	0.00
GLU	0	0.00	0	0.00	0	0.00	1	2.99	1	2.99
LYS	0	0.00	0	0.00	0	0.00	1	3.32	0	0.00
ASN	1	2.25	1	4.50	0	0.00	0	0.00	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	1	2.58	1	5.16	0	0.00	1	5.16	1	5.16
HIS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
SER	1	1.65	1	3.29	0	0.00	0	0.00	0	0.00
THR	0	0.00	0	0.00	0	0.00	0	0.00	1	3.61
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLY	0	0.00	0	0.00	0	0.00	1	2.90	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	10	1.00	5	1.00	5	1.00	5	1.00	5	1.00

cluster 9: turn-type IVb

aa	turn		position 1		position 2		before		after	
	f	P	f	P	f	f	P	f	P	f
ALA	0	0.00	0	0.00	0	0.00	> 1	6.62	0	0.00
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LEU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PRO	> 2	11.09	0	0.00	> 2	22.18	0	0.00	0	0.00
VAL	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ARG	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASN	0	0.00	0	0.00	0	0.00	> 1	11.25	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
HIS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
SER	0	0.00	0	0.00	0	0.00	0	0.00	> 1	8.22
THR	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	0	0.00	0	0.00	0	0.00	0	0.00	> 1	13.97
GLY	> 2	7.24	> 2	14.48	0	0.00	0	0.00	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00

ALL	4	1.00	2	1.00	2	1.00	2	1.00	2	1.00
-----	---	------	---	------	---	------	---	------	---	------

13.2 three residue turns

13.2.1 γ -turns

cluster 1: turn type inverse

aa	turn		position 1		position 2		position 3		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	< 2955	0.66	< 1060	0.71	< 820	0.55	< 1075	0.72	< 1241	0.83	< 1192	0.80
ILE	> 5773	1.67	> 1728	1.50	> 2337	2.03	> 1708	1.48	> 1296	1.13	< 931	0.81
LEU	> 6343	1.19	> 1924	1.08	> 2146	1.21	> 2273	1.28	< 1574	0.89	< 1228	0.69
MET	1052	0.92	< 316	0.83	< 329	0.86	407	1.06	354	0.93	< 254	0.66
PHE	> 2900	1.16	> 1024	1.23	> 896	1.07	> 980	1.17	859	1.03	< 757	0.91
PRO	< 1391	0.52	< 803	0.90	< 588	0.66	< 0	0.00	> 962	1.08	> 1317	1.48
VAL	> 7012	1.69	> 2181	1.58	> 2397	1.73	> 2434	1.76	> 1773	1.28	1371	0.99
ARG	< 2530	0.86	< 883	0.90	< 769	0.78	< 878	0.89	< 776	0.79	< 868	0.88
ASP	> 3664	1.07	< 750	0.66	> 1906	1.67	< 1008	0.88	< 991	0.87	> 1457	1.27
GLU	< 2914	0.74	< 1003	0.76	< 950	0.72	< 961	0.73	< 1055	0.80	< 1243	0.94
LYS	< 3076	0.86	< 1091	0.92	< 949	0.80	< 1036	0.87	< 1074	0.91	1162	0.98
ASN	> 2853	1.09	< 790	0.90	> 1223	1.40	840	0.96	915	1.05	> 1019	1.16
CYS	1055	1.09	309	0.96	> 359	1.12	> 387	1.20	349	1.08	325	1.01
GLN	< 1864	0.81	< 640	0.84	< 632	0.83	< 592	0.78	< 635	0.83	< 628	0.82
HIS	1493	1.04	514	1.07	491	1.02	488	1.02	509	1.06	484	1.01
SER	< 2478	0.69	< 822	0.69	< 640	0.53	< 1016	0.85	< 1109	0.93	> 1419	1.19
THR	> 3479	1.06	> 1164	1.07	< 897	0.82	> 1418	1.30	> 1198	1.10	> 1364	1.25
TRP	> 1052	1.18	324	1.09	325	1.10	> 403	1.36	305	1.03	268	0.91
TYR	> 2478	1.17	> 925	1.31	733	1.04	> 820	1.16	> 791	1.12	< 646	0.92
GLY	< 2592	0.64	1406	1.03	< 268	0.20	< 918	0.68	> 1809	1.33	> 1658	1.22
OTH	< 153	0.71	< 47	0.66	< 47	0.66	59	0.83	> 131	1.83	> 115	1.61
ALL	59107	1.00	19704	1.00	19702	1.00	19701	1.00	19706	1.00	19706	1.00

cluster 2: turn-type normal

aa	turn		position 1		position 2		position 3		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	73	0.72	25	0.74	< 16	0.48	32	0.95	34	1.01	29	0.86
ILE	< 45	0.58	< 11	0.42	< 11	0.42	23	0.89	< 13	0.50	26	1.00
LEU	< 71	0.59	38	0.95	< 6	0.15	< 27	0.67	34	0.85	35	0.87
MET	18	0.69	7	0.81	3	0.35	8	0.93	9	1.04	5	0.58
PHE	34	0.60	22	1.17	< 3	0.16	< 9	0.48	17	0.90	15	0.80
PRO	< 22	0.37	22	1.10	< 0	0.00	< 0	0.00	19	0.95	26	1.30
VAL	79	0.84	28	0.90	< 20	0.64	31	0.99	34	1.09	34	1.09
ARG	59	0.88	22	0.99	15	0.67	22	0.99	29	1.30	25	1.12
ASP	103	1.33	31	1.20	29	1.12	> 43	1.66	32	1.24	19	0.74
GLU	76	0.85	26	0.87	< 17	0.57	33	1.11	30	1.01	22	0.74
LYS	82	1.02	29	1.08	23	0.86	30	1.12	18	0.67	25	0.93
ASN	66	1.11	25	1.26	16	0.81	25	1.26	22	1.11	18	0.91
CYS	10	0.46	6	0.83	< 1	0.14	3	0.41	> 13	1.79	> 14	1.93
GLN	46	0.89	19	1.10	< 9	0.52	18	1.04	15	0.87	10	0.58

13.2.2 ϵ -turns

cluster 1: turn-type I

aa	turn		position 1		position 2		position 3		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	17	1.34	8	1.89	5	1.18	4	0.95	1	0.24	5	1.18
ILE	4	0.41	1	0.31	1	0.31	2	0.61	3	0.92	3	0.92
LEU	9	0.59	2	0.40	4	0.79	3	0.59	2	0.40	1	0.20
MET	2	0.61	1	0.92	1	0.92	0	0.00	0	0.00	1	0.92
PHE	4	0.56	0	0.00	4	1.69	0	0.00	5	2.11	2	0.84
PRO	1	0.13	0	0.00	0	0.00	1	0.40	3	1.19	> 14	5.55
VAL	8	0.68	< 0	0.00	4	1.02	4	1.02	5	1.27	2	0.51
ARG	3	0.36	1	0.36	1	0.36	1	0.36	2	0.71	0	0.00
ASP	10	1.02	2	0.62	4	1.23	4	1.23	2	0.62	6	1.85
GLU	15	1.33	4	1.07	7	1.87	4	1.07	< 0	0.00	1	0.27
LYS	9	0.89	2	0.59	4	1.19	3	0.89	2	0.59	3	0.89
ASN	9	1.21	3	1.21	4	1.61	2	0.80	3	1.21	2	0.80
CYS	1	0.36	0	0.00	0	0.00	1	1.09	0	0.00	1	1.09
GLN	7	1.08	1	0.46	0	0.00	> 6	2.76	2	0.92	1	0.46
HIS	6	1.47	1	0.73	2	1.47	3	2.20	> 4	2.93	> 4	2.93
SER	10	0.98	2	0.59	3	0.88	5	1.47	4	1.18	4	1.18
THR	15	1.61	4	1.29	> 7	2.26	4	1.29	5	1.61	3	0.97
TRP	3	1.19	2	2.38	0	0.00	1	1.19	1	1.19	2	2.38
TYR	6	1.00	1	0.50	2	1.00	3	1.50	> 5	2.50	0	0.00
GLY	> 29	2.50	> 21	5.43	3	0.78	5	1.29	7	1.81	1	0.26
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	168	1.00	56	1.00	56	1.00	56	1.00	56	1.00	56	1.00

cluster 2: turn-type II

aa	turn		position 1		position 2		position 3		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	4	0.80	1	0.60	2	1.20	1	0.60	1	0.60	1	0.60
ILE	8	2.08	0	0.00	2	1.56	> 6	4.67	0	0.00	0	0.00
LEU	5	0.84	0	0.00	3	1.51	2	1.01	0	0.00	1	0.50
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	1.07
PRO	0	0.00	0	0.00	0	0.00	0	0.00	2	2.02	2	2.02
VAL	4	0.86	0	0.00	1	0.65	3	1.94	3	1.94	1	0.65
ARG	1	0.30	0	0.00	1	0.91	0	0.00	2	1.82	0	0.00
ASP	3	0.78	2	1.57	0	0.00	1	0.78	3	2.35	2	1.57
GLU	3	0.68	0	0.00	2	1.36	1	0.68	2	1.36	1	0.68
LYS	3	0.75	0	0.00	1	0.75	2	1.51	> 4	3.02	> 4	3.02
ASN	3	1.02	2	2.05	1	1.02	0	0.00	1	1.02	1	1.02
CYS	1	0.93	0	0.00	1	2.78	0	0.00	0	0.00	1	2.78
GLN	1	0.39	0	0.00	1	1.17	0	0.00	1	1.17	1	1.17
HIS	3	1.87	1	1.87	> 2	3.73	0	0.00	0	0.00	0	0.00
SER	3	0.75	0	0.00	0	0.00	3	2.24	1	0.75	> 5	3.74
THR	3	0.82	0	0.00	1	0.82	2	1.64	1	0.82	0	0.00

TRP	2	2.02	1	3.02	1	3.02	0	0.00	1	3.02	0	0.00
TYR	3	1.27	1	1.27	1	1.27	1	1.27	0	0.00	0	0.00
GLY	> 16	3.51	> 14	9.22	2	1.32	0	0.00	0	0.00	1	0.66
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	66	1.00	22	1.00	22	1.00	22	1.00	22	1.00	22	1.00

cluster 3 : turn-type III

aa	turn		position 1		position 2		position 3		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	4	1.18	2	1.77	2	1.77	0	0.00	2	1.77	1	0.88
ILE	1	0.38	1	1.14	0	0.00	0	0.00	0	0.00	0	0.00
LEU	1	0.25	1	0.74	0	0.00	0	0.00	0	0.00	1	0.74
MET	1	1.14	1	3.43	0	0.00	0	0.00	0	0.00	0	0.00
PHE	0	0.00	0	0.00	0	0.00	0	0.00	> 3	4.72	0	0.00
PRO	1	0.49	0	0.00	0	0.00	1	1.48	1	1.48	2	2.96
VAL	2	0.63	1	0.95	0	0.00	1	0.95	0	0.00	1	0.95
ARG	3	1.33	0	0.00	1	1.33	2	2.67	1	1.33	2	2.67
ASP	2	0.77	0	0.00	0	0.00	2	2.30	1	1.15	1	1.15
GLU	3	1.00	1	1.00	0	0.00	2	1.99	0	0.00	1	1.00
LYS	2	0.74	0	0.00	0	0.00	2	2.21	0	0.00	1	1.11
ASN	2	1.00	1	1.50	1	1.50	0	0.00	0	0.00	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	0	0.00	0	0.00	0	0.00	0	0.00	1	1.72	0	0.00
HIS	1	0.91	1	2.74	0	0.00	0	0.00	1	2.74	1	2.74
SER	4	1.46	> 3	3.29	0	0.00	1	1.10	1	1.10	> 3	3.29
THR	2	0.80	0	0.00	1	1.20	1	1.20	1	1.20	1	1.20
TRP	3	4.43	1	4.43	0	0.00	> 2	8.87	0	0.00	0	0.00
TYR	1	0.62	0	0.00	1	1.86	0	0.00	> 2	3.73	0	0.00
GLY	> 12	3.86	2	1.93	> 9	8.69	1	0.97	1	0.97	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	45	1.00	15	1.00	15	1.00	15	1.00	15	1.00	15	1.00

cluster 4: turn-type IV

aa	turn		position 1		position 2		position 3		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	0	0.00	0	0.00	0	0.00	0	0.00	> 2	3.78	0	0.00
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LEU	1	0.53	1	1.58	0	0.00	0	0.00	2	3.17	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PRO	1	1.06	0	0.00	1	3.17	0	0.00	0	0.00	0	0.00
VAL	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	2.04
ARG	1	0.95	1	2.86	0	0.00	0	0.00	0	0.00	1	2.86
ASP	> 5	4.10	0	0.00	> 3	7.38	> 2	4.92	0	0.00	0	0.00
GLU	2	1.42	1	2.13	0	0.00	1	2.13	0	0.00	0	0.00
LYS	2	1.58	0	0.00	1	2.37	1	2.37	0	0.00	1	2.37
ASN	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	> 1	8.74	> 1	8.74
GLN	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	3.69

HIS	2	3.91	0	0.00	> 1	5.86	> 1	5.86	0	0.00	0	0.00
SER	2	1.57	> 2	4.70	0	0.00	0	0.00	0	0.00	0	0.00
THR	2	1.72	0	0.00	1	2.58	1	2.58	1	2.58	1	2.58
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	0	0.00	0	0.00	0	0.00	0	0.00	1	3.99	0	0.00
GLY	3	2.07	> 2	4.14	0	0.00	1	2.07	0	0.00	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	> 1	39.36
ALL	21	1.00	7	1.00	7	1.00	7	1.00	7	1.00	7	1.00

cluster 5: turn-type V

aa	turn		position 1		position 2		position 3		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	1	2.21	0	0.00	> 1	6.62	0	0.00	0	0.00	0	0.00
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LEU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	> 1	11.80
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
VAL	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ARG	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASP	1	2.87	0	0.00	> 1	8.61	0	0.00	0	0.00	0	0.00
GLU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASN	0	0.00	0	0.00	0	0.00	0	0.00	> 1	11.25	> 1	11.25
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
HIS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
SER	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
THR	0	0.00	0	0.00	0	0.00	0	0.00	> 1	9.03	0	0.00
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	> 2	9.31	> 2	27.94	0	0.00	0	0.00	0	0.00	0	0.00
GLY	2	4.83	0	0.00	0	0.00	> 2	14.48	0	0.00	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	6	1.00	2	1.00	2	1.00	2	1.00	2	1.00	2	1.00

13.3 four residue turns

13.3.1 normal β -turns

cluster 1: turn-type I

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P	f	P
ALA	> 7630	1.09	1726	0.99	> 2432	1.39	> 1960	1.12	< 1512	0.87	> 2041	1.17	< 1315	0.75
ILE	< 3559	0.66	< 1140	0.85	< 847	0.63	< 484	0.36	< 1088	0.81	1401	1.04	> 1452	1.08
LEU	8294	1.00	2127	1.02	< 1970	0.95	< 1954	0.94	> 2243	1.08	> 2577	1.24	> 2210	1.06
MET	1698	0.95	< 391	0.87	420	0.94	444	0.99	443	0.99	> 516	1.15	< 386	0.86
PHE	< 3524	0.90	< 869	0.89	< 656	0.67	< 889	0.91	> 1110	1.13	> 1075	1.10	> 1100	1.12
PRO	4312	1.03	> 1524	1.46	> 2454	2.36	< 334	0.32	< 0	0.00	> 1109	1.07	> 1673	1.61
VAL	< 3524	0.54	< 1114	0.69	< 841	0.52	< 499	0.31	< 1070	0.66	< 1504	0.93	< 1495	0.92
ARG	4581	0.99	< 1015	0.88	> 1232	1.07	> 1246	1.08	< 1088	0.94	< 1088	0.94	1157	1.00
ASP	> 7237	1.35	> 2211	1.65	1397	1.04	> 2254	1.68	1375	1.02	1400	1.04	1301	0.97
GLU	> 6537	1.06	< 1374	0.89	> 2105	1.36	> 1914	1.24	< 1144	0.74	< 1333	0.86	< 1255	0.81
LYS	> 5912	1.06	< 1285	0.92	> 1697	1.22	> 1557	1.12	1373	0.99	< 1282	0.92	> 1512	1.09
ASN	> 5341	1.30	> 1330	1.30	< 813	0.79	> 1871	1.82	> 1327	1.29	1003	0.98	1008	0.98
CYS	1373	0.91	376	1.00	< 295	0.78	< 298	0.79	404	1.07	> 426	1.13	< 306	0.81
GLN	> 3943	1.10	< 771	0.86	> 960	1.07	> 1168	1.31	> 1044	1.17	< 806	0.90	844	0.94
HIS	> 2634	1.17	> 679	1.21	< 493	0.88	> 797	1.42	> 665	1.18	579	1.03	540	0.96
SER	> 7024	1.25	> 1624	1.16	> 1920	1.37	> 2069	1.47	1411	1.00	< 1282	0.91	< 1300	0.93
THR	< 4046	0.79	< 1089	0.85	< 832	0.65	< 1117	0.87	< 1008	0.79	< 1109	0.87	1323	1.03
TRP	1457	1.05	> 386	1.11	342	0.99	374	1.08	355	1.02	> 398	1.15	> 387	1.11
TYR	< 3015	0.91	< 715	0.87	< 546	0.66	835	1.01	> 919	1.11	809	0.98	> 922	1.12
GLY	6476	1.01	< 1281	0.80	< 784	0.49	< 963	0.60	> 3448	2.16	< 1247	0.78	< 1478	0.93
OTH	277	0.83	72	0.86	< 63	0.75	71	0.85	71	0.85	> 115	1.37	> 136	1.62
ALL	92394	1.00	23099	1.00	23099	1.00	23098	1.00	23098	1.00	23100	1.00	23100	1.00

cluster 2: turn-type I'

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	P	P	f	P	f	P	
ALA	< 196	0.55	< 62	0.69	< 48	0.54	< 13	0.14	73	0.82	75	0.84	< 50	0.56
ILE	< 156	0.56	> 93	1.35	< 0	0.00	< 0	0.00	63	0.91	66	0.96	> 90	1.30
LEU	< 181	0.42	< 83	0.78	< 27	0.25	< 8	0.08	< 63	0.59	96	0.90	113	1.06
MET	58	0.63	< 13	0.57	< 12	0.52	< 6	0.26	27	1.18	23	1.00	27	1.18
PHE	155	0.77	> 78	1.56	< 32	0.64	< 19	0.38	< 26	0.52	54	1.08	47	0.94
PRO	< 28	0.13	< 28	0.53	< 0	0.00	< 0	0.00	< 0	0.00	< 29	0.54	> 118	2.21
VAL	< 245	0.74	> 131	1.58	< 7	0.08	< 7	0.08	100	1.20	95	1.14	> 107	1.29
ARG	227	0.96	56	0.95	53	0.90	< 15	0.25	> 103	1.74	51	0.86	46	0.78
ASP	> 382	1.39	56	0.82	> 216	3.14	72	1.05	< 38	0.55	65	0.95	58	0.84
GLU	< 228	0.72	< 57	0.72	< 60	0.76	< 19	0.24	92	1.16	81	1.02	< 62	0.78
LYS	> 372	1.31	> 88	1.24	70	0.98	< 23	0.32	> 191	2.68	71	1.00	74	1.04
ASN	> 503	2.39	56	1.07	> 299	5.69	> 110	2.09	< 38	0.72	66	1.25	< 36	0.69
CYS	50	0.65	21	1.09	< 10	0.52	< 7	0.36	12	0.62	23	1.19	> 31	1.60
GLN	187	1.02	40	0.87	43	0.94	< 20	0.44	> 84	1.83	35	0.76	34	0.74

HIS	123	1.07	> 44	1.53	36	1.25	< 9	0.31	34	1.18	28	0.97	21	0.73
SER	< 208	0.72	57	0.79	< 48	0.67	< 38	0.53	65	0.90	74	1.03	63	0.88
THR	< 134	0.51	60	0.92	< 2	0.03	< 8	0.12	64	0.98	> 100	1.53	66	1.01
TRP	49	0.69	23	1.29	< 5	0.28	< 9	0.51	12	0.68	20	1.12	> 34	1.91
TYR	172	1.02	> 86	2.03	< 25	0.59	< 20	0.47	41	0.97	> 62	1.46	50	1.18
GLY	> 1068	3.27	< 48	0.59	> 185	2.27	> 780	9.55	< 55	0.67	< 61	0.75	< 55	0.67
OTH	10	0.58	3	0.70	5	1.17	< 0	0.00	2	0.47	8	1.86	1	0.23
ALL	4732	1.00	1183	1.00	1183	1.00	1183	1.00	1183	1.00	1183	1.00	1183	1.00

cluster 3: turn-type II

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	849	0.89	< 198	0.83	> 312	1.31	< 58	0.24	> 281	1.18	< 195	0.82	< 175	0.73
ILE	< 388	0.53	199	1.08	< 89	0.48	< 0	0.00	< 100	0.54	> 215	1.17	166	0.90
LEU	< 660	0.58	> 319	1.12	< 124	0.44	< 34	0.12	< 183	0.64	> 371	1.30	< 193	0.68
MET	199	0.81	70	1.14	< 29	0.47	< 21	0.34	> 79	1.29	70	1.14	< 42	0.69
PHE	< 393	0.73	140	1.05	< 81	0.61	< 46	0.34	126	0.94	132	0.99	114	0.85
PRO	> 929	1.63	> 265	1.86	> 659	4.63	< 5	0.04	< 0	0.00	> 194	1.36	> 246	1.73
VAL	< 589	0.67	229	1.03	< 143	0.65	< 0	0.00	217	0.98	> 296	1.34	221	1.00
ARG	< 498	0.79	166	1.05	162	1.03	< 58	0.37	< 112	0.71	< 126	0.80	> 183	1.16
ASP	672	0.92	< 84	0.46	185	1.01	191	1.04	> 212	1.16	< 146	0.80	199	1.09
GLU	791	0.94	192	0.91	> 319	1.51	< 57	0.27	223	1.06	< 131	0.62	212	1.00
LYS	828	1.09	> 247	1.30	> 305	1.60	< 61	0.32	215	1.13	< 163	0.86	177	0.93
ASN	627	1.12	< 99	0.71	< 109	0.78	> 298	2.12	121	0.86	< 104	0.74	148	1.06
CYS	187	0.91	52	1.01	< 28	0.54	< 19	0.37	> 88	1.71	> 78	1.51	< 36	0.70
GLN	442	0.90	132	1.08	104	0.85	< 41	0.34	> 165	1.35	< 98	0.80	105	0.86
HIS	254	0.83	63	0.82	< 54	0.70	< 58	0.75	79	1.03	77	1.00	< 57	0.74
SER	< 656	0.85	< 151	0.79	< 154	0.80	< 70	0.37	> 281	1.46	203	1.06	> 222	1.16
THR	< 527	0.75	165	0.94	< 123	0.70	< 16	0.09	> 223	1.28	177	1.01	> 265	1.52
TRP	146	0.77	51	1.07	< 32	0.67	< 13	0.27	50	1.05	44	0.93	40	0.84
TYR	< 365	0.81	103	0.91	< 91	0.81	< 50	0.44	121	1.07	94	0.83	109	0.97
GLY	> 2603	2.99	221	1.01	< 49	0.23	> 2060	9.45	> 273	1.25	222	1.02	237	1.09
OTH	25	0.55	11	0.96	5	0.44	< 1	0.09	8	0.70	> 21	1.83	10	0.87
ALL	12628	1.00	3157	1.00	3157	1.00	3157	1.00	3157	1.00	3157	1.00	3157	1.00

cluster 4: turn-type II'

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	221	0.78	71	1.01	< 17	0.24	< 53	0.75	80	1.13	56	0.79	75	1.06
ILE	< 104	0.48	53	0.97	< 3	0.06	< 15	0.28	< 33	0.61	< 39	0.72	50	0.92
LEU	< 171	0.51	74	0.88	< 15	0.18	< 31	0.37	< 51	0.61	< 67	0.80	> 109	1.29
MET	47	0.65	11	0.61	< 3	0.17	12	0.66	21	1.16	20	1.10	19	1.05
PHE	122	0.77	> 60	1.52	< 3	0.08	< 20	0.51	39	0.99	34	0.86	43	1.09
PRO	< 53	0.32	< 21	0.50	< 0	0.00	32	0.76	< 0	0.00	39	0.93	42	1.00
VAL	< 129	0.49	63	0.96	< 4	0.06	< 19	0.29	< 43	0.66	75	1.14	> 84	1.28
ARG	160	0.86	54	1.16	< 16	0.34	42	0.90	48	1.03	> 70	1.50	38	0.81
ASP	> 323	1.49	43	0.79	45	0.83	> 178	3.28	57	1.05	< 39	0.72	< 21	0.39
GLU	243	0.97	60	0.96	< 18	0.29	> 93	1.49	72	1.15	< 39	0.62	< 43	0.69
LYS	180	0.80	< 41	0.73	< 28	0.50	52	0.92	59	1.05	69	1.23	48	0.85

ASN	> 217	1.31	32	0.77	< 24	0.58	> 104	2.51	> 57	1.37	39	0.94	< 27	0.65
CYS	51	0.84	22	1.44	< 6	0.39	11	0.72	12	0.79	12	0.79	15	0.98
GLN	121	0.84	39	1.08	< 13	0.36	30	0.83	39	1.08	43	1.19	38	1.05
HIS	83	0.91	> 32	1.41	< 8	0.35	18	0.79	25	1.10	22	0.97	< 10	0.44
SER	281	1.24	> 74	1.30	46	0.81	> 95	1.67	66	1.16	59	1.04	66	1.16
THR	169	0.82	54	1.04	< 4	0.08	43	0.83	> 68	1.32	56	1.08	65	1.26
TRP	45	0.80	13	0.93	7	0.50	12	0.85	13	0.93	13	0.93	19	1.35
TYR	120	0.90	> 61	1.83	< 7	0.21	< 16	0.48	36	1.08	41	1.23	42	1.26
GLY	> 879	3.41	51	0.79	> 666	10.33	53	0.82	> 109	1.69	> 97	1.50	78	1.21
OTH	17	1.25	5	1.48	1	0.30	5	1.48	6	1.77	5	1.48	2	0.59
ALL	3736	1.00	934	1.00	934	1.00	934	1.00	934	1.00	934	1.00	934	1.00

cluster 5: turn-type VIa1

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	27	0.96	9	1.28	11	1.57	< 0	0.00	7	1.00	8	1.14	4	0.57
ILE	4	0.18	2	0.37	1	0.18	< 0	0.00	1	0.18	3	0.55	1	0.18
LEU	18	0.54	8	0.95	5	0.60	< 0	0.00	5	0.60	5	0.60	6	0.72
MET	2	0.28	1	0.55	1	0.55	0	0.00	0	0.00	0	0.00	2	1.11
PHE	16	1.01	1	0.25	4	1.01	< 0	0.00	> 11	2.79	2	0.51	2	0.51
PRO	> 114	6.80	7	1.67	> 15	3.58	> 92	21.94	< 0	0.00	5	1.19	8	1.91
VAL	12	0.46	7	1.07	< 0	0.00	< 0	0.00	5	0.77	7	1.07	> 12	1.84
ARG	14	0.75	8	1.72	1	0.22	< 0	0.00	5	1.08	3	0.65	5	1.08
ASP	12	0.56	4	0.74	1	0.19	< 0	0.00	7	1.30	6	1.11	9	1.67
GLU	12	0.48	4	0.64	7	1.13	< 0	0.00	< 1	0.16	5	0.80	4	0.64
LYS	21	0.94	5	0.89	10	1.79	< 0	0.00	6	1.07	4	0.71	5	0.89
ASN	13	0.79	5	1.21	1	0.24	< 0	0.00	7	1.69	7	1.69	5	1.21
CYS	7	1.15	> 4	2.63	3	1.97	0	0.00	0	0.00	1	0.66	2	1.32
GLN	9	0.62	5	1.39	1	0.28	0	0.00	3	0.83	4	1.11	0	0.00
HIS	17	1.88	4	1.77	> 6	2.65	0	0.00	> 7	3.09	4	1.77	4	1.77
SER	12	0.53	3	0.53	5	0.88	< 0	0.00	4	0.71	8	1.42	6	1.06
THR	11	0.53	4	0.78	1	0.19	< 0	0.00	6	1.17	6	1.17	> 11	2.14
TRP	8	1.43	1	0.72	3	2.15	1	0.72	3	2.15	1	0.72	1	0.72
TYR	> 28	2.10	4	1.20	> 16	4.81	0	0.00	> 8	2.40	> 7	2.10	1	0.30
GLY	14	0.55	6	0.93	< 1	0.16	< 0	0.00	7	1.09	7	1.09	5	0.78
OTH	1	0.74	1	2.96	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	372	1.00	93	1.00	93	1.00	93	1.00	93	1.00	93	1.00	93	1.00

cluster 6: turn-type VIb

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	3	0.83	1	1.10	1	1.10	0	0.00	1	1.10	1	1.10	0	0.00
ILE	4	1.43	1	1.43	0	0.00	0	0.00	> 3	4.28	1	1.43	0	0.00
LEU	2	0.46	0	0.00	1	0.92	0	0.00	1	0.92	2	1.85	1	0.92
MET	1	1.07	1	4.29	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	4	1.97	> 3	5.90	0	0.00	0	0.00	1	1.97	0	0.00	0	0.00
PRO	> 10	4.62	0	0.00	0	0.00	> 10	18.49	0	0.00	0	0.00	1	1.85
VAL	2	0.59	1	1.19	0	0.00	0	0.00	1	1.19	0	0.00	1	1.19
ARG	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2	3.33	0	0.00

13.3.2 open β -turns

cluster 1: turn-type I

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P	f	P
ALA	> 3788	1.14	> 1043	1.26	> 965	1.17	882	1.07	> 898	1.08	> 977	1.18	< 633	0.76
ILE	< 2054	0.80	< 446	0.70	608	0.95	< 577	0.90	< 423	0.66	663	1.03	< 461	0.72
LEU	> 4397	1.11	< 898	0.91	> 1197	1.21	> 1449	1.47	< 853	0.86	> 1216	1.23	< 813	0.82
MET	920	1.08	211	0.99	> 243	1.14	> 246	1.16	220	1.03	219	1.03	187	0.88
PHE	1797	0.97	< 396	0.85	441	0.95	> 546	1.18	< 414	0.89	478	1.03	< 358	0.77
PRO	< 1051	0.53	454	0.92	< 202	0.41	< 38	0.08	< 357	0.72	< 342	0.69	> 1168	2.36
VAL	< 2161	0.70	< 415	0.54	< 639	0.83	< 600	0.78	< 507	0.66	< 681	0.89	< 539	0.70
ARG	> 2413	1.10	> 693	1.26	> 643	1.17	575	1.05	< 502	0.92	> 606	1.11	> 657	1.20
ASP	2538	1.00	> 789	1.24	668	1.05	< 549	0.86	< 532	0.84	> 824	1.29	625	0.98
GLU	> 3387	1.15	> 1040	1.42	> 1096	1.49	737	1.00	< 514	0.70	761	1.04	694	0.95
LYS	> 3214	1.22	> 851	1.29	> 918	1.39	> 754	1.14	691	1.05	623	0.94	> 880	1.33
ASN	> 2224	1.14	> 534	1.10	453	0.93	> 568	1.17	> 669	1.37	462	0.95	489	1.00
CYS	655	0.91	< 125	0.70	< 140	0.78	201	1.12	189	1.05	> 231	1.29	< 146	0.82
GLN	> 1991	1.17	> 504	1.19	> 542	1.27	460	1.08	> 485	1.14	433	1.02	> 489	1.15
HIS	1104	1.03	< 232	0.87	274	1.03	> 321	1.20	277	1.04	257	0.96	289	1.08
SER	2844	1.07	> 845	1.27	< 588	0.88	674	1.01	> 737	1.11	669	1.00	> 793	1.19
THR	2288	0.94	< 529	0.87	< 519	0.86	> 818	1.35	< 422	0.70	< 468	0.77	629	1.04
TRP	597	0.91	< 139	0.84	168	1.02	178	1.08	< 112	0.68	> 206	1.25	< 123	0.75
TYR	1443	0.92	< 291	0.74	< 342	0.87	> 495	1.26	< 315	0.80	386	0.98	< 288	0.73
GLY	2841	0.94	< 488	0.65	< 285	0.38	< 256	0.34	> 1812	2.39	< 419	0.55	< 601	0.79
OTH	157	0.99	43	1.08	35	0.88	42	1.06	37	0.93	46	1.16	> 105	2.64
ALL	43864	1.00	10966	1.00	10966	1.00	10966	1.00	10966	1.00	10967	1.00	10967	1.00

cluster 2: turn-type I'

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	43	0.89	> 20	1.66	< 4	0.33	8	0.66	11	0.91	10	0.83	< 4	0.33
ILE	15	0.40	6	0.64	< 1	0.11	< 0	0.00	8	0.86	9	0.96	8	0.86
LEU	< 29	0.50	15	1.04	< 4	0.28	< 5	0.35	< 5	0.35	17	1.18	13	0.90
MET	9	0.72	5	1.61	0	0.00	2	0.64	2	0.64	4	1.29	1	0.32
PHE	17	0.63	2	0.30	< 1	0.15	7	1.03	7	1.03	9	1.33	7	1.03
PRO	28	0.97	8	1.11	< 0	0.00	< 1	0.14	> 19	2.63	5	0.69	> 13	1.80
VAL	< 15	0.33	5	0.45	< 1	0.09	< 0	0.00	9	0.80	8	0.71	9	0.80
ARG	29	0.91	7	0.88	3	0.38	3	0.38	> 16	2.00	6	0.75	8	1.00
ASP	36	0.97	13	1.40	13	1.40	9	0.97	< 1	0.11	12	1.29	8	0.86
GLU	34	0.79	14	1.31	8	0.75	7	0.65	5	0.47	7	0.65	9	0.84
LYS	38	0.99	8	0.83	4	0.42	4	0.42	> 22	2.28	8	0.83	9	0.93
ASN	> 60	2.11	5	0.70	> 27	3.80	> 22	3.09	6	0.84	10	1.41	8	1.13
CYS	7	0.67	3	1.15	0	0.00	3	1.15	1	0.38	> 6	2.29	3	1.15
GLN	26	1.05	4	0.65	8	1.29	2	0.32	> 12	1.94	9	1.45	6	0.97
HIS	18	1.16	6	1.54	3	0.77	4	1.03	5	1.28	1	0.26	6	1.54
SER	36	0.93	11	1.13	6	0.62	9	0.93	10	1.03	11	1.13	10	1.03
THR	21	0.59	10	1.13	< 2	0.23	< 0	0.00	9	1.02	7	0.79	11	1.24
TRP	6	0.62	0	0.00	2	0.83	4	1.66	0	0.00	4	1.66	2	0.83
TYR	23	1.00	9	1.57	< 1	0.18	6	1.05	7	1.22	8	1.40	10	1.75

GLY	> 150	3.40	9	0.82	> 72	6.52	> 64	5.79	5	0.45	8	0.72	15	1.36
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	1.72	0	0.00
ALL	640	1.00	160	1.00	160	1.00	160	1.00	160	1.00	160	1.00	160	1.00

cluster 3: turn-type VIb

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	73	0.75	21	0.87	< 13	0.54	< 0	0.00	> 39	1.61	21	0.87	< 12	0.50
ILE	< 40	0.53	17	0.91	< 9	0.48	< 0	0.00	14	0.75	17	0.91	14	0.75
LEU	< 74	0.64	24	0.83	22	0.76	< 1	0.04	27	0.93	34	1.18	23	0.80
MET	20	0.80	9	1.44	5	0.80	< 0	0.00	6	0.96	5	0.80	5	0.80
PHE	59	1.08	13	0.96	> 21	1.54	< 0	0.00	> 25	1.84	13	0.96	20	1.47
PRO	> 384	6.63	> 25	1.73	12	0.83	> 308	21.28	> 39	2.70	> 25	1.73	> 30	2.07
VAL	64	0.71	20	0.89	19	0.84	< 1	0.04	24	1.07	26	1.15	23	1.02
ARG	35	0.55	12	0.75	11	0.69	< 1	0.06	11	0.69	15	0.94	10	0.62
ASP	< 36	0.48	12	0.64	17	0.91	< 0	0.00	< 7	0.38	< 10	0.54	20	1.07
GLU	< 45	0.52	17	0.79	20	0.93	< 1	0.05	< 7	0.33	13	0.61	13	0.61
LYS	53	0.69	13	0.67	21	1.09	< 1	0.05	18	0.93	14	0.72	19	0.98
ASN	58	1.02	13	0.91	> 30	2.10	< 2	0.14	13	0.91	20	1.40	20	1.40
CYS	17	0.81	6	1.14	6	1.14	< 0	0.00	5	0.95	3	0.57	7	1.33
GLN	45	0.90	10	0.80	> 20	1.61	< 1	0.08	14	1.13	8	0.64	13	1.05
HIS	26	0.83	11	1.41	4	0.51	< 0	0.00	11	1.41	8	1.02	9	1.15
SER	71	0.91	> 28	1.44	23	1.18	< 0	0.00	20	1.02	23	1.18	20	1.02
THR	55	0.77	21	1.18	17	0.96	< 3	0.17	14	0.79	17	0.96	> 30	1.69
TRP	16	0.83	1	0.21	> 13	2.69	< 0	0.00	2	0.41	3	0.62	4	0.83
TYR	59	1.28	12	1.05	> 27	2.35	< 2	0.17	18	1.57	16	1.39	14	1.22
GLY	< 53	0.60	> 36	1.62	< 10	0.45	< 0	0.00	< 7	0.32	28	1.26	15	0.68
OTH	1	0.22	0	0.00	1	0.86	0	0.00	0	0.00	2	1.72	0	0.00
ALL	1284	1.00	321	1.00	321	1.00	321	1.00	321	1.00	321	1.00	321	1.00

cluster 4: turn-type VIc

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	60	0.63	< 14	0.59	< 1	0.04	> 33	1.40	< 12	0.51	19	0.80	21	0.89
ILE	47	0.64	< 9	0.49	< 2	0.11	19	1.04	17	0.93	14	0.77	> 29	1.59
LEU	79	0.70	20	0.71	< 1	0.04	33	1.17	25	0.89	21	0.74	26	0.92
MET	12	0.49	4	0.66	< 0	0.00	4	0.66	4	0.66	> 11	1.81	7	1.15
PHE	50	0.94	19	1.43	< 0	0.00	20	1.51	11	0.83	18	1.36	15	1.13
PRO	> 380	6.73	13	0.92	> 294	20.84	> 36	2.55	> 37	2.62	> 25	1.77	> 23	1.63
VAL	76	0.87	17	0.77	< 1	0.05	30	1.37	28	1.27	20	0.91	> 33	1.50
ARG	< 29	0.46	10	0.64	< 0	0.00	< 6	0.38	13	0.83	13	0.83	21	1.34
ASP	< 38	0.52	18	0.99	< 0	0.00	< 3	0.17	17	0.94	< 9	0.50	< 10	0.55
GLU	< 34	0.41	15	0.72	< 2	0.10	< 8	0.38	< 9	0.43	18	0.86	19	0.91
LYS	47	0.62	15	0.80	< 2	0.11	12	0.64	18	0.96	< 10	0.53	11	0.58
ASN	57	1.02	> 29	2.09	< 1	0.07	9	0.65	18	1.29	15	1.08	17	1.22
CYS	17	0.83	4	0.78	< 0	0.00	7	1.37	6	1.17	5	0.98	9	1.76
GLN	43	0.89	16	1.32	< 2	0.17	15	1.24	10	0.82	11	0.91	9	0.74
HIS	20	0.66	3	0.39	< 0	0.00	10	1.31	7	0.92	6	0.79	7	0.92
SER	69	0.91	25	1.31	< 2	0.11	20	1.05	22	1.16	27	1.42	< 10	0.53

THR	66	0.95	17	0.98	< 4	0.23	17	0.98	> 28	1.62	> 30	1.73	12	0.69
TRP	13	0.69	8	1.70	< 0	0.00	2	0.43	3	0.64	1	0.21	4	0.85
TYR	60	1.34	> 20	1.79	< 1	0.09	> 24	2.14	15	1.34	13	1.16	14	1.25
GLY	55	0.64	> 37	1.71	< 0	0.00	< 5	0.23	13	0.60	24	1.11	15	0.69
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	3	2.64	1	0.88
ALL	1252	1.00	313	1.00	313	1.00	313	1.00	313	1.00	313	1.00	313	1.00

cluster 5: turn-type VIIla

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	< 4813	0.89	1352	0.99	< 1076	0.79	1412	1.04	< 973	0.72	< 1182	0.87	< 1072	0.79
ILE	< 3758	0.89	< 855	0.81	< 624	0.59	> 1207	1.15	1072	1.02	< 781	0.74	> 1180	1.12
LEU	< 5337	0.82	< 1195	0.74	< 1383	0.85	1649	1.02	< 1110	0.68	< 1462	0.90	< 1470	0.91
MET	< 1095	0.78	< 253	0.72	< 221	0.63	369	1.06	< 252	0.72	< 296	0.85	< 283	0.81
PHE	2869	0.94	< 592	0.78	< 559	0.73	> 1124	1.47	< 594	0.78	< 703	0.92	> 874	1.15
PRO	> 4692	1.45	> 1520	1.87	< 637	0.79	< 32	0.04	> 2503	3.08	> 1010	1.24	> 1203	1.48
VAL	< 4693	0.93	< 1079	0.85	< 855	0.68	> 1402	1.11	> 1357	1.07	< 913	0.72	> 1549	1.23
ARG	3490	0.97	< 795	0.88	> 985	1.09	857	0.95	853	0.95	> 983	1.09	< 799	0.89
ASP	> 5339	1.28	> 1180	1.13	> 1883	1.80	> 1168	1.12	> 1108	1.06	> 1302	1.25	1048	1.00
GLU	< 4472	0.93	1205	1.00	> 1335	1.11	< 1036	0.86	< 896	0.74	< 1133	0.94	< 1095	0.91
LYS	4574	1.05	1046	0.96	> 1418	1.31	< 908	0.84	> 1202	1.11	> 1162	1.07	< 926	0.85
ASN	> 3817	1.19	> 863	1.08	> 1300	1.63	> 941	1.18	< 713	0.89	> 995	1.24	777	0.97
CYS	1168	0.99	318	1.08	< 193	0.66	> 402	1.37	< 255	0.87	279	0.95	272	0.92
GLN	2707	0.97	< 614	0.88	> 809	1.16	683	0.98	< 601	0.86	682	0.98	< 571	0.82
HIS	1864	1.06	< 381	0.87	> 568	1.30	> 566	1.29	< 349	0.80	> 508	1.16	440	1.00
SER	> 5359	1.22	> 1478	1.35	> 1395	1.27	> 1311	1.20	> 1175	1.07	> 1280	1.17	1120	1.02
THR	> 4638	1.16	939	0.94	> 1320	1.32	> 1085	1.09	> 1294	1.30	1011	1.01	1018	1.02
TRP	1001	0.93	< 181	0.67	< 211	0.78	> 351	1.30	258	0.95	246	0.91	> 315	1.16
TYR	2409	0.94	< 465	0.72	< 564	0.88	> 914	1.42	< 466	0.72	627	0.97	> 743	1.15
GLY	< 3691	0.74	> 1647	1.33	< 614	0.49	< 502	0.40	< 928	0.75	> 1341	1.08	< 1137	0.92
OTH	216	0.83	< 45	0.69	52	0.80	80	1.22	< 39	0.60	> 111	1.70	> 115	1.76
ALL	72002	1.00	18003	1.00	18002	1.00	17999	1.00	17998	1.00	18007	1.00	18007	1.00

cluster 6: turn-type VIIIb

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	117	0.87	34	1.01	29	0.86	28	0.83	26	0.77	24	0.71	31	0.92
ILE	< 62	0.59	< 16	0.61	18	0.69	< 5	0.19	23	0.88	< 12	0.46	18	0.69
LEU	< 95	0.59	< 28	0.70	32	0.79	< 6	0.15	29	0.72	30	0.74	44	1.09
MET	45	1.30	13	1.50	13	1.50	6	0.69	13	1.50	6	0.69	7	0.81
PHE	83	1.10	> 29	1.53	19	1.00	13	0.69	22	1.16	12	0.63	20	1.06
PRO	66	0.82	> 34	1.69	15	0.74	< 1	0.05	16	0.79	28	1.39	23	1.14
VAL	< 69	0.55	28	0.89	22	0.70	< 3	0.10	< 16	0.51	26	0.83	> 43	1.37
ARG	83	0.93	17	0.76	31	1.39	< 9	0.40	26	1.16	19	0.85	16	0.72
ASP	113	1.09	< 16	0.62	> 36	1.39	32	1.23	29	1.12	28	1.08	23	0.89
GLU	81	0.68	21	0.70	29	0.97	< 12	0.40	< 19	0.64	31	1.04	26	0.87
LYS	103	0.96	26	0.97	> 41	1.52	< 14	0.52	22	0.82	31	1.15	19	0.71
ASN	82	1.03	19	0.96	> 31	1.56	13	0.65	19	0.96	26	1.31	14	0.71
CYS	28	0.96	7	0.96	5	0.68	11	1.51	5	0.68	6	0.82	> 15	2.05

GLN	46	0.66	< 9	0.52	20	1.15	< 7	0.40	10	0.58	21	1.21	< 9	0.52
HIS	40	0.92	5	0.46	11	1.01	11	1.01	13	1.19	12	1.10	17	1.56
SER	> 167	1.54	> 41	1.51	28	1.03	> 54	1.99	> 44	1.62	> 42	1.55	37	1.36
THR	123	1.24	< 15	0.61	> 42	1.70	29	1.17	> 37	1.50	27	1.09	32	1.29
TRP	27	1.00	11	1.64	3	0.45	2	0.30	11	1.64	5	0.74	7	1.04
TYR	56	0.88	18	1.13	10	0.63	10	0.63	18	1.13	17	1.06	20	1.25
GLY	> 301	2.44	> 60	1.94	< 12	0.39	> 181	5.87	> 48	1.56	> 43	1.39	< 20	0.65
OTH	1	0.15	0	0.00	0	0.00	0	0.00	1	0.62	1	0.62	> 6	3.70
ALL	1788	1.00	447	1.00	447	1.00	447	1.00	447	1.00	447	1.00	447	1.00

cluster 7: turn-type IX

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P	f	P
ALA	< 1734	0.90	> 697	1.44	504	1.04	< 136	0.28	< 397	0.82	501	1.04	< 334	0.69
ILE	< 851	0.57	< 176	0.47	< 153	0.41	< 7	0.02	> 515	1.38	< 309	0.83	382	1.02
LEU	< 1847	0.80	< 364	0.63	> 724	1.26	< 74	0.13	> 685	1.19	< 477	0.83	< 508	0.88
MET	< 395	0.80	< 95	0.77	137	1.10	< 44	0.36	119	0.96	< 97	0.78	< 93	0.75
PHE	< 736	0.68	< 132	0.49	< 234	0.86	< 70	0.26	300	1.11	< 168	0.62	< 239	0.88
PRO	< 459	0.40	300	1.04	< 55	0.19	< 0	0.00	< 104	0.36	< 233	0.81	> 568	1.97
VAL	< 973	0.54	< 234	0.52	< 175	0.39	< 18	0.04	> 546	1.22	< 323	0.72	416	0.93
ARG	1318	1.03	> 420	1.32	314	0.98	< 229	0.72	> 355	1.11	> 415	1.30	324	1.01
ASP	1630	1.10	> 452	1.22	> 628	1.69	< 295	0.80	< 255	0.69	> 615	1.66	> 419	1.13
GLU	1627	0.95	> 761	1.78	< 343	0.80	< 169	0.40	< 354	0.83	> 472	1.10	< 384	0.90
LYS	> 1818	1.18	> 598	1.55	< 336	0.87	364	0.95	> 520	1.35	> 548	1.42	422	1.10
ASN	> 1733	1.53	301	1.06	> 572	2.01	> 599	2.11	261	0.92	> 365	1.29	260	0.92
CYS	370	0.89	< 70	0.67	> 150	1.44	< 50	0.48	100	0.96	< 84	0.80	119	1.14
GLN	1016	1.03	> 332	1.34	260	1.05	< 199	0.80	225	0.91	> 280	1.13	< 213	0.86
HIS	642	1.03	152	0.98	> 215	1.38	142	0.91	133	0.85	155	1.00	149	0.96
SER	1467	0.94	> 453	1.17	> 523	1.35	< 152	0.39	< 339	0.87	416	1.07	> 439	1.13
THR	1304	0.92	< 274	0.78	> 604	1.71	< 30	0.09	> 396	1.12	372	1.05	> 441	1.25
TRP	< 223	0.58	< 50	0.52	< 72	0.75	< 13	0.14	88	0.92	85	0.89	100	1.04
TYR	< 697	0.76	< 143	0.63	245	1.07	< 46	0.20	> 263	1.15	< 177	0.77	239	1.05
GLY	> 4659	2.64	< 376	0.85	< 123	0.28	> 3741	8.48	419	0.95	< 269	0.61	< 322	0.73
OTH	61	0.66	< 10	0.43	23	0.99	< 12	0.52	16	0.69	29	1.25	19	0.82
ALL	25560	1.00	6390	1.00	6390	1.00	6390	1.00	6390	1.00	6390	1.00	6390	1.00

cluster 8: turn-type X

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P	f	P
ALA	589	1.00	156	1.06	< 77	0.53	> 234	1.60	< 122	0.83	139	0.95	159	1.09
ILE	< 268	0.59	< 48	0.42	< 82	0.72	101	0.89	< 37	0.33	< 82	0.72	102	0.90
LEU	636	0.91	< 94	0.54	> 262	1.50	188	1.07	< 92	0.53	< 135	0.77	> 224	1.28
MET	125	0.83	31	0.82	38	1.01	40	1.06	< 16	0.42	32	0.85	43	1.14
PHE	333	1.01	< 43	0.52	> 147	1.79	71	0.86	72	0.88	< 63	0.77	> 107	1.30
PRO	304	0.87	> 115	1.31	< 17	0.19	< 5	0.06	> 167	1.91	> 145	1.66	< 66	0.75
VAL	< 283	0.52	< 46	0.34	< 108	0.79	< 82	0.60	< 47	0.35	117	0.86	< 86	0.63
ARG	390	1.00	112	1.15	< 70	0.72	> 126	1.30	82	0.85	91	0.94	90	0.93
ASP	> 694	1.54	> 243	2.15	> 139	1.23	< 92	0.82	> 220	1.95	> 147	1.30	> 134	1.19
GLU	586	1.13	> 165	1.27	< 94	0.72	132	1.02	> 195	1.50	> 171	1.32	144	1.11

LYS	480	1.03	136	1.16	< 71	0.61	> 140	1.20	133	1.14	120	1.03	133	1.14
ASN	> 474	1.37	> 125	1.45	> 160	1.86	< 51	0.59	> 138	1.60	91	1.06	80	0.93
CYS	114	0.90	28	0.88	> 53	1.67	21	0.66	< 12	0.38	35	1.10	26	0.82
GLN	319	1.06	79	1.05	70	0.93	> 94	1.25	76	1.01	86	1.14	> 101	1.34
HIS	> 243	1.29	51	1.08	> 63	1.33	42	0.89	> 87	1.84	45	0.95	58	1.23
SER	> 573	1.21	> 164	1.39	99	0.84	> 197	1.67	113	0.96	131	1.11	< 85	0.72
THR	366	0.85	< 84	0.78	124	1.15	100	0.93	< 58	0.54	< 76	0.71	90	0.84
TRP	137	1.17	< 16	0.55	> 41	1.41	> 42	1.44	38	1.30	26	0.89	36	1.23
TYR	325	1.17	< 40	0.58	> 149	2.15	70	1.01	66	0.95	64	0.92	> 98	1.41
GLY	509	0.95	> 162	1.21	< 74	0.55	< 109	0.81	> 164	1.22	136	1.01	< 74	0.55
OTH	14	0.50	3	0.43	3	0.43	3	0.43	5	0.71	9	1.28	5	0.71
ALL	7762	1.00	1941	1.00	1941	1.00	1940	1.00	1940	1.00	1941	1.00	1941	1.00

cluster 9: turn-type XI

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P	f	P
ALA	< 282	0.78	89	0.98	< 59	0.65	< 40	0.44	94	1.03	< 65	0.72	< 60	0.66
ILE	278	0.99	> 112	1.59	69	0.98	< 5	0.07	> 92	1.31	> 96	1.37	< 41	0.58
LEU	378	0.87	> 138	1.27	< 74	0.68	< 28	0.26	> 138	1.27	< 89	0.82	< 56	0.52
MET	84	0.90	25	1.07	22	0.94	< 11	0.47	26	1.11	19	0.81	< 10	0.43
PHE	204	1.00	> 74	1.45	44	0.86	< 30	0.59	56	1.10	61	1.20	< 27	0.53
PRO	218	1.00	57	1.05	54	1.00	< 0	0.00	> 107	1.97	60	1.11	> 95	1.75
VAL	345	1.02	> 132	1.56	71	0.84	< 5	0.06	> 137	1.62	> 102	1.21	71	0.84
ARG	204	0.85	48	0.80	> 86	1.43	< 40	0.66	< 30	0.50	51	0.85	50	0.83
ASP	301	1.08	< 38	0.54	72	1.03	> 160	2.29	< 31	0.44	< 46	0.66	> 96	1.37
GLU	< 223	0.69	< 39	0.48	96	1.19	< 54	0.67	< 34	0.42	72	0.89	92	1.14
LYS	233	0.80	< 45	0.62	> 102	1.41	< 49	0.68	< 37	0.51	< 51	0.70	> 91	1.25
ASN	> 430	2.01	< 36	0.67	51	0.95	> 295	5.51	48	0.90	57	1.07	62	1.16
CYS	100	1.27	22	1.12	16	0.81	12	0.61	> 50	2.54	21	1.07	16	0.81
GLN	153	0.82	< 30	0.64	45	0.96	< 26	0.56	52	1.11	59	1.26	58	1.24
HIS	119	1.01	23	0.78	31	1.06	> 46	1.57	19	0.65	35	1.19	27	0.92
SER	289	0.99	58	0.79	> 107	1.46	< 51	0.70	73	1.00	70	0.96	> 102	1.39
THR	252	0.95	70	1.05	> 104	1.56	< 8	0.12	70	1.05	> 92	1.38	> 109	1.64
TRP	54	0.75	20	1.11	13	0.72	< 7	0.39	14	0.77	18	0.99	17	0.94
TYR	135	0.78	41	0.95	47	1.09	< 15	0.35	32	0.74	48	1.11	33	0.77
GLY	> 526	1.58	> 104	1.25	< 39	0.47	> 322	3.87	< 61	0.73	88	1.06	87	1.05
OTH	8	0.46	3	0.69	2	0.46	< 0	0.00	3	0.69	5	1.14	5	1.14
ALL	4816	1.00	1204	1.00	1204	1.00	1204	1.00	1204	1.00	1205	1.00	1205	1.00

cluster 10: turn-type XII

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	327	0.84	88	0.90	< 43	0.44	110	1.13	86	0.88	104	1.07	91	0.93
ILE	< 217	0.72	< 43	0.57	< 2	0.03	73	0.97	> 99	1.31	64	0.85	< 50	0.66
LEU	391	0.84	< 92	0.79	< 22	0.19	> 152	1.31	125	1.07	108	0.93	< 83	0.71
MET	87	0.87	20	0.80	< 14	0.56	26	1.04	27	1.08	25	1.00	< 15	0.60
PHE	190	0.87	< 27	0.49	< 27	0.49	57	1.04	> 79	1.44	42	0.77	55	1.00
PRO	< 125	0.54	> 81	1.39	< 0	0.00	< 31	0.53	< 13	0.22	69	1.19	72	1.24

VAL	< 210 0.58	< 40 0.44	< 6 0.07	84 0.93	80 0.88	< 59 0.65	< 63 0.70
ARG	219 0.85	77 1.19	< 43 0.67	59 0.91	< 40 0.62	64 0.99	67 1.04
ASP	321 1.07	> 133 1.77	72 0.96	< 44 0.59	72 0.96	> 92 1.23	> 97 1.29
GLU	298 0.86	85 0.98	< 44 0.51	81 0.94	88 1.02	97 1.12	102 1.18
LYS	287 0.92	78 1.00	< 54 0.69	> 118 1.52	< 37 0.48	> 108 1.39	> 111 1.43
ASN	> 305 1.33	> 96 1.67	> 128 2.23	< 34 0.59	47 0.82	47 0.82	58 1.01
CYS	88 1.04	20 0.95	18 0.85	25 1.19	25 1.19	21 1.00	< 11 0.52
GLN	188 0.94	52 1.04	< 32 0.64	59 1.18	45 0.90	57 1.14	60 1.20
HIS	137 1.09	35 1.11	41 1.30	25 0.80	36 1.15	29 0.92	34 1.08
SER	319 1.02	> 101 1.29	< 34 0.43	> 107 1.36	77 0.98	< 58 0.74	91 1.16
THR	255 0.89	> 96 1.34	< 6 0.08	82 1.15	71 0.99	61 0.85	87 1.22
TRP	57 0.73	14 0.72	< 4 0.21	20 1.03	19 0.98	15 0.77	23 1.19
TYR	167 0.90	< 30 0.65	< 23 0.50	48 1.04	> 66 1.43	41 0.89	36 0.78
GLY	> 961 2.70	78 0.88	> 677 7.60	< 54 0.61	> 152 1.71	> 124 1.39	78 0.88
OTH	15 0.80	5 1.07	1 0.21	2 0.43	7 1.49	6 1.28	7 1.49
ALL	5164 1.00	1291 1.00	1291 1.00	1291 1.00	1291 1.00	1291 1.00	1291 1.00

cluster 11: turn-type XIII

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	91	0.71	33	1.03	< 9	0.28	< 15	0.47	34	1.06	30	0.94	34	1.06
ILE	< 48	0.49	< 6	0.24	< 0	0.00	17	0.69	25	1.01	20	0.81	23	0.93
LEU	< 101	0.66	< 24	0.63	< 3	0.08	36	0.94	38	1.00	34	0.89	29	0.76
MET	25	0.76	6	0.73	3	0.37	9	1.10	7	0.85	3	0.37	6	0.73
PHE	42	0.59	12	0.67	< 9	0.50	< 6	0.34	15	0.84	16	0.89	11	0.61
PRO	< 33	0.43	> 33	1.73	< 0	0.00	< 0	0.00	< 0	0.00	19	1.00	> 35	1.84
VAL	82	0.69	< 4	0.14	< 0	0.00	> 43	1.45	35	1.18	32	1.08	29	0.98
ARG	92	1.09	> 33	1.56	14	0.66	> 33	1.56	< 12	0.57	22	1.04	25	1.18
ASP	121	1.23	> 61	2.48	18	0.73	20	0.81	22	0.90	29	1.18	> 38	1.55
GLU	95	0.84	30	1.06	< 7	0.25	< 17	0.60	> 41	1.45	28	0.99	24	0.85
LYS	108	1.06	21	0.82	25	0.98	> 46	1.81	16	0.63	34	1.33	30	1.18
ASN	> 126	1.68	> 31	1.65	> 44	2.34	> 32	1.70	19	1.01	19	1.01	21	1.12
CYS	35	1.27	8	1.16	4	0.58	8	1.16	> 15	2.17	7	1.01	9	1.30
GLN	67	1.02	16	0.98	14	0.85	20	1.22	17	1.04	23	1.40	11	0.67
HIS	41	1.00	9	0.87	11	1.07	11	1.07	10	0.97	10	0.97	6	0.58
SER	96	0.93	25	0.97	< 15	0.58	30	1.17	26	1.01	22	0.86	30	1.17
THR	109	1.16	> 33	1.41	< 0	0.00	> 51	2.18	25	1.07	24	1.02	< 13	0.56
TRP	13	0.51	4	0.63	< 1	0.16	6	0.94	2	0.31	> 13	2.04	> 17	2.67
TYR	38	0.63	14	0.93	< 3	0.20	10	0.66	11	0.73	10	0.66	11	0.73
GLY	> 327	2.80	19	0.65	> 243	8.32	< 12	0.41	> 53	1.82	28	0.96	22	0.75
OTH	2	0.33	1	0.65	0	0.00	1	0.65	0	0.00	1	0.65	0	0.00
ALL	1692	1.00	423	1.00	423	1.00	423	1.00	423	1.00	424	1.00	424	1.00

cluster 12: turn-type VIIa

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	< 1739	0.68	618	0.97	< 143	0.23	607	0.96	< 371	0.58	> 749	1.18	< 507	0.80
ILE	< 1360	0.69	< 194	0.40	< 10	0.02	> 654	1.33	502	1.02	< 389	0.79	> 751	1.53
LEU	< 2091	0.69	< 672	0.89	< 89	0.12	777	1.02	< 553	0.73	< 607	0.80	> 851	1.12

cluster 14: turn-type VII2

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	< 1284	0.77	< 292	0.70	< 362	0.87	< 235	0.56	395	0.95	408	0.98	381	0.91
ILE	1337	1.03	< 258	0.80	> 570	1.76	> 370	1.15	< 139	0.43	304	0.94	> 443	1.37
LEU	< 1651	0.83	< 357	0.72	509	1.02	> 575	1.15	< 210	0.42	< 414	0.83	< 438	0.88
MET	< 300	0.70	< 66	0.61	< 72	0.67	93	0.87	< 69	0.64	97	0.90	104	0.97
PHE	831	0.89	< 193	0.82	> 274	1.17	< 151	0.64	213	0.91	> 264	1.13	224	0.96
PRO	> 1240	1.24	244	0.98	> 370	1.48	> 609	2.44	< 17	0.07	> 294	1.18	> 388	1.56
VAL	> 1738	1.12	405	1.04	> 653	1.68	> 467	1.20	< 213	0.55	406	1.05	> 492	1.27
ARG	1057	0.96	> 336	1.22	< 172	0.62	282	1.02	267	0.97	< 223	0.81	< 208	0.75
ASP	> 1627	1.27	293	0.91	> 498	1.55	315	0.98	> 521	1.62	< 266	0.83	< 285	0.89
GLU	< 1312	0.89	> 417	1.13	< 200	0.54	< 268	0.72	> 427	1.15	< 304	0.82	336	0.91
LYS	1391	1.04	> 472	1.42	< 156	0.47	> 419	1.26	344	1.03	301	0.90	317	0.95
ASN	> 1173	1.19	> 277	1.13	> 346	1.41	< 190	0.77	> 360	1.46	249	1.01	< 194	0.79
CYS	412	1.14	< 55	0.61	> 203	2.25	< 62	0.69	92	1.02	101	1.12	92	1.02
GLN	824	0.96	> 246	1.15	< 159	0.74	200	0.93	219	1.02	< 179	0.84	201	0.94
HIS	575	1.07	150	1.11	131	0.97	< 109	0.81	> 185	1.37	113	0.84	130	0.97
SER	1421	1.06	< 292	0.87	< 267	0.79	325	0.97	> 537	1.60	331	0.98	323	0.96
THR	> 1426	1.16	> 382	1.25	< 184	0.60	> 528	1.72	332	1.08	280	0.91	311	1.01
TRP	307	0.92	< 64	0.77	71	0.85	74	0.89	98	1.18	> 129	1.55	88	1.06
TYR	723	0.91	192	0.97	222	1.12	< 111	0.56	198	1.00	213	1.08	186	0.94
GLY	1438	0.94	> 529	1.39	< 96	0.25	< 136	0.36	> 677	1.77	> 627	1.64	356	0.93
OTH	61	0.76	12	0.60	17	0.85	13	0.65	19	0.95	> 29	1.44	> 35	1.74
ALL	22128	1.00	5532	1.00	5532	1.00	5532	1.00	5532	1.00	5532	1.00	5532	1.00

cluster 15: turn-type VII3

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	1119	0.94	< 251	0.84	> 349	1.17	< 256	0.86	< 263	0.88	281	0.95	265	0.89
ILE	< 589	0.64	< 200	0.87	< 164	0.71	< 112	0.49	< 113	0.49	> 261	1.14	> 275	1.20
LEU	< 942	0.66	< 285	0.80	< 226	0.64	< 238	0.67	< 193	0.54	334	0.94	< 308	0.87
MET	< 206	0.67	65	0.85	< 50	0.65	< 49	0.64	< 42	0.55	67	0.88	64	0.84
PHE	697	1.04	170	1.02	> 229	1.37	< 91	0.55	> 207	1.24	> 211	1.26	173	1.04
PRO	> 958	1.35	> 213	1.20	> 358	2.02	> 382	2.15	< 5	0.03	> 217	1.22	> 408	2.30
VAL	< 839	0.76	245	0.89	264	0.96	< 136	0.49	< 194	0.70	> 327	1.18	> 370	1.34
ARG	747	0.95	207	1.05	< 137	0.70	198	1.01	205	1.04	187	0.95	< 146	0.74
ASP	> 1288	1.41	< 186	0.81	< 181	0.79	> 565	2.47	> 356	1.56	< 171	0.75	203	0.89
GLU	< 883	0.84	258	0.98	< 102	0.39	264	1.00	259	0.98	< 213	0.81	< 185	0.70
LYS	871	0.92	> 279	1.18	< 98	0.41	> 300	1.26	< 194	0.82	245	1.03	< 188	0.79
ASN	> 816	1.17	166	0.95	158	0.90	> 242	1.38	> 250	1.43	165	0.94	160	0.91
CYS	274	1.06	74	1.15	> 113	1.76	< 30	0.47	57	0.89	67	1.04	55	0.85
GLN	518	0.85	148	0.97	< 89	0.58	134	0.88	147	0.96	< 111	0.73	< 94	0.62
HIS	409	1.07	104	1.08	103	1.07	77	0.80	> 125	1.30	91	0.95	83	0.87
SER	> 1490	1.56	234	0.98	> 538	2.25	> 327	1.37	> 391	1.63	225	0.94	248	1.04
THR	> 1176	1.35	226	1.04	> 420	1.93	> 275	1.26	> 255	1.17	190	0.87	238	1.09
TRP	211	0.89	65	1.10	< 44	0.74	< 40	0.68	62	1.05	58	0.98	49	0.83
TYR	586	1.04	153	1.09	> 189	1.34	< 81	0.57	163	1.16	> 171	1.21	156	1.11
GLY	1090	1.00	> 397	1.46	< 116	0.43	< 130	0.48	> 447	1.64	> 319	1.17	253	0.93
OTH	40	0.70	12	0.84	9	0.63	10	0.70	9	0.63	> 27	1.89	17	1.19

ALL	15749	1.00	3938	1.00	3937	1.00	3937	1.00	3937	1.00	3938	1.00	3938	1.00
-----	-------	------	------	------	------	------	------	------	------	------	------	------	------	------

cluster 16: turn-type VII4a

aa	turn		position 1		position 2		position 3		position 4		before		after							
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P						
ALA	<	138	0.73	47	0.99	38	0.80	<	10	0.21	43	0.91	49	1.03	39	0.82				
ILE		119	0.81	30	0.82	36	0.98	<	1	0.03	>	52	1.42	26	0.71	>	49	1.33		
LEU		175	0.77	61	1.08	>	74	1.31	<	6	0.11	<	34	0.60	48	0.85	<	33	0.58	
MET		42	0.86	14	1.15	14	1.15	<	2	0.16	12	0.98	12	0.98			7	0.57		
PHE		99	0.93	30	1.13	34	1.27	<	5	0.19	30	1.13	25	0.94			27	1.01		
PRO		135	1.19	>	48	1.69	22	0.78	<	1	0.04	>	64	2.26	31	1.09			33	1.16
VAL	<	118	0.67	39	0.88	36	0.82	<	1	0.02	42	0.95	35	0.79			45	1.02		
ARG	<	76	0.60	24	0.76	29	0.92	<	1	0.03	22	0.70	35	1.11			27	0.86		
ASP		101	0.69	34	0.93	33	0.90	<	6	0.16	28	0.77	38	1.04			>	58	1.59	
GLU	<	115	0.68	38	0.90	39	0.93	<	3	0.07	35	0.83	45	1.07			45	1.07		
LYS		112	0.74	36	0.95	33	0.87	<	6	0.16	37	0.98	40	1.06			28	0.74		
ASN		93	0.83	25	0.89	>	40	1.43	<	5	0.18	23	0.82	27	0.97			25	0.89	
CYS		32	0.78	7	0.68	13	1.26	<	3	0.29	9	0.88	9	0.88			10	0.97		
GLN		66	0.68	22	0.90	26	1.07	<	4	0.16	<	14	0.57	26	1.07			24	0.98	
HIS		44	0.72	10	0.65	19	1.24	<	2	0.13	13	0.85	15	0.98			13	0.85		
SER		141	0.92	>	51	1.33	40	1.05	<	9	0.24	41	1.07	46	1.20			45	1.18	
THR		129	0.93	32	0.92	31	0.89	<	6	0.17	>	60	1.72	>	47	1.35			43	1.24
TRP		35	0.93	13	1.37	12	1.27	<	1	0.11	9	0.95	4	0.42			7	0.74		
TYR		79	0.88	25	1.11	31	1.38	<	4	0.18	19	0.84	18	0.80			29	1.29		
GLY	>	662	3.81	40	0.92	<	29	0.67	>	553	12.73	40	0.92	45	1.04			41	0.94	
OTH		5	0.55	3	1.31	0	0.00	0	0.00	2	0.88	>	8	3.50			1	0.44		
ALL		2516	1.00	629	1.00	629	1.00	629	1.00	629	1.00	629	1.00	629	1.00			629	1.00	

cluster 17: turn-type VII4b

aa	turn		position 1		position 2		position 3		position 4		before		after						
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P					
ALA	<	90	0.61	29	0.78	30	0.81	<	4	0.11	27	0.73	29	0.78	33	0.89			
ILE	<	68	0.59	<	15	0.52	20	0.70	<	0	0.00	33	1.15	25	0.87	34	1.18		
LEU	<	106	0.60	<	13	0.29	>	63	1.42	<	1	0.02	<	29	0.65	33	0.74	36	0.81
MET		34	0.89	6	0.63	13	1.36	<	1	0.11	14	1.47	9	0.94	13	1.36			
PHE		86	1.03	16	0.77	>	40	1.92	<	0	0.00	>	30	1.44	21	1.01	19	0.91	
PRO		79	0.89	>	33	1.49	19	0.86	<	0	0.00	27	1.22	>	33	1.49	>	33	1.49
VAL	<	66	0.48	<	14	0.41	27	0.78	<	0	0.00	25	0.72	<	23	0.67	>	46	1.33
ARG		75	0.76	24	0.98	28	1.14	<	4	0.16	19	0.77	26	1.06			31	1.26	
ASP		105	0.92	>	52	1.82	<	18	0.63	<	2	0.07	33	1.16	32	1.12	28	0.98	
GLU		109	0.83	42	1.28	29	0.88	<	3	0.09	35	1.06	39	1.18	28	0.85			
LYS		93	0.78	34	1.15	31	1.05	<	7	0.24	21	0.71	32	1.08	31	1.05			
ASN		86	0.98	>	32	1.46	20	0.92	<	1	0.05	>	33	1.51	24	1.10	21	0.96	
CYS		20	0.62	5	0.62	8	1.00	<	0	0.00	7	0.87	7	0.87	9	1.12			
GLN		50	0.66	20	1.05	20	1.05	<	1	0.05	<	9	0.47	15	0.79	13	0.68		
HIS		41	0.86	8	0.67	16	1.34	<	2	0.17	15	1.25	9	0.75	11	0.92			
SER		104	0.87	40	1.34	28	0.94	<	2	0.07	34	1.14	33	1.10	21	0.70			
THR		84	0.77	31	1.14	29	1.07	<	0	0.00	24	0.88	>	41	1.51	33	1.21		
TRP		23	0.78	6	0.81	8	1.08	<	1	0.14	8	1.08	8	1.08	7	0.95			
TYR		59	0.84	15	0.85	22	1.25	<	0	0.00	22	1.25	16	0.91	10	0.57			

13.3.3 reverse β -turns

cluster 1: turn-type I'

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	P	f	P	f	P	f	P
ALA	<	90 0.41	<	29 0.53	<	23 0.42	<	5 0.09	<	33 0.61	<	30 0.55	<	26 0.48
ILE		131 0.78	>	87 2.07	<	0 0.00	<	0 0.00		44 1.05	>	55 1.31	>	70 1.66
LEU	<	89 0.34	<	49 0.75	<	9 0.14	<	2 0.03	<	29 0.45		51 0.78	>	82 1.26
MET		30 0.54	<	6 0.43	<	6 0.43	<	2 0.14		16 1.14		16 1.14		19 1.36
PHE		80 0.65	>	49 1.60	<	14 0.46	<	5 0.16	<	12 0.39		39 1.28		29 0.95
PRO	<	0 0.00	<	0 0.00	<	0 0.00	<	0 0.00	<	0 0.00	<	11 0.34	>	87 2.68
VAL		201 0.99	>	120 2.37	<	2 0.04	<	2 0.04	>	77 1.52	>	70 1.38	>	72 1.42
ARG		130 0.90		37 1.03		28 0.78	<	7 0.19	>	58 1.61		31 0.86		28 0.78
ASP	>	238 1.42	<	21 0.50	>	159 3.80		47 1.12	<	11 0.26	<	27 0.65	<	25 0.60
GLU		186 0.96		53 1.10		47 0.97	<	13 0.27	>	73 1.51	>	63 1.31		43 0.89
LYS	>	290 1.67	>	67 1.54		42 0.97	<	18 0.41	>	163 3.75		48 1.11		47 1.08
ASN	>	319 2.49	<	17 0.53	>	222 6.93	>	59 1.84	<	21 0.66		34 1.06	<	20 0.62
CYS		22 0.47		11 0.93	<	4 0.34	<	1 0.09		6 0.51		14 1.19	>	19 1.61
GLN		127 1.14		28 1.00		31 1.11	<	9 0.32	>	59 2.11		21 0.75	<	13 0.47
HIS		76 1.08	>	27 1.54		20 1.14	<	3 0.17	>	26 1.48		11 0.63	<	9 0.51
SER	<	96 0.55	<	29 0.66	<	19 0.43	<	22 0.50	<	26 0.59		49 1.12	<	28 0.64
THR	<	54 0.34	<	14 0.35	<	0 0.00	<	4 0.10		36 0.90	>	68 1.70		38 0.95
TRP	<	15 0.35		10 0.92	<	1 0.09	<	0 0.00	<	4 0.37		12 1.11	>	27 2.49
TYR		97 0.94	>	63 2.44	<	10 0.39	<	3 0.12		21 0.81	>	54 2.09		28 1.09
GLY	>	611 3.07	<	4 0.08	>	82 1.65	>	519 10.43	<	6 0.12	<	14 0.28	<	11 0.22
OTH		2 0.19		0 0.00		2 0.76		0 0.00		0 0.00		3 1.15		0 0.00
ALL		2884 1.00		721 1.00		721 1.00		721 1.00		721 1.00		721 1.00		721 1.00

cluster 2: turn-type II'

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	<	54 0.50		22 0.81	<	2 0.07	<	14 0.52	<	16 0.59	<	11 0.41	<	15 0.55
ILE		50 0.59		27 1.28	<	2 0.10	<	5 0.24		16 0.76		19 0.90		27 1.28
LEU	<	56 0.43		24 0.74	<	3 0.09	<	8 0.25	<	21 0.65	<	19 0.59	>	67 2.06
MET		11 0.39		3 0.43		2 0.29	<	0 0.00		6 0.86		11 1.57		9 1.29
PHE		39 0.64		19 1.25	<	1 0.07	<	3 0.20		16 1.05		11 0.72		17 1.11
PRO	<	10 0.15	<	0 0.00	<	0 0.00		10 0.62	<	0 0.00	<	7 0.43	<	8 0.49
VAL		78 0.77	>	41 1.62	<	5 0.20	<	2 0.08		30 1.19	>	38 1.50	>	46 1.82
ARG		70 0.97		24 1.33	<	5 0.28		20 1.11		21 1.17	>	39 2.17		10 0.56
ASP	>	172 2.06	>	31 1.48		27 1.29	>	94 4.50		20 0.96	<	5 0.24	<	6 0.29
GLU		123 1.28	>	35 1.45	<	10 0.42	>	41 1.70	>	37 1.54		16 0.66		22 0.91
LYS		97 1.12		24 1.11	<	12 0.55		23 1.06	>	38 1.75	>	45 2.08		15 0.69
ASN	>	102 1.59		18 1.13		15 0.94	>	46 2.88		23 1.44		11 0.69	<	7 0.44
CYS		16 0.68		8 1.36	<	1 0.17		2 0.34		5 0.85		8 1.36		7 1.19
GLN		57 1.02		21 1.51	<	5 0.36		12 0.86		19 1.36	>	22 1.58		13 0.93
HIS		26 0.74		12 1.37	<	1 0.11	<	3 0.34		10 1.14		10 1.14	<	3 0.34
SER		75 0.86		15 0.69	<	6 0.27	>	32 1.46		22 1.00		20 0.91		17 0.78
THR		68 0.85	<	9 0.45	<	0 0.00		22 1.10	>	37 1.86		28 1.41		21 1.05

HIS	1	0.93	0	0.00	0	0.00	1	3.73	0	0.00	0	0.00	0	0.00
SER	6	2.24	2	2.99	> 3	4.49	1	1.50	0	0.00	2	2.99	> 3	4.49
THR	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TRP	1	1.51	0	0.00	0	0.00	> 1	6.05	0	0.00	0	0.00	> 1	6.05
TYR	1	0.64	0	0.00	0	0.00	1	2.54	0	0.00	0	0.00	0	0.00
GLY	> 17	5.60	> 5	6.58	1	1.32	0	0.00	> 11	14.48	1	1.32	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	44	1.00	11	1.00	11	1.00	11	1.00	11	1.00	11	1.00	11	1.00

cluster 5: turn-type IVa

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	15	0.78	< 0	0.00	> 9	1.86	6	1.24	< 0	0.00	4	0.83	7	1.45
ILE	3	0.20	< 0	0.00	2	0.54	1	0.27	< 0	0.00	6	1.61	> 8	2.14
LEU	6	0.26	< 1	0.17	3	0.52	2	0.35	< 0	0.00	3	0.52	10	1.73
MET	2	0.40	0	0.00	0	0.00	2	1.61	0	0.00	2	1.61	2	1.61
PHE	4	0.37	1	0.37	0	0.00	3	1.11	0	0.00	2	0.74	2	0.74
PRO	3	0.26	0	0.00	3	1.04	0	0.00	0	0.00	1	0.35	4	1.39
VAL	4	0.22	< 0	0.00	4	0.89	< 0	0.00	< 0	0.00	6	1.34	6	1.34
ARG	6	0.47	0	0.00	1	0.31	5	1.56	0	0.00	4	1.25	2	0.63
ASP	29	1.95	> 16	4.30	> 9	2.42	4	1.08	< 0	0.00	1	0.27	2	0.54
GLU	17	0.99	< 0	0.00	8	1.87	> 9	2.10	< 0	0.00	< 0	0.00	5	1.17
LYS	18	1.17	1	0.26	7	1.82	> 10	2.59	< 0	0.00	> 9	2.33	2	0.52
ASN	17	1.49	> 8	2.81	2	0.70	> 7	2.46	0	0.00	0	0.00	4	1.41
CYS	1	0.24	0	0.00	1	0.96	0	0.00	0	0.00	1	0.96	1	0.96
GLN	6	0.61	0	0.00	4	1.61	2	0.81	0	0.00	> 6	2.42	0	0.00
HIS	3	0.48	1	0.64	0	0.00	2	1.28	0	0.00	2	1.28	1	0.64
SER	23	1.48	> 8	2.06	> 9	2.31	6	1.54	< 0	0.00	7	1.80	2	0.51
THR	13	0.92	> 10	2.82	1	0.28	2	0.56	0	0.00	5	1.41	2	0.56
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	3	0.33	1	0.44	0	0.00	2	0.87	0	0.00	3	1.31	3	1.31
GLY	> 82	4.64	> 17	3.85	1	0.23	1	0.23	> 63	14.26	2	0.45	1	0.23
OTH	1	1.08	0	0.00	0	0.00	0	0.00	1	4.31	0	0.00	0	0.00
ALL	256	1.00	64	1.00	64	1.00	64	1.00	64	1.00	64	1.00	64	1.00

cluster 6: turn-type IVb

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	7	0.43	1	0.25	3	0.74	3	0.74	< 0	0.00	3	0.74	1	0.25
ILE	2	0.16	0	0.00	1	0.32	1	0.32	0	0.00	3	0.95	5	1.59
LEU	7	0.36	2	0.41	< 0	0.00	5	1.03	< 0	0.00	5	1.03	8	1.64
MET	4	0.95	2	1.91	1	0.95	1	0.95	0	0.00	1	0.95	1	0.95
PHE	7	0.77	4	1.75	1	0.44	2	0.87	0	0.00	2	0.87	2	0.87
PRO	14	1.44	0	0.00	> 14	5.75	0	0.00	0	0.00	1	0.41	> 7	2.88
VAL	4	0.26	1	0.26	1	0.26	2	0.53	< 0	0.00	6	1.58	5	1.32
ARG	7	0.65	0	0.00	2	0.74	4	1.48	1	0.37	4	1.48	1	0.37
ASP	12	0.96	2	0.64	4	1.27	6	1.91	0	0.00	5	1.59	0	0.00
GLU	17	1.18	< 0	0.00	> 11	3.04	6	1.66	< 0	0.00	6	1.66	2	0.55
LYS	8	0.62	2	0.62	4	1.23	2	0.62	0	0.00	5	1.54	1	0.31

ASN	6	0.63	2	0.83	1	0.42	3	1.25	0	0.00	1	0.42	2	0.83
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2	2.27	0	0.00
GLN	9	1.08	1	0.48	3	1.43	> 5	2.39	0	0.00	4	1.91	1	0.48
HIS	6	1.14	> 4	3.04	0	0.00	2	1.52	0	0.00	1	0.76	0	0.00
SER	14	1.07	5	1.52	6	1.83	3	0.91	0	0.00	2	0.61	4	1.22
THR	> 27	2.26	> 22	7.36	2	0.67	3	1.00	0	0.00	1	0.33	> 7	2.34
TRP	2	0.62	1	1.23	0	0.00	1	1.23	0	0.00	0	0.00	2	2.46
TYR	4	0.52	1	0.52	0	0.00	3	1.55	0	0.00	1	0.52	2	1.03
GLY	> 59	3.96	4	1.07	< 0	0.00	2	0.54	> 53	14.22	1	0.27	3	0.81
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	216	1.00	54	1.00	54	1.00	54	1.00	54	1.00	54	1.00	54	1.00

cluster 7: turn-type Va

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	52	0.87	8	0.54	19	1.28	10	0.67	15	1.01	9	0.61	< 3	0.20
ILE	< 12	0.26	< 0	0.00	< 4	0.35	< 2	0.17	6	0.52	6	0.52	13	1.13
LEU	< 28	0.39	10	0.56	< 2	0.11	< 7	0.39	< 9	0.51	21	1.18	15	0.84
MET	9	0.59	3	0.78	2	0.52	2	0.52	2	0.52	3	0.78	3	0.78
PHE	25	0.75	8	0.96	4	0.48	5	0.60	8	0.96	9	1.08	5	0.60
PRO	> 60	1.69	< 0	0.00	> 60	6.76	< 0	0.00	< 0	0.00	9	1.01	> 32	3.60
VAL	< 21	0.38	< 2	0.14	< 6	0.43	< 4	0.29	9	0.65	19	1.37	16	1.16
ARG	31	0.79	6	0.61	6	0.61	10	1.01	9	0.91	6	0.61	7	0.71
ASP	69	1.51	9	0.79	13	1.14	> 33	2.88	14	1.22	7	0.61	8	0.70
GLU	69	1.31	7	0.53	> 24	1.82	> 27	2.05	11	0.83	< 5	0.38	11	0.83
LYS	43	0.91	< 2	0.17	13	1.10	14	1.18	14	1.18	13	1.10	14	1.18
ASN	> 62	1.77	14	1.60	6	0.69	> 17	1.94	> 25	2.86	11	1.26	9	1.03
CYS	7	0.54	4	1.24	0	0.00	2	0.62	1	0.31	3	0.93	4	1.24
GLN	31	1.01	8	1.05	5	0.66	9	1.18	9	1.18	10	1.31	7	0.92
HIS	31	1.62	7	1.46	3	0.63	9	1.88	> 12	2.50	> 13	2.71	5	1.04
SER	> 89	1.86	> 42	3.51	9	0.75	17	1.42	> 21	1.75	12	1.00	9	0.75
THR	> 79	1.81	> 52	4.77	8	0.73	13	1.19	6	0.55	12	1.10	> 20	1.83
TRP	11	0.93	0	0.00	4	1.35	4	1.35	3	1.01	4	1.35	2	0.68
TYR	29	1.03	5	0.71	3	0.43	5	0.71	> 16	2.27	7	0.99	5	0.71
GLY	28	0.52	10	0.74	< 6	0.44	< 6	0.44	< 6	0.44	18	1.32	9	0.66
OTH	2	0.70	0	0.00	0	0.00	1	1.40	1	1.40	0	0.00	0	0.00
ALL	788	1.00	197	1.00	197	1.00	197	1.00	197	1.00	197	1.00	197	1.00

cluster 8: turn-type Vb

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	< 30	0.50	< 4	0.27	15	0.99	< 1	0.07	10	0.66	15	0.99	13	0.86
ILE	< 15	0.32	< 0	0.00	< 3	0.26	< 1	0.09	11	0.94	> 19	1.63	> 20	1.71
LEU	< 16	0.22	< 4	0.22	< 3	0.17	< 3	0.17	< 6	0.33	13	0.72	25	1.39
MET	7	0.45	1	0.26	< 0	0.00	1	0.26	5	1.29	3	0.77	5	1.29
PHE	22	0.65	< 2	0.24	< 2	0.24	5	0.59	13	1.53	13	1.53	> 15	1.77
PRO	< 7	0.19	< 0	0.00	7	0.78	< 0	0.00	< 0	0.00	4	0.44	10	1.11
VAL	< 27	0.48	< 3	0.21	< 4	0.28	< 6	0.43	14	1.00	11	0.78	20	1.43
ARG	32	0.80	4	0.40	8	0.80	9	0.90	11	1.10	14	1.40	10	1.00

ASP	> 110 2.37	> 35 3.01	> 44 3.79	> 26 2.24	< 5 0.43	< 4 0.34	8 0.69
GLU	68 1.27	< 3 0.22	> 32 2.39	18 1.35	15 1.12	16 1.20	< 4 0.30
LYS	51 1.06	< 1 0.08	17 1.41	18 1.49	15 1.25	17 1.41	< 5 0.42
ASN	> 87 2.45	> 34 3.83	> 17 1.91	> 22 2.48	14 1.58	4 0.45	7 0.79
CYS	4 0.31	1 0.31	0 0.00	1 0.31	2 0.61	4 1.22	> 7 2.14
GLN	27 0.87	3 0.39	5 0.65	11 1.42	8 1.03	8 1.03	3 0.39
HIS	29 1.49	2 0.41	5 1.03	> 10 2.05	> 12 2.46	5 1.03	3 0.62
SER	> 83 1.71	> 35 2.88	18 1.48	14 1.15	16 1.32	16 1.32	14 1.15
THR	> 90 2.03	> 35 3.16	7 0.63	> 31 2.80	17 1.54	11 0.99	11 0.99
TRP	8 0.67	0 0.00	3 1.00	2 0.67	3 1.00	5 1.66	2 0.67
TYR	30 1.05	< 2 0.28	< 2 0.28	12 1.68	> 14 1.96	9 1.26	10 1.40
GLY	53 0.96	> 29 2.10	7 0.51	9 0.65	8 0.58	9 0.65	8 0.58
OTH	4 1.38	2 2.76	1 1.38	0 0.00	1 1.38	0 0.00	0 0.00
ALL	800 1.00	200 1.00	200 1.00	200 1.00	200 1.00	200 1.00	200 1.00

cluster 9: turn-type VIa1

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	2	0.95	1	1.89	1	1.89	0	0.00	0	0.00	0	0.00	0	0.00
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	2.45	0	0.00
LEU	3	1.19	0	0.00	1	1.58	0	0.00	2	3.17	0	0.00	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	1	0.84	0	0.00	0	0.00	0	0.00	1	3.37	0	0.00	0	0.00
PRO	> 9	7.13	0	0.00	> 3	9.51	> 6	19.01	0	0.00	0	0.00	1	3.17
VAL	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	> 3	6.11
ARG	1	0.71	1	2.86	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	2.46
GLU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LYS	2	1.19	> 2	4.74	0	0.00	0	0.00	0	0.00	> 2	4.74	0	0.00
ASN	2	1.61	1	3.22	0	0.00	0	0.00	1	3.22	> 2	6.43	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	> 2	17.48
GLN	2	1.84	1	3.69	0	0.00	1	3.69	0	0.00	0	0.00	0	0.00
HIS	1	1.47	> 1	5.86	0	0.00	0	0.00	0	0.00	> 1	5.86	0	0.00
SER	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	2.35	0	0.00
THR	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	4	3.99	0	0.00	> 2	7.98	0	0.00	> 2	7.98	0	0.00	0	0.00
GLY	1	0.52	0	0.00	0	0.00	0	0.00	1	2.07	0	0.00	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	28	1.00	7	1.00	7	1.00	7	1.00	7	1.00	7	1.00	7	1.00

cluster 10: turn-type VIIa

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	P3	pos 4	P4	before	Pb	after	Pa
ALA	13	0.80	5	1.23	3	0.74	1	0.25	4	0.98	3	0.74	<	0 0.00
ILE	5	0.40	0	0.00	3	0.95	0	0.00	2	0.63	4	1.27	>	7 2.22
LEU	7	0.36	4	0.82	2	0.41	<	0 0.00	1	0.21	1	0.21	<	0 0.00
MET	3	0.72	2	1.91	0	0.00	0	0.00	1	0.95	0	0.00	0	0.00
PHE	5	0.55	1	0.44	1	0.44	1	0.44	2	0.87	3	1.31	1	0.44

PRO	14	1.44	0	0.00	>	14	5.75	0	0.00	0	0.00	4	1.64	>	6	2.47	
VAL	11	0.73	6	1.58		1	0.26	<	0	0.00	4	1.06	7	1.85	>	9	2.37
ARG	13	1.20	>	7	2.59	1	0.37	1	0.37	4	1.48	3	1.11	5	1.85		
ASP	5	0.40	0	0.00	2	0.64	2	0.64	1	0.32	3	0.96	3	0.96			
GLU	15	1.04	6	1.66	5	1.38	<	0	0.00	4	1.11	2	0.55	2	0.55		
LYS	20	1.54	2	0.62	>	8	2.46	3	0.92	>	7	2.15	3	0.92	4	1.23	
ASN	13	1.35	1	0.42	3	1.25	4	1.67	5	2.08	2	0.83	2	0.83			
CYS	5	1.42	2	2.27	0	0.00	0	0.00	>	3	3.40	1	1.13	2	2.27		
GLN	8	0.96	3	1.43	3	1.43	1	0.48	1	0.48	1	0.48	0	0.00			
HIS	4	0.76	1	0.76	1	0.76	0	0.00	2	1.52	0	0.00	1	0.76			
SER	14	1.07	6	1.83	3	0.91	3	0.91	2	0.61	5	1.52	4	1.22			
THR	5	0.42	1	0.33	2	0.67	0	0.00	2	0.67	5	1.67	1	0.33			
TRP	5	1.54	1	1.23	1	1.23	1	1.23	2	2.46	1	1.23	2	2.46			
TYR	7	0.91	1	0.52	0	0.00	2	1.03	4	2.07	0	0.00	3	1.55			
GLY	>	44	2.95	5	1.34	1	0.27	>	35	9.39	3	0.81	5	1.34	2	0.54	
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	5.10	0	0.00			
ALL	216	1.00	54	1.00	54	1.00	54	1.00	54	1.00	54	1.00	54	1.00			

cluster 11: turn-type VIIIb

aa	turn		position 1		position 2		position 3		position 4		before		after		
	f	P	f	P	f	P	f								
ALA	3	0.41	0	0.00	2	1.10	0	0.00	1	0.55	0	0.00	2	1.10	
ILE	2	0.36	0	0.00	1	0.71	0	0.00	1	0.71	2	1.43	2	1.43	
LEU	5	0.58	2	0.92	2	0.92	0	0.00	1	0.46	0	0.00	2	0.92	
MET	2	1.07	0	0.00	1	2.15	0	0.00	1	2.15	0	0.00	1	2.15	
PHE	4	0.98	2	1.97	0	0.00	0	0.00	2	1.97	1	0.98	0	0.00	
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.92	1	0.92	
VAL	2	0.30	0	0.00	0	0.00	0	0.00	2	1.19	2	1.19	0	0.00	
ARG	3	0.63	0	0.00	0	0.00	1	0.83	2	1.67	0	0.00	>	4	3.33
ASP	14	2.51	2	1.44	>	9	6.46	3	2.15	0	0.00	1	0.72	1	0.72
GLU	3	0.47	0	0.00	1	0.62	2	1.25	0	0.00	1	0.62	1	0.62	
LYS	8	1.38	3	2.08	1	0.69	2	1.38	2	1.38	2	1.38	2	1.38	
ASN	7	1.64	0	0.00	2	1.88	3	2.81	2	1.88	2	1.88	0	0.00	
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	2.55	0	0.00	
GLN	6	1.61	1	1.08	2	2.15	1	1.08	2	2.15	>	3	3.23	1	1.08
HIS	1	0.43	0	0.00	1	1.71	0	0.00	0	0.00	0	0.00	1	1.71	
SER	2	0.34	1	0.69	0	0.00	0	0.00	1	0.69	>	5	3.43	1	0.69
THR	8	1.51	3	2.26	2	1.51	0	0.00	3	2.26	1	0.75	1	0.75	
TRP	2	1.39	0	0.00	0	0.00	0	0.00	>	2	5.54	0	0.00	1	2.77
TYR	2	0.58	0	0.00	0	0.00	0	0.00	2	2.33	1	1.16	2	2.33	
GLY	>	22	3.32	>	10	6.04	0	0.00	>	12	7.24	0	0.00	1	0.60
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	>	1	11.48	0	0.00
ALL	96	1.00	24	1.00	24	1.00	24	1.00	24	1.00	24	1.00	24	1.00	

cluster 12: turn-type VIII

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	P3	pos 4	P4	before	Pb	after	Pa
ALA	4	0.44	0	0.00	0	0.00	1	0.44	3	1.32	0	0.00	2	0.88
ILE	5	0.71	0	0.00	0	0.00	2	1.14	3	1.71	0	0.00	3	1.71

GLY	> 13	3.14	> 5	4.83	> 3	2.90	0	0.00	> 5	4.83	1	0.97	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	> 1	18.37
ALL	60	1.00	15	1.00	15	1.00	15	1.00	15	1.00	15	1.00	15	1.00

cluster 16: turn-type XI

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	P3	pos 4	P4	before	Pb	after	Pa
ALA	3	0.55	2	1.47	0	0.00	0	0.00	1	0.74	2	1.47	0	0.00
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.95	0	0.00
LEU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2	1.23
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	> 2	5.72
PHE	2	0.66	1	1.31	0	0.00	0	0.00	1	1.31	1	1.31	1	1.31
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	> 4	4.93	0	0.00
VAL	1	0.20	1	0.79	0	0.00	0	0.00	0	0.00	1	0.79	2	1.58
ARG	3	0.83	1	1.11	1	1.11	1	1.11	0	0.00	> 4	4.45	1	1.11
ASP	5	1.20	2	1.91	0	0.00	> 3	2.87	0	0.00	0	0.00	0	0.00
GLU	4	0.83	0	0.00	1	0.83	3	2.49	0	0.00	1	0.83	1	0.83
LYS	2	0.46	1	0.92	0	0.00	0	0.00	1	0.92	1	0.92	1	0.92
ASN	5	1.56	0	0.00	0	0.00	> 5	6.25	0	0.00	0	0.00	1	1.25
CYS	1	0.85	0	0.00	1	3.40	0	0.00	0	0.00	0	0.00	0	0.00
GLN	5	1.79	2	2.87	1	1.43	0	0.00	2	2.87	0	0.00	1	1.43
HIS	2	1.14	1	2.28	1	2.28	0	0.00	0	0.00	0	0.00	0	0.00
SER	9	2.06	3	2.74	1	0.91	3	2.74	2	1.83	2	1.83	0	0.00
THR	3	0.75	0	0.00	0	0.00	2	2.01	1	1.00	0	0.00	1	1.00
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	5	1.94	> 4	6.21	0	0.00	0	0.00	1	1.55	0	0.00	> 4	6.21
GLY	> 22	4.43	0	0.00	> 12	9.66	1	0.81	> 9	7.24	1	0.81	1	0.81
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	72	1.00	18	1.00	18	1.00	18	1.00	18	1.00	18	1.00	18	1.00

cluster 17: turn-type XII

aa	turn		position 1		position 2		position 3		position 4		before		after	
	f	P	f	P	f	P	f	aa	f	P	f	P	f	P
ALA	2	0.66	2	2.65	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LEU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	> 3	3.33	1	1.11
MET	1	1.29	1	5.15	0	0.00	0	0.00	0	0.00	0	0.00	1	5.15
PHE	1	0.59	0	0.00	0	0.00	1	2.36	0	0.00	0	0.00	0	0.00
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	> 2	4.44	0	0.00
VAL	1	0.36	1	1.43	0	0.00	0	0.00	0	0.00	1	1.43	2	2.85
ARG	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASP	3	1.29	0	0.00	> 3	5.17	0	0.00	0	0.00	0	0.00	0	0.00
GLU	1	0.37	0	0.00	0	0.00	0	0.00	1	1.49	2	2.99	2	2.99
LYS	1	0.42	1	1.66	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASN	> 7	3.94	0	0.00	> 5	11.25	> 2	4.50	0	0.00	0	0.00	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	2	1.29	1	2.58	0	0.00	1	2.58	0	0.00	0	0.00	0	0.00
HIS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
SER	4	1.65	0	0.00	0	0.00	0	0.00	> 4	6.58	0	0.00	0	0.00

13.4 five residue turns

13.4.1 normal α -turns

cluster 1: turn-type I

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	>49624	1.44	> 8939	1.30	>10027	1.46	>10114	1.47	>10224	1.49	>10320	1.50
ILE	>29259	1.10	5392	1.01	> 5773	1.09	> 5963	1.12	> 6125	1.15	> 6006	1.13
LEU	>54907	1.34	> 9674	1.18	>10697	1.30	>10903	1.33	>11698	1.43	>11935	1.45
MET	>11601	1.31	> 2047	1.16	> 2251	1.27	> 2318	1.31	> 2454	1.39	> 2531	1.43
PHE	19026	0.99	< 3583	0.93	3816	0.99	3753	0.97	3921	1.02	3953	1.02
PRO	< 5809	0.28	< 2409	0.59	< 2093	0.51	< 905	0.22	< 402	0.10	< 0	0.00
VAL	<29644	0.93	< 5708	0.89	< 6019	0.94	< 6014	0.94	< 6098	0.95	< 5805	0.91
ARG	>27498	1.21	> 4959	1.09	> 5315	1.17	> 5548	1.22	> 5780	1.27	> 5896	1.30
ASP	<23173	0.88	> 5905	1.12	< 4705	0.89	< 4757	0.90	< 4126	0.78	< 3680	0.70
GLU	>38545	1.26	> 7708	1.26	> 8261	1.36	> 8422	1.38	> 7505	1.23	> 6649	1.09
LYS	>30638	1.12	5449	0.99	> 5943	1.08	> 6295	1.15	> 6376	1.16	> 6575	1.20
ASN	<16055	0.79	< 3594	0.89	< 2850	0.70	< 3000	0.74	< 3069	0.76	< 3542	0.88
CYS	< 5909	0.79	< 1158	0.78	< 1086	0.73	< 1144	0.77	< 1247	0.84	< 1274	0.86
GLN	>22925	1.30	> 4395	1.25	> 4596	1.30	> 4727	1.34	> 4671	1.32	> 4536	1.29
HIS	< 9736	0.88	< 1941	0.88	< 1842	0.83	< 1932	0.87	< 1962	0.89	< 2059	0.93
SER	<21773	0.79	< 5231	0.95	< 4139	0.75	< 4038	0.73	< 4094	0.74	< 4271	0.77
THR	<19985	0.79	< 4731	0.94	< 3824	0.76	< 3752	0.74	< 3975	0.79	< 3703	0.74
TRP	6831	1.00	1364	1.00	> 1445	1.06	1367	1.00	1350	0.99	1305	0.95
TYR	<15061	0.92	< 2883	0.89	< 3011	0.92	< 2944	0.90	< 3082	0.95	< 3141	0.96
GLY	<15481	0.49	< 3667	0.58	< 3014	0.48	< 2803	0.45	< 2516	0.40	< 3481	0.55
OTH	1773	1.07	316	0.96	345	1.04	353	1.07	> 373	1.13	> 386	1.17
ALL	455253	1.00	91053	1.00	91052	1.00	91052	1.00	91048	1.00	91048	1.00

cluster 2: turn-type I'

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	8	0.88	4	2.21	3	1.66	1	0.55	0	0.00	0	0.00
ILE	5	0.71	2	1.43	0	0.00	2	1.43	0	0.00	1	0.71
LEU	4	0.37	2	0.92	1	0.46	0	0.00	1	0.46	0	0.00
MET	1	0.43	1	2.15	0	0.00	0	0.00	0	0.00	0	0.00
PHE	1	0.20	1	0.98	0	0.00	0	0.00	0	0.00	0	0.00
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
VAL	1	0.12	1	0.59	0	0.00	0	0.00	0	0.00	0	0.00
ARG	5	0.83	0	0.00	0	0.00	1	0.83	0	0.00	> 4	3.33
ASP	12	1.72	> 5	3.59	2	1.44	> 4	2.87	0	0.00	1	0.72
GLU	7	0.87	0	0.00	1	0.62	4	2.49	0	0.00	2	1.25
LYS	5	0.69	1	0.69	3	2.08	1	0.69	0	0.00	0	0.00
ASN	> 20	3.75	2	1.88	> 6	5.63	> 5	4.69	> 6	5.63	1	0.94
CYS	4	2.04	1	2.55	1	2.55	0	0.00	1	2.55	1	2.55
GLN	6	1.29	0	0.00	0	0.00	1	1.08	0	0.00	> 5	5.38
HIS	2	0.68	0	0.00	2	3.42	0	0.00	0	0.00	0	0.00
SER	2	0.27	0	0.00	0	0.00	0	0.00	1	0.69	1	0.69
THR	8	1.20	2	1.51	1	0.75	1	0.75	2	1.51	2	1.51

TRP	4	2.22	0	0.00	0	0.00	1	2.77	1	2.77	>	2	5.54
TYR	7	1.63	2	2.33	1	1.16	1	1.16	2	2.33		1	1.16
GLY	18	2.17	0	0.00	3	1.81	2	1.21	>	10	6.04	3	1.81
OTH	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00	0	0.00
ALL	120	1.00	24	1.00	24	1.00	24	1.00		24	1.00	24	1.00

cluster 3: turn-type II

aa	turn		position 1		position 2		position 3		position 4		position 5						
	f	P	f	P	f	P	f	P	f	P	f	P					
ALA	43	0.84	9	0.88	13	1.27	6	0.58	7	0.68	8	0.78					
ILE	41	1.03	>	16	2.01	4	0.50	<	1	0.13	12	1.51	8	1.01			
LEU	42	0.69	13	1.06	7	0.57	<	2	0.16	9	0.73	11	0.90				
MET	7	0.53	0	0.00	2	0.76	1	0.38	4	1.51	0	0.00					
PHE	17	0.59	4	0.69	2	0.35	<	1	0.17	7	1.21	3	0.52				
PRO	30	0.98	4	0.65	>	25	4.08	<	0	0.00	<	1	0.16	<	0	0.00	
VAL	58	1.22	12	1.26	11	1.15	<	1	0.11	>	20	2.10	14	1.47			
ARG	27	0.79	7	1.03	7	1.03	2	0.29	7	1.03	4	0.59					
ASP	36	0.91	6	0.76	7	0.89	8	1.01	4	0.51	11	1.39					
GLU	39	0.86	13	1.43	9	0.99	<	3	0.33	7	0.77	7	0.77				
LYS	44	1.07	6	0.73	11	1.34	3	0.37	>	16	1.95	8	0.98				
ASN	30	0.99	5	0.83	2	0.33	>	11	1.82	3	0.50	9	1.49				
CYS	3	0.27	2	0.90	0	0.00	1	0.45	0	0.00	0	0.00					
GLN	26	0.99	7	1.33	8	1.52	1	0.19	5	0.95	5	0.95					
HIS	13	0.79	2	0.60	5	1.51	1	0.30	2	0.60	3	0.91					
SER	23	0.56	<	2	0.24	4	0.48	3	0.36	9	1.09	5	0.61				
THR	36	0.96	6	0.80	6	0.80	<	1	0.13	>	14	1.86	9	1.20			
TRP	13	1.27	>	6	2.93	4	1.96	0	0.00	2	0.98	1	0.49				
TYR	23	0.95	3	0.62	8	1.64	1	0.21	5	1.03	6	1.23					
GLY	>	128	2.73	12	1.28	<	1	0.11	>	89	9.48	<	2	0.21	>	24	2.56
OTH	1	0.41	1	2.03	0	0.00	0	0.00	0	0.00	0	0.00					
ALL	680	1.00	136	1.00	136	1.00	136	1.00	136	1.00	136	1.00					

cluster 4: turn-type IIIa

aa	turn		position 1		position 2		position 3		position 4		position 5					
	f	P	f	P	f	P	f	P	f	P	f	P				
ALA	51	1.19	8	0.93	<	2	0.23	10	1.16	12	1.39	>	19	2.21		
ILE	15	0.45	5	0.75	<	0	0.00	6	0.90	3	0.45	<	1	0.15		
LEU	31	0.60	<	4	0.39	<	3	0.29	10	0.97	5	0.49	9	0.88		
MET	10	0.90	2	0.90	0	0.00	2	0.90	3	1.36	3	1.36				
PHE	14	0.58	4	0.83	<	0	0.00	6	1.24	2	0.41	2	0.41			
PRO	4	0.16	<	0	0.00	<	0	0.00	2	0.39	2	0.39	<	0	0.00	
VAL	15	0.38	7	0.88	<	0	0.00	4	0.50	<	2	0.25	<	2	0.25	
ARG	33	1.16	9	1.58	<	1	0.18	7	1.23	6	1.05	10	1.75			
ASP	28	0.85	<	1	0.15	<	0	0.00	>	12	1.81	>	12	1.81	3	0.45
GLU	45	1.18	5	0.66	<	2	0.26	>	16	2.10	>	13	1.70	9	1.18	
LYS	28	0.82	6	0.87	<	1	0.15	5	0.73	10	1.46	6	0.87			
ASN	18	0.71	5	0.99	<	0	0.00	3	0.59	7	1.38	3	0.59			
CYS	9	0.97	1	0.54	2	1.07	4	2.15	1	0.54	1	0.54				
GLN	21	0.95	5	1.13	1	0.23	3	0.68	>	10	2.26	2	0.45			
HIS	8	0.58	3	1.08	0	0.00	1	0.36	3	1.08	1	0.36				
SER	25	0.72	2	0.29	9	1.30	7	1.01	6	0.87	<	1	0.14			
THR	30	0.95	11	1.74	<	1	0.16	6	0.95	8	1.27	4	0.63			

TRP	9	1.05	1	0.58	0	0.00	3	1.75	1	0.58	4	2.33					
TYR	13	0.64	5	1.23	<	0	0.00	5	1.23	1	0.25	2	0.49				
GLY	>	161	4.09	>	30	3.81	>	91	11.56	<	2	0.25	7	0.89	>	31	3.94
OTH	2	0.97	0	0.00	1	2.42	0	0.00	0	0.00	0	0.00	1	2.42			
ALL	570	1.00	114	1.00	114	1.00	114	1.00	114	1.00	114	1.00	114	1.00			

cluster 5: turn-type IIIb

aa	turn		position 1		position 2		position 3		position 4		position 5				
	f	P	f	P	f	P	f	P	f	P	f	P			
ALA	8	0.82	1	0.51	0	0.00	3	1.53	3	1.53	1	0.51			
ILE	6	0.79	3	1.98	0	0.00	3	1.98	0	0.00	0	0.00			
LEU	8	0.68	4	1.71	0	0.00	1	0.43	1	0.43	2	0.85			
MET	1	0.40	0	0.00	0	0.00	1	1.98	0	0.00	0	0.00			
PHE	1	0.18	0	0.00	0	0.00	0	0.00	1	0.91	0	0.00			
PRO	2	0.34	0	0.00	0	0.00	2	1.71	0	0.00	0	0.00			
VAL	9	0.99	>	5	2.74	0	0.00	4	2.19	0	0.00	0	0.00		
ARG	4	0.62	1	0.77	0	0.00	1	0.77	1	0.77	1	0.77			
ASP	6	0.80	0	0.00	0	0.00	1	0.66	>	5	3.31	0	0.00		
GLU	11	1.26	0	0.00	0	0.00	2	1.15	>	5	2.87	4	2.30		
LYS	7	0.89	3	1.92	0	0.00	1	0.64	0	0.00	3	1.92			
ASN	4	0.69	1	0.87	0	0.00	1	0.87	1	0.87	1	0.87			
CYS	4	1.88	>	3	7.06	0	0.00	0	0.00	1	2.35	0	0.00		
GLN	4	0.79	0	0.00	1	0.99	2	1.99	0	0.00	1	0.99			
HIS	5	1.58	2	3.16	0	0.00	0	0.00	1	1.58	2	3.16			
SER	10	1.26	0	0.00	2	1.26	2	1.26	3	1.90	3	1.90			
THR	5	0.70	0	0.00	2	1.39	0	0.00	2	1.39	1	0.70			
TRP	1	0.51	1	2.56	0	0.00	0	0.00	0	0.00	0	0.00			
TYR	3	0.65	1	1.08	0	0.00	1	1.08	1	1.08	0	0.00			
GLY	>	31	3.45	1	0.56	>	21	11.70	1	0.56	1	0.56	>	7	3.90
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00			
ALL	130	1.00	26	1.00	26	1.00	26	1.00	26	1.00	26	1.00			

cluster 6: turn-type III'

aa	turn		position 1		position 2		position 3		position 4		position 5			
	f	P	f	P	f	P	f	P	f	P	f	P		
ALA	15	0.99	2	0.66	>	7	2.32	1	0.33	2	0.66	3	0.99	
ILE	4	0.34	3	1.28	1	0.43	0	0.00	0	0.00	0	0.00		
LEU	13	0.72	6	1.66	3	0.83	<	0	0.00	1	0.28	3	0.83	
MET	1	0.26	0	0.00	0	0.00	0	0.00	0	0.00	1	1.29		
PHE	5	0.59	1	0.59	1	0.59	1	0.59	1	0.59	1	0.59		
PRO	5	0.56	0	0.00	>	5	2.77	0	0.00	0	0.00	0	0.00	
VAL	3	0.21	1	0.36	1	0.36	0	0.00	1	0.36	0	0.00		
ARG	15	1.50	3	1.50	3	1.50	1	0.50	2	1.00	>	6	3.00	
ASP	11	0.95	3	1.29	2	0.86	4	1.72	2	0.86	0	0.00		
GLU	12	0.90	3	1.12	3	1.12	1	0.37	5	1.87	0	0.00		
LYS	13	1.08	3	1.25	3	1.25	1	0.42	0	0.00	>	6	2.49	
ASN	18	2.03	0	0.00	1	0.56	>	8	4.50	>	9	5.06	0	0.00
CYS	1	0.31	1	1.53	0	0.00	0	0.00	0	0.00	0	0.00		
GLN	11	1.42	3	1.94	1	0.65	2	1.29	1	0.65	>	4	2.58	
HIS	4	0.82	0	0.00	0	0.00	>	4	4.11	0	0.00	0	0.00	
SER	10	0.82	1	0.41	1	0.41	0	0.00	2	0.82	>	6	2.47	
THR	11	0.99	2	0.90	4	1.81	1	0.45	1	0.45	3	1.36		

TRP	2	0.67	1	1.66	0	0.00	0	0.00	1	1.66	0	0.00
TYR	9	1.26	2	1.40	1	0.70	1	0.70	> 4	2.79	1	0.70
GLY	> 37	2.68	5	1.81	3	1.09	> 15	5.43	> 8	2.90	> 6	2.17
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	200	1.00	40	1.00	40	1.00	40	1.00	40	1.00	40	1.00

cluster 7: turn-type IV

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	2	0.18	0	0.00	2	0.91	0	0.00	0	0.00	0	0.00
ILE	14	1.65	> 6	3.54	0	0.00	0	0.00	2	1.18	> 6	3.54
LEU	10	0.77	4	1.53	1	0.38	0	0.00	2	0.77	3	1.15
MET	4	1.42	0	0.00	0	0.00	2	3.55	0	0.00	2	3.55
PHE	4	0.65	1	0.81	0	0.00	1	0.81	0	0.00	2	1.63
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
VAL	10	0.98	> 6	2.95	0	0.00	0	0.00	2	0.98	2	0.98
ARG	7	0.97	2	1.38	3	2.07	0	0.00	2	1.38	0	0.00
ASP	14	1.66	0	0.00	> 6	3.56	4	2.38	3	1.78	1	0.59
GLU	9	0.93	1	0.52	2	1.03	2	1.03	3	1.55	1	0.52
LYS	10	1.15	1	0.57	1	0.57	2	1.15	4	2.29	2	1.15
ASN	13	2.02	0	0.00	> 4	3.10	> 4	3.10	3	2.33	2	1.55
CYS	2	0.84	0	0.00	1	2.11	0	0.00	0	0.00	1	2.11
GLN	5	0.89	0	0.00	1	0.89	0	0.00	3	2.67	1	0.89
HIS	3	0.85	0	0.00	2	2.83	0	0.00	1	1.42	0	0.00
SER	12	1.36	1	0.57	3	1.70	> 5	2.84	3	1.70	0	0.00
THR	4	0.50	3	1.87	0	0.00	0	0.00	0	0.00	1	0.62
TRP	1	0.46	0	0.00	0	0.00	0	0.00	0	0.00	1	2.29
TYR	9	1.73	> 3	2.89	1	0.96	1	0.96	1	0.96	> 3	2.89
GLY	12	1.20	1	0.50	2	1.00	> 8	4.00	0	0.00	1	0.50
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	145	1.00	29	1.00	29	1.00	29	1.00	29	1.00	29	1.00

cluster 8: turn-type V

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	4	0.44	0	0.00	3	1.66	0	0.00	0	0.00	1	0.55
ILE	4	0.57	1	0.71	0	0.00	0	0.00	1	0.71	2	1.43
LEU	10	0.92	3	1.39	2	0.92	4	1.85	0	0.00	1	0.46
MET	1	0.43	0	0.00	0	0.00	1	2.15	0	0.00	0	0.00
PHE	5	0.98	1	0.98	2	1.97	1	0.98	0	0.00	1	0.98
PRO	1	0.19	1	0.92	0	0.00	0	0.00	0	0.00	0	0.00
VAL	4	0.48	1	0.59	1	0.59	2	1.19	0	0.00	0	0.00
ARG	6	1.00	2	1.67	2	1.67	0	0.00	0	0.00	2	1.67
ASP	6	0.86	2	1.44	1	0.72	1	0.72	2	1.44	0	0.00
GLU	10	1.25	3	1.87	4	2.49	1	0.62	2	1.25	0	0.00
LYS	6	0.83	0	0.00	1	0.69	2	1.38	1	0.69	2	1.38
ASN	13	2.44	1	0.94	1	0.94	> 7	6.56	3	2.81	1	0.94
CYS	3	1.53	1	2.55	> 2	5.10	0	0.00	0	0.00	0	0.00
GLN	2	0.43	1	1.08	0	0.00	0	0.00	1	1.08	0	0.00
HIS	7	2.40	0	0.00	2	3.42	2	3.42	1	1.71	2	3.42
SER	8	1.10	> 4	2.74	0	0.00	0	0.00	3	2.06	1	0.69
THR	4	0.60	1	0.75	0	0.00	1	0.75	0	0.00	2	1.51

TRP	1	0.55	0	0.00	1	2.77	0	0.00	0	0.00	0	0.00
TYR	2	0.47	1	1.16	0	0.00	1	1.16	0	0.00	0	0.00
GLY	> 23	2.78	1	0.60	2	1.21	1	0.60	> 10	6.04	> 9	5.43
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	120	1.00	24	1.00	24	1.00	24	1.00	24	1.00	24	1.00

cluster 9: turn-type VI

aa	turn		position 1		position 2		position 3		position 4		position 5		
	f	P	f	P	f	P	f	P	f	P	f	P	
ALA	16	0.56	3	0.53	5	0.88	<	1	0.18	5	0.88	2	0.35
ILE	11	0.50	2	0.46	2	0.46	<	0	0.00	3	0.69	4	0.91
LEU	21	0.62	7	1.03	3	0.44	<	0	0.00	6	0.89	5	0.74
MET	3	0.41	0	0.00	2	1.37		0	0.00	0	0.00	1	0.69
PHE	19	1.20	2	0.63	> 7	2.20		0	0.00	6	1.89	4	1.26
PRO	> 84	4.97	3	0.89	> 9	2.66	>	72	21.30	0	0.00	0	0.00
VAL	14	0.53	4	0.76	1	0.19	<	0	0.00	3	0.57	6	1.14
ARG	11	0.59	2	0.53	4	1.07	<	0	0.00	3	0.80	2	0.53
ASP	18	0.83	> 9	2.07	1	0.23	<	0	0.00	3	0.69	5	1.15
GLU	32	1.27	7	1.40	> 10	1.99	<	0	0.00	> 11	2.19	4	0.80
LYS	9	0.40	2	0.44	1	0.22		2	0.44	1	0.22	3	0.66
ASN	26	1.56	> 7	2.10	6	1.80		0	0.00	> 8	2.40	5	1.50
CYS	5	0.82	2	1.63	2	1.63		0	0.00	0	0.00	1	0.82
GLN	9	0.62	2	0.69	3	1.03		0	0.00	3	1.03	1	0.34
HIS	6	0.66	1	0.55	2	1.10		0	0.00	1	0.55	2	1.10
SER	18	0.79	5	1.10	2	0.44	<	0	0.00	4	0.88	7	1.54
THR	12	0.58	4	0.96	2	0.48	<	0	0.00	4	0.96	2	0.48
TRP	9	1.60	2	1.77	3	2.66		0	0.00	2	1.77	2	1.77
TYR	24	1.79	3	1.12	> 9	3.35		0	0.00	> 8	2.98	4	1.49
GLY	27	1.04	8	1.55	1	0.19	<	0	0.00	3	0.58	> 15	2.90
OTH	1	0.74	0	0.00	0	0.00		0	0.00	1	3.67	0	0.00
ALL	375	1.00	75	1.00	75	1.00		75	1.00	75	1.00	75	1.00

13.4.2 open α -turns

cluster 1: turn-type Ia

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	aa	f	P	f	P	f
ALA	< 2901	0.92	< 515	0.82	< 266	0.42	> 726	1.15	> 709	1.12	> 685	1.09
ILE	< 1554	0.64	> 733	1.50	< 69	0.14	< 352	0.72	< 137	0.28	< 263	0.54
LEU	< 2753	0.73	> 1094	1.45	< 208	0.28	< 616	0.82	< 323	0.43	< 512	0.68
MET	< 638	0.79	> 231	1.42	< 51	0.31	< 100	0.62	< 92	0.57	164	1.01
PHE	< 1377	0.78	> 449	1.27	< 132	0.37	< 242	0.68	< 209	0.59	345	0.97
PRO	> 3113	1.65	> 442	1.17	> 787	2.09	> 1771	4.70	< 105	0.28	< 8	0.02
VAL	< 1889	0.64	> 805	1.37	< 95	0.16	< 446	0.76	< 156	0.27	< 387	0.66
ARG	< 1629	0.78	< 374	0.90	< 261	0.63	< 358	0.86	379	0.91	< 257	0.62
ASP	> 4201	1.73	< 322	0.66	> 1244	2.56	475	0.98	> 1121	2.31	> 1039	2.14
GLU	> 3648	1.30	< 315	0.56	< 277	0.50	> 719	1.29	> 1319	2.36	> 1018	1.82
LYS	< 2107	0.84	< 418	0.83	< 330	0.66	532	1.06	488	0.97	< 339	0.67
ASN	> 2195	1.18	< 264	0.71	> 781	2.10	< 179	0.48	> 596	1.60	375	1.01
CYS	601	0.88	159	1.16	147	1.08	< 91	0.67	< 63	0.46	141	1.03
GLN	1598	0.99	< 237	0.73	< 169	0.52	< 260	0.80	> 359	1.11	> 573	1.77
HIS	1073	1.05	198	0.97	> 267	1.31	< 146	0.72	222	1.09	> 240	1.18
SER	> 3961	1.56	< 362	0.71	> 1649	3.25	523	1.03	> 862	1.70	> 565	1.11
THR	> 2801	1.21	435	0.94	> 1204	2.60	< 278	0.60	475	1.03	< 409	0.88
TRP	604	0.96	> 148	1.18	< 50	0.40	118	0.94	> 155	1.23	133	1.06
TYR	< 1169	0.78	> 343	1.15	< 129	0.43	< 186	0.62	< 195	0.65	316	1.06
GLY	< 1880	0.65	< 479	0.83	< 234	0.41	< 224	0.39	< 377	0.65	566	0.98
OTH	103	0.68	36	1.19	< 9	0.30	< 17	0.56	< 17	0.56	24	0.79
ALL	41795	1.00	8359	1.00	8359	1.00	8359	1.00	8359	1.00	8359	1.00

cluster 2: turn-type Ib

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	203	0.80	< 32	0.63	< 5	0.10	> 71	1.40	50	0.99	45	0.89
ILE	< 120	0.61	> 57	1.45	< 0	0.00	< 20	0.51	< 24	0.61	< 19	0.49
LEU	< 206	0.68	> 96	1.59	< 2	0.03	46	0.76	< 30	0.50	< 32	0.53
MET	42	0.65	15	1.15	< 1	0.08	9	0.69	< 6	0.46	11	0.84
PHE	110	0.77	> 47	1.65	< 2	0.07	19	0.67	20	0.70	22	0.77
PRO	107	0.71	> 48	1.59	< 7	0.23	34	1.12	< 8	0.26	< 10	0.33
VAL	< 141	0.60	> 66	1.40	< 3	0.06	< 21	0.45	< 26	0.55	< 25	0.53
ARG	< 105	0.63	< 18	0.54	< 5	0.15	36	1.07	29	0.86	< 17	0.51
ASP	> 408	2.09	< 15	0.39	> 183	4.70	> 61	1.57	> 73	1.87	> 76	1.95
GLU	266	1.19	< 15	0.33	< 5	0.11	> 90	2.00	> 87	1.94	> 69	1.54
LYS	161	0.80	30	0.74	< 6	0.15	48	1.19	> 58	1.44	< 19	0.47
ASN	> 228	1.53	20	0.67	> 114	3.82	22	0.74	> 45	1.51	27	0.91
CYS	52	0.95	> 22	2.01	5	0.46	7	0.64	< 4	0.37	14	1.28
GLN	95	0.73	< 11	0.42	< 3	0.12	21	0.81	35	1.35	25	0.96
HIS	84	1.03	9	0.55	< 5	0.31	21	1.29	24	1.47	> 25	1.53
SER	253	1.24	< 20	0.49	> 80	1.96	> 54	1.32	52	1.27	47	1.15
THR	233	1.25	38	1.02	> 67	1.80	< 24	0.65	38	1.02	> 66	1.78

TRP	47	0.93	14	1.39	< 3	0.30	9	0.89	12	1.19	9	0.89
TYR	92	0.77	21	0.87	< 8	0.33	27	1.12	23	0.96	< 13	0.54
GLY	> 388	1.68	> 70	1.51	> 166	3.58	< 29	0.63	< 27	0.58	> 96	2.07
OTH	14	1.15	> 7	2.87	1	0.41	2	0.82	0	0.00	4	1.64
ALL	3355	1.00	671	1.00	671	1.00	671	1.00	671	1.00	671	1.00

cluster 3: turn-type II

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	272	0.86	59	0.93	53	0.84	< 48	0.76	51	0.81	61	0.97
ILE	229	0.94	47	0.96	42	0.86	57	1.17	60	1.23	< 23	0.47
LEU	400	1.06	80	1.06	< 51	0.68	76	1.01	> 112	1.48	81	1.07
MET	72	0.89	15	0.92	14	0.86	< 8	0.49	17	1.05	18	1.11
PHE	215	1.21	30	0.85	33	0.93	> 47	1.33	> 56	1.58	> 49	1.38
PRO	< 94	0.50	< 20	0.53	< 15	0.40	< 5	0.13	< 2	0.05	> 52	1.38
VAL	274	0.93	< 43	0.73	46	0.78	71	1.21	> 74	1.26	< 40	0.68
ARG	230	1.10	49	1.17	> 62	1.48	37	0.88	31	0.74	51	1.22
ASP	212	0.87	60	1.23	55	1.13	41	0.84	< 33	0.68	< 23	0.47
GLU	310	1.11	> 75	1.34	> 85	1.52	> 71	1.27	55	0.98	< 24	0.43
LYS	264	1.05	> 65	1.29	> 66	1.31	44	0.87	< 34	0.67	55	1.09
ASN	235	1.26	> 54	1.45	48	1.29	42	1.13	36	0.97	> 55	1.48
CYS	88	1.29	11	0.80	14	1.02	13	0.95	> 35	2.56	15	1.10
GLN	130	0.80	27	0.83	32	0.99	27	0.83	< 19	0.59	25	0.77
HIS	118	1.16	21	1.03	18	0.88	28	1.37	29	1.42	22	1.08
SER	214	0.84	43	0.85	52	1.02	48	0.94	< 31	0.61	40	0.79
THR	227	0.98	35	0.76	59	1.27	54	1.17	48	1.04	< 31	0.67
TRP	72	1.14	12	0.95	13	1.03	15	1.19	> 22	1.75	10	0.80
TYR	> 225	1.50	> 45	1.50	37	1.24	> 57	1.90	> 48	1.60	38	1.27
GLY	285	0.99	46	0.80	< 40	0.69	< 43	0.74	< 39	0.68	> 117	2.03
OTH	19	1.25	0	0.00	2	0.66	5	1.65	5	1.65	> 7	2.30
ALL	4185	1.00	837	1.00	837	1.00	837	1.00	837	1.00	837	1.00

cluster 4: turn-type III

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	251	0.96	< 37	0.71	> 79	1.51	63	1.20	< 21	0.40	51	0.97
ILE	< 129	0.64	36	0.89	< 17	0.42	< 18	0.44	32	0.79	< 26	0.64
LEU	< 209	0.67	< 45	0.72	< 43	0.69	< 47	0.75	< 21	0.34	53	0.85
MET	37	0.55	12	0.89	< 3	0.22	< 4	0.30	9	0.67	9	0.67
PHE	127	0.86	39	1.33	31	1.05	< 11	0.37	26	0.88	20	0.68
PRO	> 408	2.61	> 44	1.41	> 82	2.62	> 83	2.65	< 10	0.32	> 189	6.04
VAL	208	0.85	47	0.97	44	0.90	< 28	0.57	47	0.97	42	0.86
ARG	147	0.85	27	0.78	25	0.72	39	1.12	31	0.89	25	0.72
ASP	> 289	1.43	44	1.09	29	0.72	> 67	1.66	> 131	3.25	< 18	0.45
GLU	229	0.99	45	0.97	< 24	0.52	> 95	2.05	37	0.80	< 28	0.60
LYS	204	0.98	49	1.17	< 21	0.50	53	1.27	47	1.12	34	0.81
ASN	195	1.26	30	0.97	28	0.91	< 19	0.62	> 101	3.28	< 17	0.55
CYS	48	0.85	7	0.62	17	1.50	6	0.53	8	0.71	10	0.88
GLN	135	1.00	20	0.74	28	1.04	27	1.00	34	1.26	26	0.97

HIS	90	1.07	24	1.42	18	1.07	16	0.95	22	1.30	10	0.59
SER	233	1.10	31	0.74	> 81	1.92	53	1.26	30	0.71	38	0.90
THR	217	1.13	49	1.27	> 65	1.69	35	0.91	37	0.96	31	0.81
TRP	39	0.75	11	1.05	5	0.48	6	0.57	8	0.77	9	0.86
TYR	111	0.89	29	1.17	26	1.05	< 10	0.40	26	1.05	20	0.81
GLY	< 149	0.62	> 65	1.36	< 24	0.50	< 12	0.25	< 13	0.27	35	0.73
OTH	15	1.19	3	1.19	4	1.59	2	0.79	3	1.19	3	1.19
ALL	3470	1.00	694	1.00	694	1.00	694	1.00	694	1.00	694	1.00

cluster 5: turn-type IV

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	192	0.85	41	0.90	46	1.01	39	0.86	37	0.81	< 29	0.64
ILE	< 105	0.60	28	0.80	30	0.85	31	0.88	< 12	0.34	< 4	0.11
LEU	< 200	0.74	67	1.23	42	0.77	59	1.09	< 26	0.48	< 6	0.11
MET	40	0.68	16	1.37	10	0.86	< 5	0.43	7	0.60	< 2	0.17
PHE	142	1.11	32	1.25	< 15	0.59	26	1.02	> 63	2.47	< 6	0.24
PRO	> 430	3.17	21	0.77	21	0.77	< 1	0.04	< 1	0.04	> 386	14.22
VAL	< 118	0.56	31	0.73	< 25	0.59	40	0.95	< 15	0.36	< 7	0.17
ARG	171	1.14	37	1.23	> 49	1.63	40	1.33	24	0.80	21	0.70
ASP	206	1.18	42	1.20	29	0.83	32	0.92	> 93	2.66	< 10	0.29
GLU	204	1.01	41	1.02	> 74	1.84	52	1.29	< 18	0.45	< 19	0.47
LYS	231	1.27	> 54	1.49	> 76	2.10	> 53	1.46	33	0.91	< 15	0.41
ASN	> 196	1.47	32	1.20	31	1.16	34	1.27	> 82	3.07	17	0.64
CYS	34	0.69	4	0.41	< 1	0.10	8	0.81	> 19	1.93	< 2	0.20
GLN	115	0.99	22	0.94	> 37	1.59	32	1.37	19	0.81	< 5	0.21
HIS	92	1.25	17	1.16	11	0.75	22	1.50	> 37	2.52	< 5	0.34
SER	145	0.79	35	0.96	37	1.01	< 24	0.66	26	0.71	< 23	0.63
THR	126	0.76	< 19	0.57	28	0.84	> 48	1.44	< 18	0.54	< 13	0.39
TRP	40	0.88	8	0.88	7	0.77	5	0.55	14	1.55	6	0.66
TYR	141	1.31	> 31	1.44	16	0.74	> 37	1.72	> 50	2.32	< 7	0.33
GLY	< 77	0.37	< 23	0.55	< 17	0.41	< 13	0.31	< 7	0.17	< 17	0.41
OTH	5	0.46	1	0.46	0	0.00	1	0.46	1	0.46	2	0.92
ALL	3010	1.00	602	1.00	602	1.00	602	1.00	602	1.00	602	1.00

cluster 6: turn-type V

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	22	0.79	6	1.07	3	0.54	4	0.72	4	0.72	5	0.90
ILE	14	0.65	3	0.69	2	0.46	< 0	0.00	6	1.39	3	0.69
LEU	16	0.48	2	0.30	3	0.45	2	0.30	2	0.30	7	1.05
MET	6	0.84	> 4	2.78	0	0.00	0	0.00	0	0.00	2	1.39
PHE	10	0.64	5	1.59	0	0.00	2	0.64	0	0.00	3	0.96
PRO	28	1.68	4	1.20	4	1.20	0	0.00	> 8	2.40	> 12	3.60
VAL	21	0.81	5	0.96	< 0	0.00	< 0	0.00	4	0.77	> 12	2.31
ARG	13	0.70	3	0.81	3	0.81	4	1.08	3	0.81	< 0	0.00
ASP	34	1.58	2	0.47	> 10	2.33	7	1.63	> 14	3.26	1	0.23
GLU	23	0.93	4	0.81	5	1.01	4	0.81	5	1.01	5	1.01
LYS	16	0.72	3	0.67	4	0.90	< 0	0.00	8	1.79	1	0.22

ASN	14	0.85	2	0.61	6	1.82	3	0.91	2	0.61	1	0.30
CYS	9	1.49	> 4	3.31	2	1.65	1	0.83	0	0.00	2	1.65
GLN	11	0.77	3	1.05	2	0.70	1	0.35	4	1.40	1	0.35
HIS	11	1.22	4	2.22	2	1.11	3	1.66	1	0.56	1	0.56
SER	23	1.02	5	1.11	6	1.33	6	1.33	3	0.67	3	0.67
THR	36	1.76	> 9	2.20	> 13	3.17	5	1.22	5	1.22	4	0.98
TRP	8	1.44	3	2.70	1	0.90	0	0.00	2	1.80	2	1.80
TYR	13	0.98	1	0.38	3	1.13	4	1.51	2	0.76	3	1.13
GLY	42	1.64	2	0.39	5	0.98	> 28	5.48	1	0.20	6	1.17
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	370	1.00	74	1.00	74	1.00	74	1.00	74	1.00	74	1.00

cluster 7: turn-type VI

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	aa	f	P	f	P	f
ALA	4	0.29	0	0.00	2	0.74	1	0.37	1	0.37	0	0.00
ILE	7	0.67	0	0.00	2	0.95	2	0.95	3	1.43	0	0.00
LEU	12	0.74	0	0.00	1	0.31	2	0.62	4	1.23	5	1.54
MET	3	0.86	1	1.43	1	1.43	0	0.00	0	0.00	1	1.43
PHE	6	0.79	2	1.31	0	0.00	0	0.00	3	1.97	1	0.66
PRO	10	1.23	2	1.23	0	0.00	0	0.00	0	0.00	> 8	4.93
VAL	22	1.74	4	1.58	> 7	2.77	5	1.98	5	1.98	1	0.40
ARG	8	0.89	1	0.56	1	0.56	2	1.11	2	1.11	2	1.11
ASP	11	1.05	1	0.48	2	0.96	3	1.44	3	1.44	2	0.96
GLU	14	1.16	2	0.83	5	2.08	3	1.25	1	0.42	3	1.25
LYS	10	0.92	1	0.46	> 5	2.31	1	0.46	1	0.46	2	0.92
ASN	4	0.50	2	1.25	1	0.63	1	0.63	0	0.00	0	0.00
CYS	1	0.34	1	1.70	0	0.00	0	0.00	0	0.00	0	0.00
GLN	7	1.00	1	0.72	1	0.72	2	1.43	1	0.72	2	1.43
HIS	5	1.14	0	0.00	0	0.00	2	2.28	2	2.28	1	1.14
SER	15	1.37	4	1.83	1	0.46	3	1.37	> 5	2.29	2	0.91
THR	15	1.51	0	0.00	> 6	3.01	> 5	2.51	1	0.50	3	1.51
TRP	3	1.11	0	0.00	1	1.85	1	1.85	1	1.85	0	0.00
TYR	5	0.78	0	0.00	0	0.00	1	0.78	3	2.33	1	0.78
GLY	18	1.45	> 14	5.63	0	0.00	2	0.81	0	0.00	2	0.81
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	180	1.00	36	1.00	36	1.00	36	1.00	36	1.00	36	1.00

cluster 8: turn-type VII

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	aa	f	P	f	P	f
ALA	44	1.00	14	1.58	7	0.79	7	0.79	< 3	0.34	13	1.47
ILE	28	0.82	6	0.88	< 1	0.15	7	1.02	3	0.44	11	1.61
LEU	36	0.68	8	0.76	< 2	0.19	14	1.33	< 2	0.19	10	0.95
MET	5	0.44	2	0.88	1	0.44	2	0.88	0	0.00	0	0.00
PHE	26	1.05	9	1.82	5	1.01	5	1.01	4	0.81	3	0.61
PRO	> 71	2.69	9	1.71	8	1.52	< 0	0.00	> 35	6.64	> 19	3.60
VAL	20	0.49	5	0.61	< 2	0.24	3	0.37	5	0.61	5	0.61
ARG	33	1.13	2	0.34	6	1.03	> 15	2.56	5	0.86	5	0.86

ASP	34	1.00	3	0.44	9	1.32	4	0.59	> 16	2.35	2	0.29
GLU	30	0.77	7	0.89	6	0.77	7	0.89	5	0.64	5	0.64
LYS	45	1.28	6	0.85	9	1.28	> 14	1.99	10	1.42	6	0.85
ASN	32	1.23	3	0.58	> 16	3.08	6	1.15	4	0.77	3	0.58
CYS	8	0.84	2	1.05	0	0.00	3	1.57	3	1.57	0	0.00
GLN	27	1.19	3	0.66	6	1.32	6	1.32	4	0.88	8	1.76
HIS	18	1.26	4	1.40	2	0.70	1	0.35	3	1.05	> 8	2.81
SER	54	1.52	> 14	1.97	11	1.55	10	1.41	9	1.26	10	1.41
THR	34	1.05	8	1.24	> 15	2.32	3	0.46	3	0.46	5	0.77
TRP	10	1.14	3	1.71	2	1.14	2	1.14	1	0.57	2	1.14
TYR	12	0.57	2	0.48	4	0.96	4	0.96	1	0.24	1	0.24
GLY	18	0.45	7	0.87	5	0.62	4	0.50	< 1	0.12	< 1	0.12
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	585	1.00	117	1.00	117	1.00	117	1.00	117	1.00	117	1.00

cluster 9: turn-type VIII

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	18	1.08	2	0.60	5	1.51	4	1.20	3	0.90	4	1.20
ILE	14	1.09	3	1.17	1	0.39	2	0.78	2	0.78	> 6	2.34
LEU	14	0.71	3	0.76	5	1.26	3	0.76	1	0.25	2	0.50
MET	2	0.47	1	1.17	0	0.00	0	0.00	0	0.00	1	1.17
PHE	11	1.18	1	0.54	1	0.54	> 5	2.68	1	0.54	3	1.61
PRO	5	0.50	0	0.00	1	0.50	0	0.00	4	2.02	0	0.00
VAL	13	0.84	4	1.30	1	0.32	1	0.32	1	0.32	6	1.94
ARG	6	0.55	4	1.82	0	0.00	0	0.00	2	0.91	0	0.00
ASP	> 30	2.35	> 7	2.74	4	1.57	> 7	2.74	> 10	3.91	2	0.78
GLU	10	0.68	2	0.68	2	0.68	5	1.70	0	0.00	1	0.34
LYS	10	0.75	2	0.75	2	0.75	2	0.75	3	1.13	1	0.38
ASN	19	1.94	3	1.53	> 6	3.07	3	1.53	> 7	3.58	0	0.00
CYS	4	1.11	1	1.39	2	2.78	0	0.00	1	1.39	0	0.00
GLN	10	1.17	0	0.00	3	1.76	4	2.35	1	0.59	2	1.17
HIS	8	1.49	2	1.87	2	1.87	1	0.93	2	1.87	1	0.93
SER	13	0.97	3	1.12	3	1.12	2	0.75	3	1.12	2	0.75
THR	12	0.99	0	0.00	3	1.23	2	0.82	2	0.82	5	2.05
TRP	7	2.12	1	1.51	1	1.51	1	1.51	0	0.00	> 4	6.05
TYR	5	0.64	0	0.00	1	0.64	1	0.64	1	0.64	2	1.27
GLY	9	0.59	5	1.65	1	0.33	1	0.33	0	0.00	2	0.66
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	220	1.00	44	1.00	44	1.00	44	1.00	44	1.00	44	1.00

cluster 10: turn-type IXa

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	793	0.93	< 142	0.84	163	0.96	> 204	1.20	< 43	0.25	> 241	1.42
ILE	< 496	0.76	> 193	1.47	> 158	1.20	< 80	0.61	< 1	0.01	< 64	0.49
LEU	< 701	0.69	> 298	1.47	189	0.93	< 86	0.42	< 30	0.15	< 98	0.48
MET	195	0.89	> 57	1.31	45	1.03	< 19	0.44	< 19	0.44	55	1.26
PHE	< 352	0.74	98	1.03	< 69	0.72	< 48	0.50	< 35	0.37	102	1.07

PRO	> 1010	1.99	> 139	1.37	> 270	2.66	> 544	5.37	< 2	0.02	< 55	0.54
VAL	709	0.90	> 263	1.67	180	1.14	< 104	0.66	< 1	0.01	161	1.02
ARG	< 411	0.73	< 89	0.79	131	1.17	100	0.89	< 48	0.43	< 43	0.38
ASP	559	0.86	< 63	0.48	< 56	0.43	126	0.97	145	1.11	> 169	1.29
GLU	716	0.95	< 82	0.55	165	1.10	> 231	1.54	< 58	0.39	> 180	1.20
LYS	619	0.91	< 99	0.73	> 193	1.43	> 201	1.48	< 43	0.32	< 83	0.61
ASN	520	1.04	< 62	0.62	< 56	0.56	< 68	0.68	> 264	2.64	< 70	0.70
CYS	174	0.95	> 51	1.39	31	0.84	< 16	0.44	< 12	0.33	> 64	1.74
GLN	364	0.84	< 51	0.59	95	1.09	72	0.83	< 29	0.33	> 117	1.34
HIS	212	0.77	55	1.00	< 38	0.69	< 30	0.55	< 39	0.71	50	0.91
SER	602	0.88	< 108	0.79	< 102	0.75	< 106	0.78	< 53	0.39	> 233	1.71
THR	544	0.87	131	1.05	125	1.00	< 96	0.77	< 12	0.10	> 180	1.45
TRP	116	0.69	31	0.92	26	0.77	< 18	0.53	< 12	0.36	29	0.86
TYR	< 309	0.77	71	0.88	72	0.90	< 57	0.71	< 28	0.35	81	1.01
GLY	> 1809	2.33	153	0.99	< 78	0.50	< 38	0.25	> 1373	8.85	167	1.08
OTH	29	0.71	12	1.47	6	0.74	4	0.49	< 1	0.12	6	0.74
ALL	11240	1.00	2248	1.00	2248	1.00	2248	1.00	2248	1.00	2248	1.00

cluster 11: turn-type IXb

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	16	0.60	< 0	0.00	4	0.75	5	0.93	3	0.56	4	0.75
ILE	15	0.72	> 14	3.38	< 0	0.00	< 0	0.00	< 0	0.00	1	0.24
LEU	15	0.47	8	1.25	< 1	0.16	2	0.31	< 1	0.16	3	0.47
MET	4	0.58	1	0.73	0	0.00	2	1.45	0	0.00	1	0.73
PHE	8	0.53	5	1.66	0	0.00	0	0.00	2	0.67	1	0.33
PRO	23	1.44	6	1.88	6	1.88	> 9	2.81	0	0.00	2	0.63
VAL	31	1.24	9	1.81	2	0.40	3	0.60	< 0	0.00	> 17	3.41
ARG	10	0.56	2	0.56	1	0.28	3	0.85	3	0.85	1	0.28
ASP	13	0.63	< 0	0.00	1	0.24	6	1.46	4	0.97	2	0.49
GLU	16	0.67	< 0	0.00	3	0.63	6	1.26	1	0.21	6	1.26
LYS	20	0.94	2	0.47	4	0.94	> 11	2.57	< 0	0.00	3	0.70
ASN	25	1.59	1	0.32	0	0.00	6	1.90	> 17	5.39	1	0.32
CYS	10	1.72	> 6	5.17	0	0.00	1	0.86	0	0.00	3	2.59
GLN	11	0.80	2	0.73	2	0.73	2	0.73	0	0.00	5	1.82
HIS	11	1.27	2	1.16	2	1.16	2	1.16	4	2.31	1	0.58
SER	23	1.07	4	0.93	8	1.85	3	0.70	4	0.93	4	0.93
THR	12	0.61	3	0.76	< 0	0.00	4	1.02	2	0.51	3	0.76
TRP	3	0.56	1	0.94	0	0.00	0	0.00	1	0.94	1	0.94
TYR	7	0.55	2	0.79	0	0.00	4	1.57	0	0.00	1	0.39
GLY	> 81	3.31	2	0.41	> 37	7.55	2	0.41	> 29	5.92	> 11	2.24
OTH	1	0.78	1	3.88	0	0.00	0	0.00	0	0.00	0	0.00
ALL	355	1.00	71	1.00	71	1.00	71	1.00	71	1.00	71	1.00

cluster 12: turn-type X

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	120	0.78	30	0.98	25	0.82	30	0.98	< 8	0.26	27	0.88
ILE	< 55	0.46	< 7	0.30	20	0.84	< 13	0.55	< 1	0.04	< 14	0.59

LEU	180	0.98	32	0.87	> 83	2.27	< 22	0.60	< 8	0.22	35	0.96
MET	33	0.84	5	0.63	> 16	2.03	< 2	0.25	< 0	0.00	10	1.27
PHE	95	1.10	18	1.05	> 37	2.15	17	0.99	< 8	0.47	15	0.87
PRO	< 50	0.55	26	1.42	< 7	0.38	< 4	0.22	< 0	0.00	13	0.71
VAL	101	0.71	< 16	0.56	24	0.84	21	0.74	< 0	0.00	> 40	1.40
ARG	110	1.08	15	0.74	15	0.74	> 42	2.07	< 9	0.44	> 29	1.43
ASP	133	1.13	> 44	1.87	< 10	0.42	> 35	1.48	> 35	1.48	< 9	0.38
GLU	98	0.72	29	1.07	< 7	0.26	> 41	1.51	< 8	0.29	< 13	0.48
LYS	164	1.34	25	1.02	19	0.78	> 49	2.00	< 12	0.49	> 59	2.41
ASN	> 136	1.51	16	0.89	19	1.05	16	0.89	> 60	3.33	25	1.39
CYS	39	1.18	10	1.51	8	1.21	6	0.90	3	0.45	> 12	1.81
GLN	73	0.93	21	1.33	< 7	0.45	16	1.02	< 8	0.51	21	1.33
HIS	53	1.07	8	0.81	13	1.31	8	0.81	10	1.01	14	1.42
SER	108	0.88	> 38	1.54	19	0.77	23	0.93	< 12	0.49	16	0.65
THR	113	1.00	21	0.93	> 33	1.47	27	1.20	< 2	0.09	30	1.34
TRP	31	1.02	5	0.82	> 13	2.13	6	0.98	3	0.49	4	0.66
TYR	70	0.96	13	0.90	> 23	1.58	21	1.45	< 4	0.28	9	0.62
GLY	> 263	1.88	26	0.93	< 4	0.14	< 7	0.25	> 215	7.67	< 11	0.39
OTH	5	0.68	1	0.68	> 4	2.72	0	0.00	0	0.00	0	0.00
ALL	2030	1.00	406	1.00	406	1.00	406	1.00	406	1.00	406	1.00

cluster 13: turn-type XI

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	412	1.17	86	1.22	> 114	1.62	> 108	1.53	< 4	0.06	> 100	1.42
ILE	< 142	0.52	50	0.92	< 14	0.26	< 18	0.33	< 1	0.02	59	1.08
LEU	< 283	0.67	73	0.87	< 36	0.43	> 102	1.21	< 1	0.01	71	0.85
MET	73	0.81	18	0.99	12	0.66	21	1.16	< 2	0.11	20	1.11
PHE	< 119	0.60	< 26	0.66	< 6	0.15	28	0.71	< 6	0.15	> 53	1.34
PRO	< 79	0.38	< 13	0.31	> 64	1.52	< 0	0.00	< 0	0.00	< 2	0.05
VAL	< 172	0.53	53	0.81	< 20	0.31	< 22	0.34	< 0	0.00	77	1.18
ARG	234	1.00	> 72	1.55	> 70	1.50	50	1.07	< 4	0.09	38	0.82
ASP	322	1.19	> 69	1.27	61	1.13	> 137	2.53	< 27	0.50	< 28	0.52
GLU	268	0.86	55	0.88	> 126	2.02	53	0.85	< 3	0.05	< 31	0.50
LYS	270	0.96	62	1.10	> 94	1.67	50	0.89	< 10	0.18	54	0.96
ASN	262	1.26	42	1.01	> 59	1.42	> 102	2.46	< 21	0.51	38	0.92
CYS	49	0.64	11	0.72	< 5	0.33	17	1.12	< 2	0.13	14	0.92
GLN	180	1.00	43	1.19	> 50	1.38	> 57	1.58	< 3	0.08	27	0.75
HIS	110	0.97	31	1.37	18	0.79	> 34	1.50	< 7	0.31	20	0.88
SER	260	0.92	> 72	1.27	> 80	1.41	58	1.02	< 5	0.09	45	0.79
THR	< 169	0.66	> 68	1.32	38	0.74	< 17	0.33	< 2	0.04	44	0.85
TRP	41	0.59	10	0.71	8	0.57	< 5	0.36	< 0	0.00	18	1.28
TYR	133	0.80	26	0.78	< 10	0.30	33	0.99	< 1	0.03	> 63	1.89
GLY	> 1072	3.33	< 48	0.75	< 44	0.68	< 18	0.28	> 833	12.95	> 129	2.01
OTH	10	0.59	4	1.18	3	0.89	2	0.59	0	0.00	1	0.30
ALL	4660	1.00	932	1.00	932	1.00	932	1.00	932	1.00	932	1.00

cluster 14: turn-type XII

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	117	1.40	16	0.95	< 4	0.24	23	1.37	> 28	1.67	> 46	2.74
ILE	< 27	0.42	9	0.69	< 0	0.00	< 6	0.46	7	0.54	< 5	0.39
LEU	70	0.70	25	1.25	< 5	0.25	< 10	0.50	14	0.70	16	0.80
MET	16	0.74	4	0.93	< 0	0.00	8	1.86	3	0.70	1	0.23
PHE	38	0.81	11	1.17	< 0	0.00	6	0.64	9	0.96	12	1.27
PRO	29	0.58	4	0.40	< 0	0.00	15	1.50	5	0.50	5	0.50
VAL	< 30	0.39	14	0.90	< 0	0.00	< 5	0.32	< 2	0.13	9	0.58
ARG	44	0.79	10	0.90	< 3	0.27	8	0.72	10	0.90	13	1.17
ASP	63	0.98	< 6	0.47	< 4	0.31	> 29	2.25	> 21	1.63	< 3	0.23
GLU	61	0.82	16	1.08	< 0	0.00	20	1.35	15	1.01	10	0.67
LYS	48	0.72	11	0.82	< 1	0.08	15	1.12	10	0.75	11	0.82
ASN	34	0.69	4	0.41	< 0	0.00	9	0.91	15	1.52	6	0.61
CYS	14	0.77	4	1.10	0	0.00	4	1.10	4	1.10	2	0.55
GLN	32	0.74	8	0.93	3	0.35	4	0.47	10	1.16	7	0.81
HIS	35	1.29	9	1.66	3	0.56	8	1.48	> 12	2.22	3	0.56
SER	91	1.35	16	1.19	8	0.59	> 26	1.93	16	1.19	> 25	1.85
THR	30	0.49	12	0.98	< 2	0.16	< 5	0.41	6	0.49	< 5	0.41
TRP	16	0.96	6	1.80	2	0.60	5	1.50	3	0.90	0	0.00
TYR	36	0.91	7	0.88	< 2	0.25	8	1.01	12	1.51	7	0.88
GLY	> 277	3.61	> 29	1.89	> 185	12.07	8	0.52	19	1.24	> 36	2.35
OTH	2	0.50	1	1.24	0	0.00	0	0.00	1	1.24	0	0.00
ALL	1110	1.00	222	1.00	222	1.00	222	1.00	222	1.00	222	1.00

cluster 15: turn-type XIII

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	113	1.02	< 10	0.45	< 1	0.05	22	1.00	26	1.18	> 54	2.45
ILE	< 43	0.50	17	1.00	< 0	0.00	12	0.70	< 6	0.35	< 8	0.47
LEU	< 67	0.51	25	0.95	< 1	0.04	20	0.76	< 10	0.38	< 11	0.42
MET	12	0.42	5	0.88	< 0	0.00	3	0.53	< 1	0.18	3	0.53
PHE	56	0.91	16	1.29	< 3	0.24	18	1.45	< 5	0.40	14	1.13
PRO	62	0.94	12	0.91	< 0	0.00	> 45	3.42	< 4	0.30	< 1	0.08
VAL	64	0.63	17	0.83	< 0	0.00	< 9	0.44	13	0.63	25	1.22
ARG	48	0.66	14	0.96	< 0	0.00	10	0.69	9	0.62	15	1.03
ASP	82	0.97	18	1.06	< 6	0.35	15	0.88	> 35	2.06	< 8	0.47
GLU	93	0.95	16	0.82	< 1	0.05	20	1.02	> 38	1.95	18	0.92
LYS	81	0.92	18	1.02	< 1	0.06	> 28	1.59	23	1.31	11	0.63
ASN	73	1.13	11	0.85	9	0.69	7	0.54	19	1.46	> 27	2.08
CYS	30	1.26	8	1.68	2	0.42	4	0.84	6	1.26	> 10	2.10
GLN	45	0.80	8	0.71	< 1	0.09	11	0.97	> 19	1.68	6	0.53
HIS	34	0.96	12	1.69	< 1	0.14	4	0.56	10	1.41	7	0.98
SER	93	1.05	14	0.79	15	0.85	21	1.18	24	1.35	19	1.07
THR	68	0.84	14	0.87	< 6	0.37	21	1.30	13	0.80	14	0.87
TRP	18	0.82	6	1.37	< 0	0.00	3	0.68	4	0.91	5	1.14
TYR	42	0.80	15	1.44	< 0	0.00	7	0.67	7	0.67	13	1.24
GLY	> 333	3.30	> 35	1.74	> 245	12.15	< 11	0.55	20	0.99	22	1.09
OTH	3	0.57	1	0.94	0	0.00	1	0.94	0	0.00	1	0.94

ALL	1460	1.00	292	1.00	292	1.00	292	1.00	292	1.00	292	1.00
-----	------	------	-----	------	-----	------	-----	------	-----	------	-----	------

cluster 16: turn-type XIV

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	33	0.99	8	1.20	9	1.35	2	0.30	6	0.90	8	1.20
ILE	8	0.31	< 0	0.00	6	1.17	1	0.20	1	0.20	< 0	0.00
LEU	29	0.73	> 16	2.02	4	0.50	< 1	0.13	3	0.38	5	0.63
MET	3	0.35	0	0.00	0	0.00	0	0.00	2	1.17	1	0.59
PHE	9	0.48	4	1.07	5	1.34	< 0	0.00	< 0	0.00	< 0	0.00
PRO	10	0.50	3	0.76	3	0.76	< 0	0.00	4	1.01	< 0	0.00
VAL	12	0.39	2	0.32	5	0.81	< 0	0.00	2	0.32	3	0.49
ARG	14	0.64	3	0.68	6	1.36	2	0.46	1	0.23	2	0.46
ASP	29	1.14	2	0.39	1	0.20	5	0.98	> 14	2.74	7	1.37
GLU	27	0.92	< 1	0.17	6	1.02	2	0.34	10	1.70	8	1.36
LYS	18	0.68	4	0.75	4	0.75	2	0.38	4	0.75	4	0.75
ASN	26	1.33	< 0	0.00	< 0	0.00	3	0.77	> 16	4.09	7	1.79
CYS	6	0.83	3	2.09	1	0.70	2	1.39	0	0.00	0	0.00
GLN	13	0.76	2	0.59	4	1.17	1	0.29	2	0.59	4	1.17
HIS	17	1.59	4	1.87	> 8	3.73	0	0.00	2	0.93	3	1.40
SER	37	1.38	5	0.94	7	1.31	9	1.68	> 11	2.06	5	0.94
THR	19	0.78	5	1.03	7	1.44	3	0.62	1	0.21	3	0.62
TRP	4	0.61	1	0.76	2	1.51	0	0.00	0	0.00	1	0.76
TYR	15	0.95	2	0.64	> 9	2.86	0	0.00	3	0.95	1	0.32
GLY	> 110	3.62	> 23	3.79	< 1	0.17	> 55	9.05	5	0.82	> 26	4.28
OTH	1	0.63	0	0.00	0	0.00	0	0.00	1	3.13	0	0.00
ALL	440	1.00	88	1.00	88	1.00	88	1.00	88	1.00	88	1.00

cluster 17: turn-type XVa

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	24	0.68	7	0.99	5	0.70	< 0	0.00	4	0.56	8	1.13
ILE	23	0.84	7	1.27	5	0.91	< 0	0.00	2	0.36	9	1.64
LEU	38	0.90	13	1.53	6	0.71	< 0	0.00	11	1.30	8	0.94
MET	5	0.55	1	0.55	1	0.55	0	0.00	1	0.55	2	1.10
PHE	15	0.75	> 8	2.01	1	0.25	< 0	0.00	2	0.50	4	1.00
PRO	27	1.27	6	1.42	5	1.18	< 0	0.00	8	1.89	8	1.89
VAL	32	0.97	5	0.76	5	0.76	< 0	0.00	9	1.36	> 13	1.97
ARG	27	1.15	7	1.49	7	1.49	1	0.21	8	1.70	4	0.85
ASP	19	0.70	< 1	0.18	4	0.73	< 0	0.00	> 11	2.02	3	0.55
GLU	17	0.54	3	0.48	7	1.11	< 0	0.00	6	0.95	< 1	0.16
LYS	17	0.60	3	0.53	5	0.88	< 0	0.00	3	0.53	6	1.06
ASN	15	0.72	1	0.24	6	1.44	< 0	0.00	5	1.20	3	0.72
CYS	7	0.91	3	1.95	1	0.65	0	0.00	0	0.00	3	1.95
GLN	17	0.93	2	0.55	> 8	2.20	0	0.00	2	0.55	5	1.37
HIS	12	1.05	2	0.87	4	1.75	0	0.00	4	1.75	2	0.87
SER	24	0.84	4	0.70	9	1.58	< 0	0.00	8	1.40	3	0.53
THR	23	0.88	3	0.58	8	1.54	< 0	0.00	5	0.96	7	1.35
TRP	6	0.85	3	2.12	0	0.00	0	0.00	3	2.12	0	0.00

TYR	16	0.95	5	1.49	5	1.49	0	0.00	2	0.59	4	1.19
GLY	> 106	3.27	10	1.54	2	0.31	> 93	14.33	< 0	0.00	< 1	0.15
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	470	1.00	94	1.00	94	1.00	94	1.00	94	1.00	94	1.00

cluster 18: turn-type XVb

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	9	0.60	2	0.66	2	0.66	0	0.00	1	0.33	4	1.32
ILE	7	0.60	0	0.00	3	1.28	0	0.00	4	1.71	0	0.00
LEU	8	0.44	3	0.83	1	0.28	< 0	0.00	< 0	0.00	4	1.11
MET	6	1.55	> 3	3.86	0	0.00	0	0.00	0	0.00	> 3	3.86
PHE	10	1.18	3	1.77	1	0.59	0	0.00	> 6	3.54	0	0.00
PRO	15	1.66	1	0.56	3	1.66	0	0.00	> 7	3.88	4	2.22
VAL	6	0.43	3	1.07	2	0.71	0	0.00	1	0.36	0	0.00
ARG	12	1.20	0	0.00	> 6	3.00	0	0.00	1	0.50	> 5	2.50
ASP	6	0.52	1	0.43	1	0.43	0	0.00	2	0.86	2	0.86
GLU	12	0.90	3	1.12	5	1.87	0	0.00	4	1.49	0	0.00
LYS	9	0.75	2	0.83	1	0.42	0	0.00	4	1.66	2	0.83
ASN	6	0.68	0	0.00	1	0.56	0	0.00	2	1.13	3	1.69
CYS	2	0.61	1	1.53	0	0.00	0	0.00	0	0.00	1	1.53
GLN	6	0.77	1	0.65	2	1.29	0	0.00	1	0.65	2	1.29
HIS	3	0.62	2	2.05	0	0.00	0	0.00	0	0.00	1	1.03
SER	12	0.99	3	1.23	> 6	2.47	0	0.00	2	0.82	1	0.41
THR	14	1.26	3	1.36	3	1.36	0	0.00	4	1.81	4	1.81
TRP	2	0.67	1	1.66	0	0.00	0	0.00	0	0.00	1	1.66
TYR	6	0.84	3	2.10	2	1.40	0	0.00	0	0.00	1	0.70
GLY	> 49	3.55	5	1.81	1	0.36	> 40	14.48	1	0.36	2	0.72
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	200	1.00	40	1.00	40	1.00	40	1.00	40	1.00	40	1.00

cluster 19: turn-type XVI

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5
ALA	12	0.79	1	0.33	6	1.99	2	0.66	0	0.00	3	0.99
ILE	5	0.43	4	1.71	1	0.43	0	0.00	0	0.00	0	0.00
LEU	9	0.50	2	0.56	2	0.56	< 0	0.00	1	0.28	4	1.11
MET	3	0.77	1	1.29	1	1.29	0	0.00	0	0.00	1	1.29
PHE	4	0.47	2	1.18	1	0.59	1	0.59	0	0.00	0	0.00
PRO	8	0.89	1	0.56	> 7	3.88	0	0.00	0	0.00	0	0.00
VAL	4	0.28	2	0.71	0	0.00	0	0.00	0	0.00	2	0.71
ARG	10	1.00	0	0.00	2	1.00	2	1.00	1	0.50	> 5	2.50
ASP	10	0.86	1	0.43	1	0.43	4	1.72	3	1.29	1	0.43
GLU	7	0.52	2	0.75	2	0.75	3	1.12	0	0.00	0	0.00
LYS	16	1.33	3	1.25	5	2.08	> 7	2.91	0	0.00	1	0.42
ASN	16	1.80	> 5	2.81	1	0.56	> 6	3.38	3	1.69	1	0.56
CYS	1	0.31	0	0.00	1	1.53	0	0.00	0	0.00	0	0.00
GLN	8	1.03	2	1.29	2	1.29	2	1.29	0	0.00	2	1.29
HIS	4	0.82	1	1.03	2	2.05	0	0.00	0	0.00	1	1.03

SER	6	0.49	1	0.41	1	0.41	1	0.41	0	0.00	3	1.23
THR	7	0.63	4	1.81	0	0.00	1	0.45	1	0.45	1	0.45
TRP	2	0.67	1	1.66	1	1.66	0	0.00	0	0.00	0	0.00
TYR	7	0.98	0	0.00	1	0.70	> 4	2.79	1	0.70	1	0.70
GLY	> 59	4.27	> 6	2.17	2	0.72	> 7	2.54	> 30	10.86	> 14	5.07
OTH	2	2.76	> 1	6.89	> 1	6.89	0	0.00	0	0.00	0	0.00
ALL	200	1.00	40	1.00	40	1.00	40	1.00	40	1.00	40	1.00

cluster 20: turn-type XVII

aa	turn		position 1		position 2		position 3		position 4		position 5					
	f	P	f	P	f	P	pos 3	P3	pos 4	P4	pos 5	P5				
ALA	15	0.58	8	1.56	3	0.58	<	0	0.00	2	0.39	2	0.39			
ILE	11	0.55	2	0.50	1	0.25	<	0	0.00	4	1.01	4	1.01			
LEU	14	0.46	<	0	0.00	2	0.33	<	1	0.16	9	1.47	2	0.33		
MET	5	0.76	1	0.76	1	0.76	0	0.00	1	0.76	2	1.51				
PHE	9	0.63	3	1.04	1	0.35	0	0.00	1	0.35	4	1.39				
PRO	5	0.33	3	0.98	1	0.33	0	0.00	0	0.00	1	0.33				
VAL	6	0.25	1	0.21	2	0.42	<	0	0.00	1	0.21	2	0.42			
ARG	22	1.29	4	1.18	3	0.88	2	0.59	>	10	2.94	3	0.88			
ASP	38	1.92	5	1.27	>	26	6.58	1	0.25	3	0.76	3	0.76			
GLU	17	0.75	5	1.10	<	0	0.00	<	0	0.00	3	0.66	>	9	1.98	
LYS	15	0.73	3	0.73	1	0.24	2	0.49	>	8	1.95	1	0.24			
ASN	19	1.26	5	1.66	>	7	2.32	0	0.00	3	0.99	4	1.32			
CYS	3	0.54	2	1.80	1	0.90	0	0.00	0	0.00	0	0.00				
GLN	11	0.84	4	1.52	1	0.38	0	0.00	4	1.52	2	0.76				
HIS	6	0.72	0	0.00	0	0.00	2	1.21	1	0.60	3	1.81				
SER	18	0.87	4	0.97	7	1.69	<	0	0.00	4	0.97	3	0.73			
THR	31	1.65	5	1.33	>	8	2.13	<	0	0.00	>	9	2.39	>	9	2.39
TRP	3	0.59	>	3	2.93	0	0.00	0	0.00	0	0.00	0	0.00			
TYR	12	0.99	3	1.23	1	0.41	1	0.41	4	1.64	3	1.23				
GLY	>	80	3.41	7	1.49	2	0.43	>	59	12.57	1	0.21	>	11	2.34	
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00				
ALL	340	1.00	68	1.00	68	1.00	68	1.00	68	1.00	68	1.00				

cluster 21: turn-type XVIII

aa	turn		position 1		position 2		position 3		position 4		position 5						
	f	P	f	P	f	P	aa	f	P	f	P	f					
ALA	26	0.84	5	0.81	5	0.81	2	0.32	<	0	0.00	>	14	2.26			
ILE	15	0.63	6	1.25	5	1.04	3	0.63	<	0	0.00	1	0.21				
LEU	22	0.60	5	0.68	6	0.81	4	0.54	<	0	0.00	7	0.95				
MET	9	1.13	1	0.63	2	1.26	4	2.51	0	0.00	2	1.26					
PHE	17	0.98	2	0.57	>	8	2.30	5	1.44	0	0.00	2	0.57				
PRO	>	97	5.25	6	1.62	<	0	0.00	<	0	0.00	>	81	21.91	>	10	2.71
VAL	29	1.01	6	1.04	9	1.56	6	1.04	<	0	0.00	8	1.39				
ARG	16	0.78	>	9	2.19	3	0.73	1	0.24	<	0	0.00	3	0.73			
ASP	13	0.55	3	0.63	1	0.21	8	1.68	<	0	0.00	1	0.21				
GLU	13	0.47	4	0.73	3	0.55	4	0.73	<	0	0.00	2	0.36				
LYS	20	0.81	7	1.42	7	1.42	4	0.81	<	0	0.00	2	0.41				
ASN	19	1.04	6	1.65	2	0.55	5	1.37	1	0.27	5	1.37					

CYS	8	1.19	1	0.75	3	2.24	1	0.75	0	0.00	3	2.24			
GLN	18	1.13	3	0.94	2	0.63	> 9	2.83	0	0.00	4	1.26			
HIS	3	0.30	1	0.50	1	0.50	1	0.50	0	0.00	0	0.00			
SER	31	1.24	5	1.00	8	1.61	9	1.81	<	0	0.00	9	1.81		
THR	26	1.15	3	0.66	> 11	2.42	8	1.76	<	0	0.00	4	0.88		
TRP	3	0.49	0	0.00	0	0.00	3	2.43	0	0.00	0	0.00			
TYR	16	1.09	3	1.02	5	1.70	5	1.70	0	0.00	3	1.02			
GLY	9	0.32	6	1.06	<	1	0.18	<	0	0.00	<	0	0.00	2	0.35
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00			
ALL	410	1.00	82	1.00	82	1.00	82	1.00	82	1.00	82	1.00			

cluster 22: kink-1

aa	turn		position 1		position 2		position 3		position 4		position 5		
	f	P	f	P	f	P	aa	f	P	f	P	f	
ALA	19	1.29	2	0.68	> 7	2.38	1	0.34	2	0.68	> 7	2.38	
ILE	4	0.35	0	0.00	1	0.44	0	0.00	1	0.44	2	0.88	
LEU	15	0.85	4	1.14	7	1.99	2	0.57	2	0.57	<	0	0.00
MET	2	0.53	0	0.00	1	1.32	1	1.32	0	0.00	0	0.00	
PHE	4	0.48	1	0.61	1	0.61	0	0.00	1	0.61	1	0.61	
PRO	7	0.80	4	2.28	3	1.71	0	0.00	0	0.00	0	0.00	
VAL	2	0.15	1	0.37	0	0.00	0	0.00	1	0.37	0	0.00	
ARG	6	0.62	1	0.51	3	1.54	0	0.00	1	0.51	1	0.51	
ASP	15	1.32	4	1.77	0	0.00	0	0.00	> 8	3.53	3	1.32	
GLU	2	0.15	1	0.38	0	0.00	0	0.00	1	0.38	0	0.00	
LYS	13	1.11	1	0.43	3	1.28	4	1.70	3	1.28	2	0.85	
ASN	7	0.81	3	1.73	0	0.00	0	0.00	2	1.15	2	1.15	
CYS	5	1.57	0	0.00	> 3	4.71	0	0.00	2	3.14	0	0.00	
GLN	3	0.40	0	0.00	1	0.66	1	0.66	0	0.00	1	0.66	
HIS	3	0.63	> 3	3.16	0	0.00	0	0.00	0	0.00	0	0.00	
SER	18	1.52	2	0.84	5	2.11	1	0.42	> 8	3.37	2	0.84	
THR	11	1.02	> 5	2.32	1	0.46	0	0.00	2	0.93	3	1.39	
TRP	2	0.68	0	0.00	0	0.00	0	0.00	2	3.41	0	0.00	
TYR	7	1.00	2	1.43	1	0.72	1	0.72	0	0.00	3	2.15	
GLY	> 47	3.49	5	1.86	1	0.37	> 28	10.40	2	0.74	> 11	4.09	
OTH	3	4.24	0	0.00	> 1	7.07	0	0.00	> 1	7.07	> 1	7.07	
ALL	195	1.00	39	1.00	39	1.00	39	1.00	39	1.00	39	1.00	

13.4.3 reverse α -turns

cluster 1: turn-type Ia

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	< 144	0.67	< 20	0.46	> 65	1.50	< 24	0.56	< 7	0.16	< 28	0.65
ILE	< 33	0.20	< 0	0.00	< 10	0.30	< 3	0.09	< 0	0.00	< 20	0.60
LEU	< 93	0.36	45	0.87	< 16	0.31	< 10	0.19	< 4	0.08	< 18	0.35
MET	< 18	0.32	8	0.72	< 2	0.18	< 2	0.18	< 3	0.27	< 3	0.27
PHE	< 45	0.37	18	0.74	< 7	0.29	< 7	0.29	< 6	0.25	< 7	0.29
PRO	100	0.77	< 0	0.00	> 100	3.87	< 0	0.00	< 0	0.00	< 0	0.00
VAL	< 44	0.22	< 1	0.03	< 12	0.30	< 5	0.12	< 0	0.00	< 26	0.65
ARG	118	0.82	19	0.66	30	1.05	< 13	0.45	< 7	0.24	> 49	1.71
ASP	> 403	2.42	> 112	3.37	32	0.96	> 222	6.67	< 13	0.39	24	0.72
GLU	151	0.79	< 11	0.29	> 56	1.46	37	0.97	< 5	0.13	42	1.10
LYS	166	0.96	< 22	0.64	> 46	1.33	< 19	0.55	< 6	0.17	> 73	2.12
ASN	> 250	1.96	> 50	1.96	29	1.14	> 104	4.08	24	0.94	> 43	1.69
CYS	21	0.45	11	1.17	< 3	0.32	5	0.53	< 2	0.21	< 0	0.00
GLN	97	0.87	20	0.90	18	0.81	18	0.81	< 5	0.23	> 36	1.62
HIS	56	0.80	13	0.93	13	0.93	9	0.65	< 6	0.43	15	1.08
SER	> 258	1.48	> 71	2.04	> 64	1.84	> 47	1.35	< 4	0.12	> 72	2.07
THR	> 220	1.39	> 91	2.87	29	0.91	22	0.69	< 3	0.10	> 75	2.36
TRP	26	0.60	9	1.05	6	0.70	4	0.46	< 1	0.12	6	0.70
TYR	< 56	0.55	17	0.83	< 8	0.39	< 10	0.49	< 3	0.15	18	0.88
GLY	> 559	2.83	33	0.83	< 25	0.63	< 10	0.25	> 474	11.98	< 17	0.43
OTH	7	0.67	2	0.96	2	0.96	2	0.96	0	0.00	1	0.48
ALL	2865	1.00	573	1.00	573	1.00	573	1.00	573	1.00	573	1.00

cluster 2: turn-type Ib

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	< 45	0.51	< 0	0.00	> 29	1.64	10	0.57	< 1	0.06	< 5	0.28
ILE	< 11	0.16	< 0	0.00	< 6	0.44	< 0	0.00	< 0	0.00	< 5	0.37
LEU	< 22	0.21	< 1	0.05	< 6	0.28	< 6	0.28	< 2	0.10	< 7	0.33
MET	9	0.40	< 0	0.00	1	0.22	1	0.22	5	1.10	2	0.44
PHE	< 15	0.30	< 0	0.00	4	0.40	7	0.71	< 2	0.20	< 2	0.20
PRO	< 9	0.17	< 0	0.00	8	0.76	< 1	0.10	< 0	0.00	< 0	0.00
VAL	< 25	0.30	< 1	0.06	< 4	0.24	< 3	0.18	< 1	0.06	16	0.97
ARG	53	0.91	< 2	0.17	12	1.03	14	1.20	< 3	0.26	> 22	1.88
ASP	> 209	3.08	> 132	9.71	> 26	1.91	> 33	2.43	7	0.52	11	0.81
GLU	95	1.21	< 0	0.00	> 36	2.30	> 27	1.72	< 0	0.00	> 32	2.04
LYS	84	1.19	< 1	0.07	21	1.49	> 25	1.77	< 5	0.36	> 32	2.27
ASN	> 176	3.39	> 63	6.06	12	1.15	> 45	4.33	> 18	1.73	> 38	3.65
CYS	6	0.31	< 0	0.00	< 0	0.00	< 0	0.00	3	0.78	3	0.78
GLN	46	1.01	< 1	0.11	> 16	1.76	9	0.99	< 3	0.33	> 17	1.87
HIS	30	1.05	2	0.35	5	0.88	> 14	2.46	2	0.35	7	1.23
SER	61	0.86	< 5	0.35	> 24	1.69	19	1.34	< 4	0.28	9	0.63
THR	35	0.54	9	0.70	7	0.54	< 6	0.46	< 0	0.00	13	1.00

TRP	6 0.34	1 0.28	1 0.28	1 0.28	1 0.28	2 0.57
TYR	21 0.50	< 0 0.00	< 1 0.12	10 1.19	< 1 0.12	9 1.08
GLY	> 210 2.60	15 0.93	15 0.93	< 3 0.19	> 175 10.83	< 2 0.12
OTH	2 0.47	1 1.18	0 0.00	0 0.00	1 1.18	0 0.00
ALL	1170 1.00	234 1.00	234 1.00	234 1.00	234 1.00	234 1.00

cluster 3: turn-type Ic

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	6	1.22	2	2.04	> 3	3.06	0	0.00	0	0.00	1	1.02
ILE	1	0.26	1	1.32	0	0.00	0	0.00	0	0.00	0	0.00
LEU	1	0.17	0	0.00	0	0.00	1	0.85	0	0.00	0	0.00
MET	1	0.79	0	0.00	1	3.96	0	0.00	0	0.00	0	0.00
PHE	2	0.73	0	0.00	0	0.00	> 2	3.63	0	0.00	0	0.00
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
VAL	1	0.22	0	0.00	1	1.10	0	0.00	0	0.00	0	0.00
ARG	6	1.85	2	3.08	2	3.08	0	0.00	0	0.00	2	3.08
ASP	9	2.38	1	1.32	2	2.65	> 5	6.62	0	0.00	1	1.32
GLU	2	0.46	0	0.00	1	1.15	0	0.00	1	1.15	0	0.00
LYS	1	0.26	0	0.00	0	0.00	0	0.00	0	0.00	1	1.28
ASN	3	1.04	1	1.73	0	0.00	2	3.46	0	0.00	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	2	0.79	> 2	3.97	0	0.00	0	0.00	0	0.00	0	0.00
HIS	2	1.26	0	0.00	1	3.16	1	3.16	0	0.00	0	0.00
SER	10	2.53	2	2.53	2	2.53	1	1.26	2	2.53	> 3	3.80
THR	4	1.11	2	2.78	0	0.00	1	1.39	0	0.00	1	1.39
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	1	0.43	0	0.00	0	0.00	0	0.00	1	2.15	0	0.00
GLY	13	2.90	0	0.00	0	0.00	0	0.00	> 9	10.03	> 4	4.46
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	65	1.00	13	1.00	13	1.00	13	1.00	13	1.00	13	1.00

cluster 4: turn-type IIa

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	4	0.29	1	0.36	1	0.36	2	0.72	0	0.00	0	0.00
ILE	14	1.30	> 13	6.02	0	0.00	0	0.00	0	0.00	1	0.46
LEU	10	0.60	> 8	2.40	0	0.00	1	0.30	0	0.00	1	0.30
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	3	0.38	0	0.00	0	0.00	1	0.64	0	0.00	2	1.27
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
VAL	11	0.85	> 6	2.31	0	0.00	0	0.00	1	0.39	4	1.54
ARG	4	0.43	1	0.54	0	0.00	1	0.54	1	0.54	1	0.54
ASP	> 32	2.98	1	0.47	> 12	5.58	> 11	5.12	> 6	2.79	2	0.93
GLU	18	1.45	1	0.40	0	0.00	> 8	3.23	3	1.21	> 6	2.42
LYS	11	0.99	3	1.35	0	0.00	1	0.45	> 6	2.69	1	0.45
ASN	18	2.19	0	0.00	> 9	5.47	1	0.61	> 6	3.65	2	1.22
CYS	1	0.33	1	1.65	0	0.00	0	0.00	0	0.00	0	0.00
GLN	6	0.84	0	0.00	0	0.00	1	0.70	2	1.40	3	2.09

HIS	6 1.33	0 0.00	0 0.00	1 1.11	2 2.22	> 3 3.33
SER	17 1.51	0 0.00	> 10 4.45	4 1.78	2 0.89	1 0.45
THR	22 2.15	1 0.49	4 1.95	2 0.98	> 8 3.91	> 7 3.42
TRP	2 0.72	0 0.00	0 0.00	0 0.00	0 0.00	2 3.60
TYR	2 0.30	1 0.76	0 0.00	1 0.76	0 0.00	0 0.00
GLY	4 0.31	0 0.00	1 0.39	2 0.78	0 0.00	1 0.39
OTH	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
ALL	185 1.00	37 1.00	37 1.00	37 1.00	37 1.00	37 1.00

cluster 5: turn-type IIb

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	6	0.59	3	1.47	0	0.00	2	0.98	0	0.00	1	0.49
ILE	11	1.40	> 7	4.44	0	0.00	0	0.00	0	0.00	> 4	2.54
LEU	4	0.33	1	0.41	0	0.00	0	0.00	2	0.82	1	0.41
MET	1	0.38	0	0.00	0	0.00	0	0.00	0	0.00	1	1.91
PHE	6	1.05	2	1.75	0	0.00	0	0.00	1	0.87	3	2.62
PRO	9	1.48	0	0.00	0	0.00	> 9	7.39	0	0.00	0	0.00
VAL	7	0.74	> 5	2.64	0	0.00	0	0.00	0	0.00	2	1.06
ARG	5	0.74	2	1.48	2	1.48	0	0.00	0	0.00	1	0.74
ASP	13	1.66	0	0.00	> 5	3.19	> 5	3.19	3	1.91	0	0.00
GLU	14	1.55	2	1.11	2	1.11	> 6	3.32	1	0.55	3	1.66
LYS	4	0.49	1	0.62	1	0.62	0	0.00	2	1.23	0	0.00
ASN	7	1.17	0	0.00	0	0.00	0	0.00	> 6	5.00	1	0.83
CYS	2	0.91	1	2.27	0	0.00	0	0.00	0	0.00	1	2.27
GLN	2	0.38	0	0.00	0	0.00	1	0.96	1	0.96	0	0.00
HIS	5	1.52	1	1.52	1	1.52	0	0.00	1	1.52	2	3.04
SER	19	2.32	0	0.00	> 10	6.09	3	1.83	3	1.83	3	1.83
THR	13	1.74	2	1.34	> 4	2.68	1	0.67	3	2.01	3	2.01
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	4	0.83	0	0.00	0	0.00	0	0.00	> 3	3.11	1	1.03
GLY	3	0.32	0	0.00	2	1.07	0	0.00	1	0.54	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	135	1.00	27	1.00	27	1.00	27	1.00	27	1.00	27	1.00

cluster 6: turn-type IIIa

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	4	1.06	1	1.32	1	1.32	2	2.65	0	0.00	0	0.00
ILE	5	1.71	2	3.42	0	0.00	0	0.00	0	0.00	> 3	5.14
LEU	2	0.44	2	2.22	0	0.00	0	0.00	0	0.00	0	0.00
MET	1	1.03	0	0.00	0	0.00	0	0.00	1	5.15	0	0.00
PHE	2	0.94	0	0.00	0	0.00	1	2.36	0	0.00	1	2.36
PRO	1	0.44	0	0.00	0	0.00	1	2.22	0	0.00	0	0.00
VAL	4	1.14	2	2.85	0	0.00	0	0.00	0	0.00	2	2.85
ARG	1	0.40	0	0.00	0	0.00	0	0.00	0	0.00	1	2.00
ASP	2	0.69	0	0.00	1	1.72	0	0.00	1	1.72	0	0.00
GLU	4	1.20	0	0.00	1	1.49	> 3	4.48	0	0.00	0	0.00
LYS	5	1.66	1	1.66	1	1.66	2	3.32	0	0.00	1	1.66

ASN	3 1.35	0 0.00	0 0.00	1 2.25	> 2 4.50	0 0.00
CYS	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
GLN	3 1.55	1 2.58	1 2.58	0 0.00	1 2.58	0 0.00
HIS	3 2.46	1 4.11	0 0.00	0 0.00	1 4.11	1 4.11
SER	3 0.99	0 0.00	2 3.29	0 0.00	1 1.65	0 0.00
THR	1 0.36	0 0.00	0 0.00	0 0.00	0 0.00	1 1.81
TRP	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
TYR	1 0.56	0 0.00	0 0.00	0 0.00	1 2.79	0 0.00
GLY	5 1.45	0 0.00	> 3 4.35	0 0.00	2 2.90	0 0.00
OTH	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
ALL	50 1.00	10 1.00	10 1.00	10 1.00	10 1.00	10 1.00

cluster 7: turn-type IIIb

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	9	1.83	2	2.04	> 5	5.09	2	2.04	0	0.00	0	0.00
ILE	2	0.53	0	0.00	0	0.00	1	1.32	0	0.00	1	1.32
LEU	3	0.51	1	0.85	0	0.00	1	0.85	0	0.00	1	0.85
MET	1	0.79	0	0.00	0	0.00	0	0.00	0	0.00	1	3.96
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PRO	3	1.02	0	0.00	0	0.00	> 3	5.12	0	0.00	0	0.00
VAL	7	1.53	2	2.19	1	1.10	1	1.10	0	0.00	> 3	3.29
ARG	2	0.62	1	1.54	1	1.54	0	0.00	0	0.00	0	0.00
ASP	3	0.80	0	0.00	0	0.00	1	1.32	1	1.32	1	1.32
GLU	6	1.38	2	2.30	0	0.00	2	2.30	1	1.15	1	1.15
LYS	5	1.28	2	2.55	0	0.00	2	2.55	0	0.00	1	1.28
ASN	2	0.69	0	0.00	0	0.00	0	0.00	1	1.73	1	1.73
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	2	0.79	0	0.00	> 2	3.97	0	0.00	0	0.00	0	0.00
HIS	2	1.26	1	3.16	0	0.00	0	0.00	1	3.16	0	0.00
SER	5	1.26	1	1.26	> 3	3.80	0	0.00	0	0.00	1	1.26
THR	2	0.56	1	1.39	0	0.00	0	0.00	0	0.00	1	1.39
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	2	0.86	0	0.00	1	2.15	0	0.00	0	0.00	1	2.15
GLY	9	2.01	0	0.00	0	0.00	0	0.00	> 9	10.03	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	65	1.00	13	1.00	13	1.00	13	1.00	13	1.00	13	1.00

cluster 8: turn-type IVa

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	6	0.51	2	0.85	3	1.28	0	0.00	0	0.00	1	0.43
ILE	3	0.33	2	1.11	1	0.55	0	0.00	0	0.00	0	0.00
LEU	9	0.64	> 6	2.15	2	0.72	0	0.00	0	0.00	1	0.36
MET	2	0.66	1	1.66	0	0.00	0	0.00	0	0.00	1	1.66
PHE	7	1.07	3	2.28	1	0.76	0	0.00	1	0.76	2	1.52
PRO	5	0.72	0	0.00	> 5	3.58	0	0.00	0	0.00	0	0.00
VAL	3	0.28	0	0.00	1	0.46	0	0.00	1	0.46	1	0.46
ARG	18	2.32	> 4	2.58	3	1.94	0	0.00	2	1.29	> 9	5.81

ASP	10	1.11	3	1.67	2	1.11	4	2.22	1	0.56	0	0.00
GLU	9	0.87	0	0.00	4	1.93	2	0.96	2	0.96	1	0.48
LYS	10	1.07	2	1.07	3	1.61	0	0.00	0	0.00	> 5	2.68
ASN	11	1.60	0	0.00	0	0.00	> 6	4.36	> 5	3.63	0	0.00
CYS	1	0.40	1	1.97	0	0.00	0	0.00	0	0.00	0	0.00
GLN	5	0.83	2	1.67	0	0.00	1	0.83	0	0.00	2	1.67
HIS	4	1.06	0	0.00	2	2.65	1	1.32	0	0.00	1	1.32
SER	6	0.64	0	0.00	3	1.59	0	0.00	1	0.53	2	1.06
THR	2	0.23	0	0.00	0	0.00	0	0.00	0	0.00	2	1.17
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	4	0.72	1	0.90	0	0.00	0	0.00	3	2.70	0	0.00
GLY	> 40	3.74	4	1.87	1	0.47	> 17	7.94	> 15	7.01	3	1.40
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	155	1.00	31	1.00	31	1.00	31	1.00	31	1.00	31	1.00

cluster 9: turn-type IVb

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	8	1.01	0	0.00	3	1.89	2	1.26	1	0.63	2	1.26
ILE	3	0.49	2	1.63	0	0.00	0	0.00	0	0.00	1	0.82
LEU	4	0.42	0	0.00	1	0.53	0	0.00	1	0.53	2	1.06
MET	1	0.49	0	0.00	0	0.00	0	0.00	0	0.00	1	2.45
PHE	3	0.67	0	0.00	0	0.00	1	1.12	0	0.00	2	2.25
PRO	2	0.42	0	0.00	2	2.11	0	0.00	0	0.00	0	0.00
VAL	2	0.27	1	0.68	0	0.00	0	0.00	0	0.00	1	0.68
ARG	3	0.57	0	0.00	2	1.91	0	0.00	0	0.00	1	0.95
ASP	9	1.48	1	0.82	0	0.00	> 5	4.10	2	1.64	1	0.82
GLU	3	0.43	0	0.00	2	1.42	0	0.00	0	0.00	1	0.71
LYS	9	1.42	1	0.79	> 5	3.95	3	2.37	0	0.00	0	0.00
ASN	8	1.71	0	0.00	2	2.14	> 3	3.22	1	1.07	2	2.14
CYS	1	0.58	0	0.00	0	0.00	1	2.91	0	0.00	0	0.00
GLN	3	0.74	1	1.23	0	0.00	1	1.23	0	0.00	1	1.23
HIS	2	0.78	0	0.00	1	1.96	1	1.96	0	0.00	0	0.00
SER	6	0.94	3	2.35	0	0.00	0	0.00	1	0.78	2	1.57
THR	4	0.69	1	0.86	1	0.86	0	0.00	1	0.86	1	0.86
TRP	3	1.90	1	3.17	0	0.00	0	0.00	0	0.00	> 2	6.33
TYR	3	0.80	0	0.00	0	0.00	1	1.33	1	1.33	1	1.33
GLY	> 28	3.86	> 10	6.90	2	1.38	3	2.07	> 13	8.97	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	105	1.00	21	1.00	21	1.00	21	1.00	21	1.00	21	1.00

cluster 10: turn-type V

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	3	0.20	0	0.00	2	0.66	0	0.00	0	0.00	1	0.33
ILE	14	1.20	> 8	3.42	0	0.00	0	0.00	1	0.43	5	2.14
LEU	14	0.78	6	1.66	1	0.28	< 0	0.00	4	1.11	3	0.83
MET	8	2.06	2	2.57	2	2.57	2	2.57	1	1.29	1	1.29
PHE	5	0.59	2	1.18	1	0.59	0	0.00	0	0.00	2	1.18

PRO	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
VAL	17 1.21	> 10 3.56	0 0.00	0 0.00	4 1.43	3 1.07	
ARG	11 1.10	2 1.00	4 2.00	1 0.50	3 1.50	1 0.50	
ASP	24 2.07	0 0.00	> 12 5.17	> 6 2.58	4 1.72	2 0.86	
GLU	9 0.67	1 0.37	2 0.75	2 0.75	2 0.75	2 0.75	
LYS	14 1.16	3 1.25	0 0.00	2 0.83	> 7 2.91	2 0.83	
ASN	20 2.25	1 0.56	> 8 4.50	> 5 2.81	4 2.25	2 1.13	
CYS	1 0.31	0 0.00	0 0.00	0 0.00	0 0.00	1 1.53	
GLN	8 1.03	0 0.00	0 0.00	3 1.94	2 1.29	3 1.94	
HIS	5 1.03	1 1.03	2 2.05	0 0.00	1 1.03	1 1.03	
SER	14 1.15	1 0.41	3 1.23	4 1.65	5 2.06	1 0.41	
THR	5 0.45	1 0.45	0 0.00	0 0.00	2 0.90	2 0.90	
TRP	1 0.33	0 0.00	0 0.00	0 0.00	0 0.00	1 1.66	
TYR	9 1.26	2 1.40	2 1.40	0 0.00	0 0.00	> 5 3.49	
GLY	18 1.30	0 0.00	1 0.36	> 15 5.43	0 0.00	2 0.72	
OTH	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	
ALL	200 1.00	40 1.00	40 1.00	40 1.00	40 1.00	40 1.00	40 1.00

cluster 11: turn-type VIa

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	4	0.71	0	0.00	3	2.65	0	0.00	1	0.88	0	0.00
ILE	3	0.69	> 3	3.42	0	0.00	0	0.00	0	0.00	0	0.00
LEU	2	0.30	0	0.00	0	0.00	1	0.74	1	0.74	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	2	0.63	0	0.00	1	1.57	0	0.00	1	1.57	0	0.00
PRO	6	1.78	0	0.00	> 6	8.87	0	0.00	0	0.00	0	0.00
VAL	6	1.14	2	1.90	0	0.00	0	0.00	2	1.90	2	1.90
ARG	4	1.07	1	1.33	0	0.00	0	0.00	1	1.33	2	2.67
ASP	3	0.69	0	0.00	1	1.15	2	2.30	0	0.00	0	0.00
GLU	5	1.00	1	1.00	1	1.00	1	1.00	2	1.99	0	0.00
LYS	9	1.99	> 3	3.32	1	1.11	0	0.00	> 3	3.32	2	2.21
ASN	5	1.50	0	0.00	0	0.00	2	3.00	1	1.50	2	3.00
CYS	1	0.82	1	4.08	0	0.00	0	0.00	0	0.00	0	0.00
GLN	3	1.03	2	3.44	1	1.72	0	0.00	0	0.00	0	0.00
HIS	1	0.55	0	0.00	1	2.74	0	0.00	0	0.00	0	0.00
SER	3	0.66	0	0.00	0	0.00	0	0.00	0	0.00	> 3	3.29
THR	5	1.20	1	1.20	0	0.00	0	0.00	1	1.20	> 3	3.61
TRP	2	1.77	1	4.43	0	0.00	0	0.00	1	4.43	0	0.00
TYR	3	1.12	0	0.00	0	0.00	1	1.86	1	1.86	1	1.86
GLY	8	1.55	0	0.00	0	0.00	> 8	7.73	0	0.00	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	75	1.00	15	1.00	15	1.00	15	1.00	15	1.00	15	1.00

cluster 12: turn-type VIb

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ILE	9	1.62	> 4	3.61	0	0.00	0	0.00	> 5	4.51	0	0.00

LEU	4	0.47	2	1.17	0	0.00	0	0.00	2	1.17	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	3	0.75	0	0.00	1	1.24	1	1.24	1	1.24	0	0.00
PRO	3	0.70	0	0.00	> 3	3.50	0	0.00	0	0.00	0	0.00
VAL	10	1.50	> 6	4.50	2	1.50	0	0.00	2	1.50	0	0.00
ARG	2	0.42	1	1.05	1	1.05	0	0.00	0	0.00	0	0.00
ASP	4	0.73	0	0.00	2	1.81	2	1.81	0	0.00	0	0.00
GLU	5	0.79	1	0.79	> 4	3.15	0	0.00	0	0.00	0	0.00
LYS	8	1.40	1	0.87	3	2.62	0	0.00	> 4	3.49	0	0.00
ASN	2	0.47	0	0.00	0	0.00	2	2.37	0	0.00	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	4	1.09	0	0.00	1	1.36	1	1.36	2	2.72	0	0.00
HIS	4	1.73	0	0.00	1	2.16	1	2.16	> 2	4.32	0	0.00
SER	1	0.17	1	0.87	0	0.00	0	0.00	0	0.00	0	0.00
THR	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TRP	1	0.70	1	3.50	0	0.00	0	0.00	0	0.00	0	0.00
TYR	3	0.88	2	2.94	1	1.47	0	0.00	0	0.00	0	0.00
GLY	> 32	4.88	0	0.00	0	0.00	> 12	9.15	1	0.76	> 19	14.48
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	95	1.00	19	1.00	19	1.00	19	1.00	19	1.00	19	1.00

cluster 13: turn-type VII

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	7	0.38	1	0.27	2	0.54	2	0.54	1	0.27	1	0.27
ILE	8	0.56	1	0.35	3	1.05	3	1.05	1	0.35	0	0.00
LEU	8	0.36	< 0	0.00	< 0	0.00	2	0.45	6	1.36	< 0	0.00
MET	2	0.42	0	0.00	1	1.05	0	0.00	1	1.05	0	0.00
PHE	10	0.96	0	0.00	0	0.00	4	1.93	> 6	2.89	0	0.00
PRO	17	1.54	0	0.00	> 16	7.24	1	0.45	0	0.00	0	0.00
VAL	5	0.29	1	0.29	2	0.58	1	0.29	0	0.00	1	0.29
ARG	11	0.90	0	0.00	2	0.82	> 6	2.45	3	1.23	0	0.00
ASP	22	1.55	> 13	4.57	2	0.70	5	1.76	2	0.70	0	0.00
GLU	14	0.85	2	0.61	5	1.53	5	1.53	2	0.61	0	0.00
LYS	17	1.15	0	0.00	4	1.36	> 7	2.37	5	1.69	1	0.34
ASN	13	1.19	5	2.30	4	1.84	3	1.38	0	0.00	1	0.46
CYS	8	2.00	> 5	6.24	0	0.00	0	0.00	> 3	3.75	0	0.00
GLN	5	0.53	1	0.53	1	0.53	1	0.53	2	1.05	0	0.00
HIS	11	1.84	> 5	4.19	1	0.84	0	0.00	> 5	4.19	0	0.00
SER	10	0.67	5	1.68	1	0.34	2	0.67	2	0.67	0	0.00
THR	14	1.03	> 7	2.58	2	0.74	4	1.47	1	0.37	0	0.00
TRP	2	0.54	0	0.00	1	1.36	0	0.00	1	1.36	0	0.00
TYR	9	1.03	0	0.00	1	0.57	2	1.14	> 6	3.42	0	0.00
GLY	> 52	3.07	3	0.89	1	0.30	1	0.30	2	0.59	> 45	13.30
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	245	1.00	49	1.00	49	1.00	49	1.00	49	1.00	49	1.00

TYR	3	1.05	2	3.49	0	0.00	1	1.75	0	0.00	0	0.00
GLY	> 17	3.08	0	0.00	> 12	10.86	0	0.00	1	0.91	> 4	3.62
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	80	1.00	16	1.00	16	1.00	16	1.00	16	1.00	16	1.00

cluster 18: turn-type XII

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ILE	3	1.14	0	0.00	0	0.00	0	0.00	0	0.00	> 3	5.71
LEU	1	0.25	0	0.00	0	0.00	1	1.23	0	0.00	0	0.00
MET	1	1.14	0	0.00	0	0.00	> 1	5.72	0	0.00	0	0.00
PHE	3	1.57	0	0.00	0	0.00	0	0.00	0	0.00	> 3	7.86
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
VAL	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ARG	2	0.89	0	0.00	0	0.00	0	0.00	1	2.22	1	2.22
ASP	9	3.44	> 7	13.39	0	0.00	1	1.91	1	1.91	0	0.00
GLU	1	0.33	0	0.00	0	0.00	1	1.66	0	0.00	0	0.00
LYS	2	0.74	0	0.00	0	0.00	> 2	3.69	0	0.00	0	0.00
ASN	6	3.00	> 2	5.00	1	2.50	0	0.00	> 3	7.50	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	2	1.15	0	0.00	0	0.00	1	2.87	0	0.00	1	2.87
HIS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
SER	3	1.10	0	0.00	0	0.00	1	1.83	> 2	3.66	0	0.00
THR	3	1.20	0	0.00	0	0.00	1	2.01	> 2	4.01	0	0.00
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	6	3.73	0	0.00	> 5	15.52	0	0.00	0	0.00	1	3.11
GLY	3	0.97	0	0.00	> 3	4.83	0	0.00	0	0.00	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	45	1.00	9	1.00	9	1.00	9	1.00	9	1.00	9	1.00

cluster 19: turn-type XIII

aa	turn		position 1		position 2		position 3		position 4		position 5	
	f	P	f	P	f	P	f	P	f	P	f	P
ALA	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ILE	1	0.57	0	0.00	0	0.00	0	0.00	0	0.00	1	2.85
LEU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
MET	1	1.72	0	0.00	0	0.00	0	0.00	0	0.00	> 1	8.58
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
VAL	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ARG	3	2.00	0	0.00	1	3.33	> 2	6.67	0	0.00	0	0.00
ASP	1	0.57	0	0.00	0	0.00	1	2.87	0	0.00	0	0.00
GLU	1	0.50	0	0.00	0	0.00	0	0.00	0	0.00	1	2.49
LYS	1	0.55	0	0.00	0	0.00	1	2.77	0	0.00	0	0.00
ASN	3	2.25	1	3.75	1	3.75	0	0.00	0	0.00	1	3.75
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	3	2.58	0	0.00	> 2	8.60	0	0.00	1	4.30	0	0.00
HIS	1	1.37	0	0.00	0	0.00	0	0.00	0	0.00	> 1	6.84

13.5 six residue turns

13.5.1 normal π -turns

cluster 1: turn-type I

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	P	f	P	f	P	f	P
ALA	> 1867	1.38	> 557	2.47	> 291	1.29	> 385	1.71	> 352	1.56	< 46	0.20	236	1.05
ILE	< 826	0.79	< 126	0.72	> 210	1.21	< 74	0.43	< 85	0.49	< 0	0.00	> 331	1.90
LEU	> 2106	1.31	> 682	2.54	260	0.97	< 199	0.74	> 477	1.77	< 20	0.07	> 468	1.74
MET	405	1.17	> 120	2.07	54	0.93	58	1.00	> 92	1.59	< 10	0.17	71	1.23
PHE	< 575	0.76	147	1.16	< 83	0.66	< 50	0.40	< 100	0.79	< 12	0.10	> 183	1.45
PRO	< 137	0.17	< 38	0.28	< 67	0.50	< 23	0.17	< 5	0.04	< 4	0.03	< 0	0.00
VAL	< 852	0.68	< 126	0.60	193	0.92	< 109	0.52	< 93	0.44	< 4	0.02	> 327	1.56
ARG	> 1039	1.16	< 104	0.70	> 252	1.69	> 247	1.66	> 205	1.38	< 107	0.72	< 124	0.83
ASP	< 818	0.79	< 140	0.81	< 128	0.74	197	1.14	< 90	0.52	168	0.97	< 95	0.55
GLU	1193	1.00	< 61	0.31	> 286	1.43	> 446	2.24	188	0.94	< 91	0.46	< 121	0.61
LYS	> 1341	1.24	< 72	0.40	> 343	1.91	> 346	1.93	> 213	1.19	186	1.03	181	1.01
ASN	915	1.15	< 96	0.72	< 84	0.63	127	0.96	> 236	1.78	> 303	2.29	< 69	0.52
CYS	275	0.94	> 111	2.28	< 29	0.60	< 20	0.41	42	0.86	< 18	0.37	55	1.13
GLN	782	1.13	< 74	0.64	> 170	1.47	> 201	1.74	> 167	1.45	98	0.85	< 72	0.62
HIS	433	0.99	< 47	0.65	86	1.18	< 53	0.73	> 128	1.76	68	0.94	< 51	0.70
SER	< 807	0.74	< 150	0.83	< 144	0.79	199	1.10	164	0.91	< 63	0.35	< 87	0.48
THR	< 509	0.51	< 56	0.34	< 118	0.72	< 89	0.54	< 123	0.75	< 11	0.07	< 112	0.68
TRP	< 179	0.67	49	1.09	39	0.87	< 15	0.34	< 28	0.62	< 2	0.05	46	1.03
TYR	< 516	0.81	95	0.89	< 80	0.75	< 57	0.53	> 129	1.21	< 7	0.07	> 148	1.39
GLY	> 2244	1.82	< 107	0.52	< 52	0.25	< 80	0.39	< 50	0.24	> 1760	8.55	195	0.95
OTH	73	1.12	> 24	2.22	13	1.20	7	0.65	15	1.39	< 4	0.37	10	0.92
ALL	17892	1.00	2982	1.00	2982	1.00	2982	1.00	2982	1.00	2982	1.00	2982	1.00

cluster 2: turn-type II

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	P	f	P	f	P	f	P
ALA	270	0.87	49	0.94	56	1.08	41	0.79	< 29	0.56	< 37	0.71	58	1.12
ILE	271	1.12	> 60	1.49	44	1.09	31	0.77	39	0.97	> 61	1.52	36	0.90
LEU	432	1.16	75	1.21	73	1.18	< 47	0.76	< 47	0.76	> 88	1.42	> 102	1.64
MET	75	0.93	19	1.42	15	1.12	10	0.75	< 6	0.45	8	0.60	17	1.27
PHE	229	1.31	> 59	2.02	29	0.99	19	0.65	32	1.10	36	1.23	> 54	1.85
PRO	< 23	0.12	< 9	0.29	< 12	0.39	< 1	0.03	< 0	0.00	< 1	0.03	< 0	0.00
VAL	336	1.16	45	0.93	52	1.08	44	0.91	55	1.14	> 90	1.86	50	1.03
ARG	217	1.05	30	0.87	> 52	1.51	> 57	1.66	27	0.78	24	0.70	27	0.78
ASP	238	0.99	38	0.95	37	0.93	48	1.20	52	1.30	39	0.97	< 24	0.60
GLU	289	1.05	35	0.76	58	1.26	> 69	1.50	> 60	1.30	44	0.95	< 23	0.50
LYS	215	0.86	< 15	0.36	37	0.89	> 61	1.47	41	0.99	30	0.72	31	0.75
ASN	194	1.06	< 17	0.56	31	1.01	> 50	1.63	> 54	1.76	< 13	0.43	29	0.95
CYS	72	1.07	14	1.24	8	0.71	8	0.71	10	0.89	17	1.51	15	1.33
GLN	127	0.79	25	0.94	19	0.71	22	0.82	24	0.90	18	0.67	19	0.71

HIS	129	1.28	>	27	1.61	22	1.31	19	1.13	>	26	1.55	18	1.07	17	1.01				
SER	196	0.78		39	0.93	33	0.79	44	1.05		35	0.84	<	24	0.57	<	21	0.50		
THR	223	0.97		30	0.79	<	26	0.68	>	55	1.44	>	54	1.42	31	0.81	27	0.71		
TRP	87	1.40	>	24	2.32	>	20	1.93		9	0.87		10	0.97	12	1.16	12	1.16		
TYR	>	254	1.72	>	59	2.39	>	40	1.62		27	1.10	>	51	2.07	>	46	1.87	31	1.26
GLY	234	0.82	<	18	0.38	<	23	0.48	<	26	0.55	<	33	0.69	46	0.97	>	88	1.85	
OTH	23	1.53		2	0.80		2	0.80		1	0.40		4	1.60	>	6	2.40	>	8	3.20
ALL	4134	1.00		689	1.00		689	1.00		689	1.00		689	1.00		689	1.00		689	1.00

cluster 3: turn-type III

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6				
	f	P	f	P	f	P	f	f	P	f	P	f	P	f			
ALA	10	0.92	4	2.21	2	1.10	1	0.55	0	0.00	2	1.10		1	0.55		
ILE	7	0.83	1	0.71	0	0.00	0	0.00	0	0.00	2	1.43	>	4	2.85		
LEU	11	0.85	4	1.85	3	1.39	3	1.39	0	0.00	1	0.46		0	0.00		
MET	2	0.72	1	2.15	0	0.00	0	0.00	0	0.00	1	2.15		0	0.00		
PHE	2	0.33	0	0.00	0	0.00	0	0.00	1	0.98	0	0.00		1	0.98		
PRO	1	0.15	1	0.92	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00		
VAL	8	0.79	0	0.00	1	0.59	1	0.59	0	0.00	1	0.59	>	5	2.97		
ARG	9	1.25	0	0.00	2	1.67	0	0.00	0	0.00	>	5	4.17	2	1.67		
ASP	9	1.08	0	0.00	>	5	3.59	3	2.15	0	0.00	0	0.00	1	0.72		
GLU	8	0.83	1	0.62	0	0.00	0	0.00	1	0.62	2	1.25		4	2.49		
LYS	8	0.92	0	0.00	1	0.69	2	1.38	0	0.00	>	4	2.77	1	0.69		
ASN	10	1.56	1	0.94	2	1.88	>	4	3.75	2	1.88	1	0.94	0	0.00		
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00		
GLN	6	1.08	1	1.08	0	0.00	0	0.00	1	1.08	2	2.15		2	2.15		
HIS	1	0.28	0	0.00	1	1.71	0	0.00	0	0.00	0	0.00		0	0.00		
SER	7	0.80	3	2.06	1	0.69	0	0.00	1	0.69	1	0.69		1	0.69		
THR	7	0.88	>	4	3.01	0	0.00	0	0.00	1	0.75	2	1.51	0	0.00		
TRP	3	1.39	0	0.00	1	2.77	>	2	5.54	0	0.00	0	0.00	0	0.00		
TYR	6	1.16	1	1.16	2	2.33	1	1.16	2	2.33	0	0.00		0	0.00		
GLY	>	29	2.92	2	1.21	3	1.81	>	7	4.22	>	15	9.05	0	0.00	2	1.21
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00		
ALL	144	1.00	24	1.00	24	1.00	24	1.00	24	1.00	24	1.00		24	1.00		

cluster 4: turn-type IV

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6			
	f	P	f	P	f	P	f	f	P	f	P	f	P	f		
ALA	9	1.10	2	1.47	0	0.00	1	0.74	2	1.47	2	1.47		2	1.47	
ILE	7	1.11	1	0.95	1	0.95	0	0.00	0	0.00	0	0.00	>	5	4.76	
LEU	9	0.92	3	1.85	2	1.23	0	0.00	1	0.62	2	1.23		1	0.62	
MET	4	1.91	1	2.86	>	2	5.72	0	0.00	0	0.00	1	2.86	0	0.00	
PHE	3	0.66	0	0.00	1	1.31	0	0.00	0	0.00	2	2.62		0	0.00	
PRO	3	0.62	0	0.00	>	3	3.70	0	0.00	0	0.00	0	0.00	0	0.00	
VAL	9	1.19	2	1.58	0	0.00	1	0.79	2	1.58	3	2.37		1	0.79	
ARG	1	0.19	0	0.00	0	0.00	0	0.00	0	0.00	1	1.11		0	0.00	
ASP	10	1.59	0	0.00	>	4	3.83	1	0.96	1	0.96	>	3	2.87	1	0.96
GLU	3	0.42	2	1.66	0	0.00	0	0.00	0	0.00	0	0.00		1	0.83	
LYS	4	0.62	0	0.00	2	1.84	0	0.00	1	0.92	1	0.92		0	0.00	

ASN	5	1.04	1	1.25	0	0.00	2	2.50	1	1.25	0	0.00	1	1.25
CYS	1	0.57	1	3.40	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	4	0.96	0	0.00	2	2.87	1	1.43	0	0.00	0	0.00	1	1.43
HIS	1	0.38	0	0.00	0	0.00	1	2.28	0	0.00	0	0.00	0	0.00
SER	4	0.61	0	0.00	0	0.00	1	0.91	1	0.91	1	0.91	1	0.91
THR	9	1.51	> 3	3.01	0	0.00	0	0.00	> 5	5.02	1	1.00	0	0.00
TRP	3	1.85	1	3.70	0	0.00	0	0.00	1	3.70	1	3.70	0	0.00
TYR	5	1.29	0	0.00	1	1.55	1	1.55	0	0.00	0	0.00	> 3	4.66
GLY	14	1.88	1	0.81	0	0.00	> 9	7.24	3	2.41	0	0.00	1	0.81
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	108	1.00	18	1.00	18	1.00	18	1.00	18	1.00	18	1.00	18	1.00

cluster 5: turn-type Va

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	2	0.49	1	1.47	0	0.00	1	1.47	0	0.00	0	0.00	0	0.00
ILE	1	0.32	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	1.90
LEU	3	0.62	0	0.00	0	0.00	0	0.00	1	1.23	1	1.23	1	1.23
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	1	0.44	0	0.00	1	2.62	0	0.00	0	0.00	0	0.00	0	0.00
PRO	1	0.41	0	0.00	0	0.00	1	2.47	0	0.00	0	0.00	0	0.00
VAL	6	1.58	> 4	6.33	0	0.00	0	0.00	1	1.58	1	1.58	0	0.00
ARG	1	0.37	0	0.00	0	0.00	0	0.00	1	2.22	0	0.00	0	0.00
ASP	4	1.27	0	0.00	0	0.00	1	1.91	1	1.91	1	1.91	1	1.91
GLU	1	0.28	0	0.00	0	0.00	1	1.66	0	0.00	0	0.00	0	0.00
LYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASN	4	1.67	0	0.00	0	0.00	0	0.00	> 3	7.50	1	2.50	0	0.00
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
GLN	4	1.91	0	0.00	0	0.00	> 2	5.73	0	0.00	0	0.00	> 2	5.73
HIS	3	2.28	1	4.56	0	0.00	1	4.56	1	4.56	0	0.00	0	0.00
SER	4	1.22	0	0.00	0	0.00	1	1.83	1	1.83	1	1.83	1	1.83
THR	1	0.33	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	2.01
TRP	1	1.23	0	0.00	0	0.00	> 1	7.39	0	0.00	0	0.00	0	0.00
TYR	4	2.07	> 2	6.21	0	0.00	0	0.00	0	0.00	0	0.00	> 2	6.21
GLY	> 13	3.49	1	1.61	> 8	12.87	0	0.00	0	0.00	> 4	6.44	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	54	1.00	9	1.00	9	1.00	9	1.00	9	1.00	9	1.00	9	1.00

cluster 6: turn-type Vb

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	3	1.32	> 2	5.30	0	0.00	0	0.00	0	0.00	0	0.00	1	2.65
ILE	1	0.57	0	0.00	0	0.00	1	3.42	0	0.00	0	0.00	0	0.00
LEU	2	0.74	0	0.00	0	0.00	0	0.00	0	0.00	1	2.22	1	2.22
MET	1	1.72	> 1	10.30	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	1	0.79	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	4.72
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
VAL	1	0.48	1	2.85	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ARG	2	1.33	0	0.00	0	0.00	1	4.00	0	0.00	1	4.00	0	0.00

ASP	1 0.57	0 0.00	0 0.00	0 0.00	1 3.44	0 0.00	0 0.00
GLU	2 1.00	0 0.00	0 0.00	1 2.99	0 0.00	0 0.00	1 2.99
LYS	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
ASN	2 1.50	1 4.50	0 0.00	0 0.00	1 4.50	0 0.00	0 0.00
CYS	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
GLN	3 2.58	0 0.00	0 0.00	1 5.16	0 0.00	> 2 10.32	0 0.00
HIS	1 1.37	0 0.00	0 0.00	0 0.00	> 1 8.21	0 0.00	0 0.00
SER	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
THR	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
TRP	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
TYR	2 1.86	0 0.00	0 0.00	> 1 5.59	> 1 5.59	0 0.00	0 0.00
GLY	8 3.86	0 0.00	> 5 14.48	0 0.00	1 2.90	1 2.90	1 2.90
OTH	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
ALL	30 1.00	5 1.00	5 1.00	5 1.00	5 1.00	5 1.00	5 1.00

cluster 7: turn-type VI

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	2	0.55	1	1.66	1	1.66	0	0.00	0	0.00	0	0.00	0	0.00
ILE	1	0.36	0	0.00	1	2.14	0	0.00	0	0.00	0	0.00	0	0.00
LEU	6	1.39	> 4	5.55	0	0.00	0	0.00	0	0.00	0	0.00	2	2.77
MET	1	1.07	> 1	6.44	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	2	0.98	0	0.00	0	0.00	0	0.00	0	0.00	1	2.95	1	2.95
PRO	1	0.46	0	0.00	0	0.00	0	0.00	0	0.00	1	2.77	0	0.00
VAL	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ARG	2	0.83	0	0.00	0	0.00	> 2	5.00	0	0.00	0	0.00	0	0.00
ASP	2	0.72	0	0.00	0	0.00	1	2.15	0	0.00	0	0.00	1	2.15
GLU	2	0.62	0	0.00	1	1.87	0	0.00	0	0.00	1	1.87	0	0.00
LYS	1	0.35	0	0.00	1	2.08	0	0.00	0	0.00	0	0.00	0	0.00
ASN	4	1.88	0	0.00	> 2	5.63	> 2	5.63	0	0.00	0	0.00	0	0.00
CYS	1	1.27	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	> 1	7.65
GLN	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
HIS	4	3.42	0	0.00	> 2	10.26	1	5.13	1	5.13	0	0.00	0	0.00
SER	1	0.34	1	2.06	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
THR	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TRP	2	2.77	0	0.00	0	0.00	0	0.00	0	0.00	> 1	8.31	> 1	8.31
TYR	2	1.16	0	0.00	0	0.00	1	3.49	0	0.00	1	3.49	0	0.00
GLY	> 13	3.92	1	1.81	0	0.00	0	0.00	> 7	12.67	> 3	5.43	> 2	3.62
OTH	1	5.74	0	0.00	0	0.00	> 1	34.44	0	0.00	0	0.00	0	0.00
ALL	48	1.00	8	1.00	8	1.00	8	1.00	8	1.00	8	1.00	8	1.00

cluster 8: turn-type VII

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	4	1.26	1	1.89	> 2	3.78	0	0.00	0	0.00	1	1.89	0	0.00
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LEU	3	0.79	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	> 3	4.75
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	1	0.56	0	0.00	1	3.37	0	0.00	0	0.00	0	0.00	0	0.00

13.5.2 open π -turns

cluster 1: turn-type I

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	126	1.10	> 35	1.83	26	1.36	> 30	1.57	< 9	0.47	< 7	0.37	19	0.99
ILE	61	0.69	11	0.74	12	0.81	12	0.81	< 6	0.41	< 1	0.07	19	1.29
LEU	113	0.83	18	0.79	20	0.88	28	1.23	30	1.32	< 0	0.00	17	0.75
MET	28	0.95	4	0.81	3	0.61	5	1.02	3	0.61	< 0	0.00	> 13	2.65
PHE	54	0.84	5	0.47	12	1.12	7	0.65	17	1.59	< 1	0.09	12	1.12
PRO	< 15	0.22	< 1	0.09	11	0.96	< 3	0.26	< 0	0.00	< 0	0.00	< 0	0.00
VAL	< 57	0.54	12	0.68	11	0.62	11	0.62	< 6	0.34	< 0	0.00	17	0.96
ARG	80	1.05	13	1.03	> 23	1.82	14	1.11	11	0.87	< 2	0.16	17	1.34
ASP	85	0.96	> 29	1.97	16	1.09	17	1.16	< 6	0.41	< 7	0.48	10	0.68
GLU	96	0.95	16	0.95	> 26	1.54	> 29	1.71	< 9	0.53	< 1	0.06	15	0.89
LYS	102	1.12	16	1.05	> 25	1.64	> 27	1.77	12	0.79	< 4	0.26	18	1.18
ASN	74	1.10	> 19	1.69	11	0.98	11	0.98	10	0.89	14	1.25	9	0.80
CYS	16	0.65	8	1.94	1	0.24	2	0.48	1	0.24	1	0.24	3	0.73
GLN	53	0.90	12	1.22	14	1.43	8	0.82	4	0.41	5	0.51	10	1.02
HIS	30	0.81	> 11	1.79	3	0.49	6	0.97	4	0.65	< 1	0.16	5	0.81
SER	88	0.95	17	1.11	9	0.59	14	0.91	> 24	1.56	10	0.65	14	0.91
THR	> 128	1.52	8	0.57	10	0.71	17	1.21	> 77	5.50	< 1	0.07	15	1.07
TRP	13	0.57	1	0.26	2	0.53	< 0	0.00	7	1.84	1	0.26	2	0.53
TYR	50	0.92	4	0.44	10	1.10	6	0.66	13	1.44	< 0	0.00	> 17	1.88
GLY	> 243	2.32	12	0.69	< 8	0.46	< 6	0.34	< 2	0.11	> 196	11.22	19	1.09
OTH	6	1.09	1	1.09	0	0.00	0	0.00	2	2.18	1	1.09	2	2.18
ALL	1518	1.00	253	1.00	253	1.00	253	1.00	253	1.00	253	1.00	253	1.00

cluster 2: turn-type II

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	106	0.99	23	1.29	20	1.12	> 27	1.52	< 0	0.00	11	0.62	25	1.40
ILE	94	1.14	> 32	2.32	13	0.94	7	0.51	< 0	0.00	14	1.02	> 28	2.03
LEU	148	1.16	19	0.89	19	0.89	24	1.13	< 0	0.00	> 34	1.60	> 52	2.44
MET	36	1.31	> 9	1.96	8	1.75	8	1.75	< 0	0.00	4	0.87	7	1.53
PHE	42	0.70	7	0.70	5	0.50	< 0	0.00	< 1	0.10	> 20	2.00	9	0.90
PRO	< 19	0.30	< 2	0.19	8	0.75	< 1	0.09	< 0	0.00	< 0	0.00	8	0.75
VAL	84	0.85	> 37	2.23	< 7	0.42	< 5	0.30	< 0	0.00	16	0.97	19	1.15
ARG	75	1.06	11	0.93	> 23	1.95	14	1.19	< 5	0.42	> 19	1.61	< 3	0.25
ASP	59	0.72	7	0.51	18	1.31	18	1.31	< 4	0.29	< 5	0.37	7	0.51
GLU	92	0.97	13	0.82	> 26	1.65	23	1.46	< 4	0.25	18	1.14	< 8	0.51
LYS	104	1.22	21	1.48	> 28	1.97	16	1.13	< 5	0.35	> 31	2.18	< 3	0.21
ASN	48	0.76	< 4	0.38	7	0.67	15	1.43	9	0.86	8	0.76	5	0.48
CYS	18	0.78	3	0.78	< 0	0.00	2	0.52	3	0.78	7	1.82	3	0.78
GLN	50	0.91	6	0.66	12	1.31	> 17	1.86	< 2	0.22	10	1.09	< 3	0.33
HIS	34	0.99	8	1.39	5	0.87	5	0.87	5	0.87	4	0.70	7	1.22
SER	74	0.86	9	0.63	18	1.25	> 26	1.81	< 3	0.21	8	0.56	10	0.70
THR	48	0.61	7	0.54	< 4	0.31	16	1.23	< 1	0.08	11	0.84	9	0.69

TRP	7 0.33	1 0.28	2 0.56	1 0.28	0 0.00	0 0.00	3 0.85
TYR	41 0.81	14 1.66	5 0.59	4 0.47	< 0 0.00	11 1.30	7 0.83
GLY	> 234 2.39	< 3 0.18	< 8 0.49	< 5 0.31	> 194 11.91	< 4 0.25	20 1.23
OTH	3 0.58	0 0.00	0 0.00	2 2.34	0 0.00	1 1.17	0 0.00
ALL	1416 1.00	236 1.00	236 1.00	236 1.00	236 1.00	236 1.00	236 1.00

cluster 3: turn-type III

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	18	0.67	4	0.90	6	1.35	<	0 0.00	4	0.90	2	0.45	2	0.45
ILE	12	0.58	3	0.87	4	1.16		0 0.00	3	0.87	1	0.29	1	0.29
LEU	15	0.47	4	0.75	6	1.13	<	0 0.00	3	0.56	<	0 0.00	2	0.38
MET	3	0.44	1	0.87	0	0.00		0 0.00	0	0.00	0	0.00	2	1.75
PHE	10	0.67	3	1.20	2	0.80		0 0.00	3	1.20	0	0.00	2	0.80
PRO	10	0.63	5	1.88	3	1.13		0 0.00	2	0.75	0	0.00	0	0.00
VAL	15	0.60	1	0.24	6	1.45	<	0 0.00	2	0.48	2	0.48	4	0.97
ARG	18	1.02	5	1.70	5	1.70		0 0.00	3	1.02	3	1.02	2	0.68
ASP	> 49	2.38	3	0.88	3	0.88	>	18 5.25	> 13	3.79	5	1.46	>	7 2.04
GLU	40	1.69	<	0 0.00	3	0.76	<	0 0.00	> 11	2.79	>	10 2.53	>	16 4.05
LYS	21	0.99	4	1.13	6	1.69		0 0.00	2	0.56	>	8 2.25	1	0.28
ASN	23	1.46	0	0.00	0	0.00	>	16 6.10	1	0.38	5	1.91	1	0.38
CYS	3	0.52	>	3 3.11	0	0.00		0 0.00	0	0.00	0	0.00	0	0.00
GLN	11	0.80	4	1.75	1	0.44		0 0.00	1	0.44	3	1.31	2	0.88
HIS	8	0.93	0	0.00	1	0.70		0 0.00	1	0.70	3	2.09	3	2.09
SER	37	1.72	4	1.12	3	0.84	>	17 4.74	4	1.12	6	1.67	3	0.84
THR	28	1.43	2	0.61	4	1.23		3 0.92	1	0.31	>	9 2.76	>	9 2.76
TRP	8	1.50	>	3 3.38	>	4 4.51		0 0.00	0	0.00	1	1.13	0	0.00
TYR	11	0.87	>	8 3.79	1	0.47		0 0.00	1	0.47	0	0.00	1	0.47
GLY	14	0.57	2	0.49	1	0.25		5 1.23	4	0.98	1	0.25	1	0.25
OTH	0	0.00	0	0.00	0	0.00		0 0.00	0	0.00	0	0.00	0	0.00
ALL	354	1.00	59	1.00	59	1.00		59 1.00	59	1.00	59	1.00	59	1.00

cluster 4: turn-type IV

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	50	0.91	13	1.42	<	0 0.00	5	0.55	5	0.55	>	17 1.86	10	1.09
ILE	17	0.40	4	0.57	<	0 0.00	<	0 0.00	5	0.71	4	0.57	4	0.57
LEU	38	0.58	6	0.55	<	0 0.00	<	2 0.18	12	1.10	12	1.10	6	0.55
MET	8	0.57	1	0.43	0	0.00	1	0.43	2	0.85	1	0.43	3	1.28
PHE	24	0.78	5	0.98	5	0.98	<	0 0.00	1	0.20	7	1.37	6	1.17
PRO	47	1.44	>	15 2.75	<	0 0.00	<	0 0.00	> 30	5.50	1	0.18	1	0.18
VAL	19	0.37	3	0.35	<	2 0.24	<	1 0.12	4	0.47	3	0.35	6	0.71
ARG	29	0.80	5	0.83	2	0.33	4	0.66	6	0.99	5	0.83	7	1.16
ASP	> 74	1.76	>	16 2.28	3	0.43	>	17 2.42	10	1.42	>	13 1.85	>	15 2.13
GLU	41	0.84	5	0.62	<	2 0.25	3	0.37	6	0.74	>	17 2.10	8	0.99
LYS	32	0.73	8	1.10	4	0.55	<	2 0.27	6	0.82	7	0.96	5	0.69
ASN	35	1.09	2	0.37	>	12 2.23	8	1.49	1	0.19	8	1.49	4	0.74
CYS	6	0.51	0	0.00	0	0.00	2	1.01	0	0.00	0	0.00	4	2.02
GLN	21	0.75	1	0.21	2	0.43	4	0.85	1	0.21	2	0.43	>	11 2.35

HIS	20	1.13	5	1.70	3	1.02	6	2.04	3	1.02	2	0.68	1	0.34			
SER	71	1.61	9	1.22	<	2	0.27	>	29	3.94	>	15	2.04	7	0.95	9	1.22
THR	41	1.02	6	0.90	<	0	0.00	>	23	3.43	5	0.75	5	0.75	2	0.30	
TRP	9	0.82	2	1.10	0	0.00	0	0.00	1	0.55	2	1.10	4	2.20			
TYR	25	0.96	7	1.62	4	0.92	4	0.92	5	1.16	<	0	0.00	5	1.16		
GLY	>	118	2.35	8	0.96	>	80	9.58	10	1.20	3	0.36	8	0.96	9	1.08	
OTH	1	0.38	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	2.28			
ALL	726	1.00	121	1.00	121	1.00	121	1.00	121	1.00	121	1.00	121	1.00			

cluster 5: turn-type V

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6						
	f	P	f	P	f	P	f	f	P	f	P	f	P	f					
ALA	36	0.69	8	0.91	4	0.46	<	1	0.11	9	1.03	6	0.69	8	0.91				
ILE	14	0.34	3	0.44	<	1	0.15	<	1	0.15	2	0.30	3	0.44	4	0.59			
LEU	42	0.67	<	4	0.38	<	3	0.29	<	3	0.29	11	1.05	12	1.15	9	0.86		
MET	3	0.22	0	0.00	1	0.44	1	0.44	1	0.44	0	0.00	0	0.00	0	0.00			
PHE	26	0.88	5	1.02	8	1.63	<	0	0.00	4	0.81	5	1.02	4	0.81				
PRO	41	1.31	>	10	1.91	<	0	0.00	<	0	0.00	>	30	5.74	<	0	0.00	1	0.19
VAL	26	0.53	5	0.61	<	0	0.00	<	2	0.25	8	0.98	<	2	0.25	9	1.11		
ARG	29	0.83	6	1.03	3	0.52	4	0.69	5	0.86	6	1.03	5	0.86					
ASP	70	1.73	>	14	2.08	10	1.48	>	22	3.27	4	0.59	>	14	2.08	6	0.89		
GLU	44	0.95	6	0.77	4	0.52	3	0.39	10	1.29	8	1.03	13	1.68					
LYS	31	0.74	8	1.15	6	0.86	3	0.43	3	0.43	6	0.86	5	0.72					
ASN	>	60	1.94	>	10	1.94	>	13	2.52	>	17	3.30	2	0.39	>	11	2.13	7	1.36
CYS	5	0.44	0	0.00	1	0.53	1	0.53	0	0.00	0	0.00	3	1.58					
GLN	21	0.78	5	1.11	2	0.45	2	0.45	1	0.22	5	1.11	6	1.33					
HIS	17	1.00	3	1.06	4	1.42	4	1.42	0	0.00	3	1.06	3	1.06					
SER	71	1.68	>	14	1.99	5	0.71	>	23	3.26	11	1.56	8	1.13	10	1.42			
THR	54	1.40	6	0.93	<	0	0.00	>	24	3.74	4	0.62	11	1.71	9	1.40			
TRP	10	0.96	1	0.57	1	0.57	0	0.00	3	1.72	>	5	2.87	0	0.00				
TYR	17	0.68	3	0.72	1	0.24	1	0.24	5	1.20	5	1.20	2	0.48					
GLY	78	1.62	5	0.62	>	49	6.12	4	0.50	3	0.38	5	0.62	12	1.50				
OTH	1	0.40	0	0.00	0	0.00	0	0.00	0	0.00	1	2.38	0	0.00					
ALL	696	1.00	116	1.00	116	1.00	116	1.00	116	1.00	116	1.00	116	1.00					

cluster 6: turn-type VI

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6				
	f	P	f	P	f	P	f	f	P	f	P	f	P	f			
ALA	31	1.07	3	0.62	5	1.03	>	9	1.86	2	0.41	6	1.24	6	1.24		
ILE	14	0.62	3	0.80	<	0	0.00	2	0.54	1	0.27	5	1.34	3	0.80		
LEU	41	1.18	8	1.39	<	0	0.00	5	0.87	7	1.21	10	1.73	>	11	1.91	
MET	7	0.94	1	0.80	2	1.61	0	0.00	0	0.00	0	0.00	>	4	3.22		
PHE	17	1.04	5	1.84	2	0.74	3	1.11	3	1.11	2	0.74	2	0.74			
PRO	20	1.16	6	2.08	0	0.00	>	10	3.47	3	1.04	1	0.35	0	0.00		
VAL	19	0.71	2	0.45	1	0.22	2	0.45	1	0.22	5	1.11	8	1.78			
ARG	17	0.89	3	0.94	3	0.94	1	0.31	3	0.94	4	1.25	3	0.94			
ASP	25	1.12	2	0.54	>	10	2.69	<	0	0.00	>	10	2.69	1	0.27	2	0.54
GLU	18	0.70	4	0.93	3	0.70	1	0.23	4	0.93	6	1.40	<	0	0.00		
LYS	24	1.04	3	0.78	2	0.52	2	0.52	7	1.82	4	1.04	6	1.56			

ASN	22 1.29	4 1.41	> 8 2.81	2 0.70	2 0.70	2 0.70	4 1.41
CYS	7 1.12	2 1.91	3 2.87	1 0.96	0 0.00	1 0.96	0 0.00
GLN	14 0.94	2 0.81	3 1.21	3 1.21	3 1.21	2 0.81	1 0.40
HIS	19 2.03	0 0.00	> 6 3.85	> 5 3.21	2 1.28	2 1.28	> 4 2.57
SER	24 1.03	5 1.29	4 1.03	3 0.77	6 1.54	3 0.77	3 0.77
THR	16 0.75	2 0.56	0 0.00	3 0.85	4 1.13	6 1.69	1 0.28
TRP	4 0.69	2 2.08	0 0.00	0 0.00	2 2.08	0 0.00	0 0.00
TYR	13 0.95	3 1.31	2 0.87	> 7 3.06	0 0.00	0 0.00	1 0.44
GLY	32 1.21	4 0.91	> 10 2.26	5 1.13	4 0.91	4 0.91	5 1.13
OTH	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
ALL	384 1.00	64 1.00	64 1.00	64 1.00	64 1.00	64 1.00	64 1.00

cluster 7: turn-type VII

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	1759	0.95	< 228	0.74	< 110	0.36	327	1.05	> 356	1.15	304	0.98	> 434	1.40
ILE	< 1110	0.77	> 351	1.46	< 47	0.20	< 166	0.69	< 112	0.47	< 166	0.69	268	1.12
LEU	< 1901	0.86	> 436	1.18	< 93	0.25	337	0.91	< 157	0.42	398	1.08	> 480	1.30
MET	413	0.86	> 111	1.39	< 20	0.25	< 53	0.66	< 40	0.50	76	0.95	> 113	1.42
PHE	950	0.91	> 267	1.53	< 65	0.37	< 134	0.77	< 95	0.55	171	0.98	> 218	1.25
PRO	> 1551	1.40	194	1.05	> 346	1.87	> 828	4.47	< 103	0.56	< 25	0.14	< 55	0.30
VAL	< 1428	0.83	> 385	1.34	< 61	0.21	< 242	0.84	< 160	0.56	286	0.99	294	1.02
ARG	1221	0.99	< 168	0.82	< 123	0.60	216	1.05	224	1.09	199	0.97	> 291	1.42
ASP	> 2368	1.65	< 150	0.63	> 1181	4.95	< 167	0.70	> 437	1.83	> 322	1.35	< 111	0.47
GLU	1826	1.11	< 165	0.60	< 180	0.66	> 325	1.18	> 687	2.50	> 349	1.27	< 120	0.44
LYS	1458	0.98	< 184	0.74	< 114	0.46	> 279	1.13	> 343	1.39	> 298	1.20	240	0.97
ASN	> 1359	1.24	< 155	0.85	> 546	2.99	< 77	0.42	193	1.06	> 229	1.25	159	0.87
CYS	432	1.07	66	0.98	> 136	2.03	< 24	0.36	< 32	0.48	> 115	1.71	59	0.88
GLN	949	0.99	< 122	0.77	< 80	0.50	153	0.96	> 197	1.24	171	1.07	> 226	1.42
HIS	591	0.99	94	0.94	> 152	1.52	< 59	0.59	106	1.06	103	1.03	< 77	0.77
SER	1629	1.09	< 175	0.70	> 481	1.93	< 206	0.83	> 324	1.30	249	1.00	< 194	0.78
THR	1299	0.95	225	0.99	< 179	0.79	< 181	0.80	212	0.93	> 351	1.54	< 151	0.66
TRP	358	0.97	> 91	1.47	< 22	0.36	> 86	1.39	< 43	0.70	52	0.84	64	1.04
TYR	790	0.90	> 234	1.59	< 61	0.42	124	0.84	< 105	0.71	130	0.88	136	0.93
GLY	< 1182	0.70	288	1.02	< 105	0.37	< 113	0.40	< 179	0.63	< 102	0.36	> 395	1.39
OTH	68	0.76	18	1.21	< 5	0.34	10	0.67	< 2	0.13	11	0.74	22	1.48
ALL	24642	1.00	4107	1.00	4107	1.00	4107	1.00	4107	1.00	4107	1.00	4107	1.00

cluster 8: turn-type VIIa

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	< 135	0.66	< 16	0.47	< 15	0.44	> 54	1.58	< 19	0.55	< 6	0.18	25	0.73
ILE	< 74	0.47	> 46	1.74	< 1	0.04	< 6	0.23	< 2	0.08	< 0	0.00	19	0.72
LEU	< 111	0.45	37	0.90	35	0.86	< 11	0.27	< 10	0.24	< 3	0.07	< 15	0.37
MET	26	0.49	13	1.47	6	0.68	< 2	0.23	< 2	0.23	< 1	0.11	< 2	0.23
PHE	69	0.60	> 36	1.87	15	0.78	< 4	0.21	< 6	0.31	< 4	0.21	< 4	0.21
PRO	92	0.75	< 11	0.54	< 2	0.10	> 79	3.86	< 0	0.00	< 0	0.00	< 0	0.00
VAL	< 104	0.54	> 64	2.01	< 2	0.06	< 12	0.38	< 4	0.13	< 0	0.00	22	0.69
ARG	114	0.84	27	1.19	16	0.71	27	1.19	< 6	0.26	< 2	0.09	> 36	1.59

ASP	> 316 2.00	< 6 0.23	> 90 3.41	26 0.99	> 164 6.22	< 10 0.38	20 0.76
GLU	133 0.73	< 18 0.59	< 7 0.23	> 41 1.35	33 1.09	< 5 0.17	29 0.95
LYS	173 1.05	> 44 1.61	< 14 0.51	33 1.21	< 16 0.59	< 5 0.18	> 61 2.23
ASN	> 213 1.76	< 8 0.40	> 39 1.93	> 29 1.44	> 88 4.36	18 0.89	> 31 1.54
CYS	21 0.47	8 1.08	8 1.08	< 1 0.14	4 0.54	< 0 0.00	< 0 0.00
GLN	100 0.95	25 1.42	14 0.80	12 0.68	17 0.97	< 3 0.17	> 29 1.65
HIS	56 0.84	9 0.81	11 1.00	13 1.18	9 0.81	< 4 0.36	10 0.90
SER	205 1.24	< 7 0.25	> 57 2.07	> 51 1.85	36 1.30	< 3 0.11	> 51 1.85
THR	> 228 1.51	> 44 1.75	> 78 3.10	26 1.03	17 0.68	< 2 0.08	> 61 2.43
TRP	17 0.42	3 0.44	2 0.29	4 0.59	3 0.44	< 1 0.15	4 0.59
TYR	66 0.68	19 1.17	15 0.92	< 6 0.37	9 0.55	< 1 0.06	16 0.99
GLY	> 464 2.47	< 10 0.32	26 0.83	< 16 0.51	< 7 0.22	> 386 12.31	< 19 0.61
OTH	7 0.71	3 1.82	1 0.61	1 0.61	2 1.21	0 0.00	0 0.00
ALL	2724 1.00	454 1.00	454 1.00	454 1.00	454 1.00	454 1.00	454 1.00

cluster 9: turn-type VIIIb

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	31	0.45	8	0.70	< 0	0.00	15	1.31	7	0.61	< 0	0.00	< 1	0.09
ILE	21	0.39	12	1.35	< 0	0.00	6	0.68	< 0	0.00	< 0	0.00	< 3	0.34
LEU	< 30	0.37	14	1.02	< 1	0.07	< 2	0.15	< 4	0.29	< 2	0.15	7	0.51
MET	10	0.56	5	1.69	0	0.00	0	0.00	1	0.34	4	1.36	0	0.00
PHE	27	0.70	> 18	2.79	< 0	0.00	3	0.47	4	0.62	< 1	0.16	< 1	0.16
PRO	< 10	0.24	4	0.58	< 0	0.00	5	0.73	< 1	0.15	< 0	0.00	< 0	0.00
VAL	44	0.69	> 30	2.81	< 0	0.00	< 3	0.28	< 2	0.19	< 0	0.00	9	0.84
ARG	31	0.68	3	0.40	< 0	0.00	4	0.53	10	1.32	< 1	0.13	> 13	1.71
ASP	> 128	2.42	< 1	0.11	> 85	9.63	> 17	1.93	> 15	1.70	< 3	0.34	7	0.79
GLU	64	1.05	< 4	0.39	< 0	0.00	> 23	2.26	16	1.57	< 0	0.00	> 21	2.06
LYS	79	1.44	> 19	2.08	< 0	0.00	> 20	2.18	> 17	1.86	< 2	0.22	> 21	2.29
ASN	> 124	3.06	4	0.59	> 49	7.25	6	0.89	> 30	4.44	9	1.33	> 26	3.85
CYS	3	0.20	2	0.81	0	0.00	0	0.00	0	0.00	0	0.00	1	0.40
GLN	37	1.05	4	0.68	< 1	0.17	> 12	2.04	5	0.85	3	0.51	> 12	2.04
HIS	22	0.99	1	0.27	1	0.27	3	0.81	> 11	2.97	0	0.00	6	1.62
SER	43	0.78	5	0.54	< 2	0.22	15	1.62	12	1.30	< 2	0.22	7	0.76
THR	33	0.65	10	1.19	5	0.59	7	0.83	5	0.59	< 0	0.00	6	0.71
TRP	4	0.29	0	0.00	0	0.00	1	0.44	1	0.44	0	0.00	2	0.88
TYR	22	0.67	6	1.10	< 0	0.00	< 0	0.00	8	1.47	1	0.18	7	1.29
GLY	> 148	2.35	< 1	0.10	8	0.76	10	0.95	< 3	0.29	> 124	11.82	< 2	0.19
OTH	1	0.30	1	1.81	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	912	1.00	152	1.00	152	1.00	152	1.00	152	1.00	152	1.00	152	1.00

cluster 10: turn-type IXa

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	30	0.64	5	0.64	4	0.51	5	0.64	5	0.64	9	1.15	< 2	0.26
ILE	25	0.69	8	1.32	< 0	0.00	< 1	0.17	< 0	0.00	6	0.99	10	1.65
LEU	41	0.73	14	1.49	4	0.43	< 2	0.21	7	0.75	6	0.64	8	0.85
MET	10	0.83	2	0.99	1	0.50	2	0.99	1	0.50	3	1.49	1	0.50
PHE	25	0.95	4	0.91	4	0.91	3	0.68	3	0.68	8	1.82	3	0.68

PRO	32	1.14	3	0.64	<	0	0.00	>	26	5.55	<	0	0.00	<	0	0.00	3	0.64	
VAL	27	0.62	9	1.23	<	0	0.00		3	0.41	<	1	0.14		5	0.69	9	1.23	
ARG	27	0.87	3	0.58		4	0.77		5	0.96		6	1.15		3	0.58	6	1.15	
ASP	48	1.32	5	0.83		9	1.49		9	1.49	>	20	3.31		4	0.66	<	1	0.17
GLU	43	1.03	3	0.43		3	0.43	>	17	2.44		9	1.29		6	0.86		5	0.72
LYS	33	0.88	10	1.60		3	0.48		6	0.96		5	0.80		3	0.48		6	0.96
ASN	34	1.23	6	1.30		8	1.73		4	0.87	>	11	2.38		4	0.87		1	0.22
CYS	8	0.78	2	1.18		0	0.00		1	0.59		2	1.18		1	0.59		2	1.18
GLN	28	1.16	7	1.74		4	0.99		3	0.74		3	0.74		4	0.99		7	1.74
HIS	21	1.38	3	1.18		4	1.58		1	0.40	>	7	2.76		4	1.58		2	0.79
SER	>	67	1.77	5	0.79	>	26	4.11	5	0.79	5	0.79	>	16	2.53		10	1.58	
THR	60	1.74	9	1.56	>	21	3.65		4	0.70		8	1.39		7	1.22	>	11	1.91
TRP	13	1.39	1	0.64		0	0.00		3	1.92		2	1.28		3	1.92	>	4	2.56
TYR	21	0.94	2	0.54		3	0.81	<	0	0.00		5	1.34	>	8	2.15		3	0.81
GLY	29	0.67	3	0.42		6	0.84		4	0.56		4	0.56		3	0.42		9	1.25
OTH	2	0.88	0	0.00		0	0.00		0	0.00		0	0.00		1	2.65		1	2.65
ALL	624	1.00	104	1.00		104	1.00		104	1.00		104	1.00		104	1.00		104	1.00

cluster 11: turn-type IXb

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6							
	f	P	f	P	f	P	f	f	P	f	P	f	P	f						
ALA	45	0.69	12	1.11	<	3	0.28	13	1.20	<	0	0.00	8	0.74	9	0.83				
ILE	40	0.80	>	17	2.04	<	0	0.00	<	2	0.24	<	1	0.12	5	0.60	>	15	1.80	
LEU	43	0.56	13	1.01	<	3	0.23	<	2	0.16	<	3	0.23	<	5	0.39		17	1.32	
MET	9	0.54	2	0.72		0	0.00		0	0.00		0	0.00		4	1.44		3	1.08	
PHE	25	0.69	4	0.66	<	0	0.00	<	1	0.17		3	0.50		8	1.32		9	1.49	
PRO	<	9	0.23	2	0.31	<	0	0.00	3	0.47	<	0	0.00	<	0	0.00		4	0.62	
VAL	47	0.78	11	1.10	<	0	0.00	<	1	0.10		6	0.60		12	1.20	>	17	1.69	
ARG	38	0.89	10	1.40	<	2	0.28		5	0.70		5	0.70		8	1.12		8	1.12	
ASP	>	102	2.05	<	1	0.12	>	30	3.61	>	37	4.45	>	23	2.77		4	0.48	7	0.84
GLU	60	1.05	13	1.36	<	2	0.21	>	24	2.51		12	1.25		8	0.84	<	1	0.10	
LYS	43	0.83	11	1.28	<	0	0.00		11	1.28		11	1.28		7	0.81	<	3	0.35	
ASN	59	1.55	2	0.32	>	22	3.46		10	1.57	>	14	2.20		6	0.94		5	0.79	
CYS	11	0.78	2	0.86		1	0.43		0	0.00		0	0.00		2	0.86	>	6	2.57	
GLN	35	1.05	6	1.08		3	0.54		5	0.90		10	1.80		8	1.44		3	0.54	
HIS	24	1.15	2	0.57		0	0.00		4	1.15		5	1.44	>	11	3.16		2	0.57	
SER	85	1.63	14	1.61	>	27	3.11	>	15	1.73		10	1.15		14	1.61		5	0.57	
THR	>	90	1.90	11	1.39	>	26	3.28	4	0.51	>	24	3.03	>	14	1.77		11	1.39	
TRP	9	0.70	1	0.47		0	0.00		1	0.47		2	0.93		2	0.93		3	1.40	
TYR	34	1.11	7	1.37		2	0.39	<	0	0.00		9	1.76	>	10	1.95		6	1.17	
GLY	47	0.79	<	2	0.20	>	20	2.03	4	0.41		5	0.51		7	0.71		9	0.91	
OTH	3	0.96	0	0.00	>	2	3.85		1	1.93		0	0.00		0	0.00		0	0.00	
ALL	858	1.00	143	1.00		143	1.00		143	1.00		143	1.00		143	1.00		143	1.00	

cluster 12: turn-type Xa

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6						
	f	P	f	P	f	P	f	f	P	f	P	f	P	f					
ALA	25	1.06	3	0.76	<	0	0.00	>	9	2.29		6	1.53	<	0	0.00		7	1.78
ILE	12	0.66	4	1.32		0	0.00		1	0.33		1	0.33		0	0.00		6	1.98

LEU	12	0.43	2	0.43	1	0.21	1	0.21	2	0.43	<	0	0.00	6	1.28		
MET	4	0.66	2	1.98	0	0.00	0	0.00	1	0.99		0	0.00	1	0.99		
PHE	6	0.45	2	0.91	1	0.45	0	0.00	2	0.91		0	0.00	1	0.45		
PRO	7	0.50	0	0.00	1	0.43	3	1.28	0	0.00		0	0.00	3	1.28		
VAL	13	0.59	6	1.64	<	0	0.00	3	0.82	<	0	0.00	<	0	0.00	4	1.10
ARG	10	0.64	4	1.54		0	0.00	1	0.39	3	1.15		0	0.00	2	0.77	
ASP	26	1.44	1	0.33	>	14	4.64	>	7	2.32	3	0.99		0	0.00	1	0.33
GLU	18	0.86	0	0.00		0	0.00	6	1.72	7	2.01		0	0.00	5	1.44	
LYS	23	1.22	6	1.92	1	0.32	6	1.92	>	8	2.55		0	0.00	2	0.64	
ASN	20	1.44	1	0.43	4	1.73	2	0.87	>	8	3.46		0	0.00	5	2.16	
CYS	2	0.39	1	1.18	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00	1	1.18
GLN	8	0.66	4	1.99	0	0.00	2	0.99	2	0.99	2	0.99		0	0.00	0	0.00
HIS	3	0.40	1	0.79	0	0.00	0	0.00	2	1.58	0	0.00		0	0.00	0	0.00
SER	26	1.37	6	1.90	5	1.58	>	9	2.85	4	1.26		0	0.00	2	0.63	
THR	17	0.98	4	1.39	>	9	3.13	1	0.35	1	0.35		0	0.00	2	0.70	
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00	0	0.00
TYR	8	0.72	3	1.61	1	0.54	0	0.00	1	0.54	0	0.00		0	0.00	3	1.61
GLY	>	71	3.30	2	0.56	>	15	4.18	1	0.28	1	0.28	>	51	14.21	1	0.28
OTH	1	0.88	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		1	5.30	0	0.00
ALL	312	1.00	52	1.00	52	1.00	52	1.00	52	1.00	52	1.00		52	1.00	52	1.00

cluster 13: turn-type Xb

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6			
	f	P	f	P	f	P	f	f	P	f	P	f	P	f		
ALA	2	0.18	1	0.53	1	0.53	0	0.00	0	0.00	0	0.00	0	0.00		
ILE	7	0.80	2	1.37	0	0.00	1	0.69	0	0.00	0	0.00	>	4	2.74	
LEU	10	0.74	1	0.44	2	0.89	0	0.00	2	0.89	0	0.00		5	2.22	
MET	3	1.03	1	2.06	1	2.06	0	0.00	0	0.00	0	0.00		1	2.06	
PHE	4	0.63	2	1.89	0	0.00	0	0.00	1	0.94	0	0.00		1	0.94	
PRO	9	1.33	0	0.00	0	0.00	>	5	4.44	0	0.00	0	0.00	>	4	3.55
VAL	6	0.57	4	2.28	0	0.00	1	0.57	0	0.00	0	0.00		1	0.57	
ARG	7	0.93	2	1.60	0	0.00	0	0.00	3	2.40	1	0.80		1	0.80	
ASP	8	0.92	3	2.07	1	0.69	1	0.69	3	2.07	0	0.00		0	0.00	
GLU	14	1.40	1	0.60	0	0.00	>	8	4.78	4	2.39	0	0.00		1	0.60
LYS	7	0.78	1	0.66	2	1.33	2	1.33	2	1.33	0	0.00		0	0.00	
ASN	3	0.45	1	0.90	0	0.00	0	0.00	2	1.80	0	0.00		0	0.00	
CYS	2	0.82	>	2	4.89	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00
GLN	6	1.03	2	2.06	1	1.03	2	2.06	1	1.03	0	0.00		0	0.00	
HIS	2	0.55	0	0.00	1	1.64	0	0.00	1	1.64	0	0.00		0	0.00	
SER	9	0.99	0	0.00	2	1.32	3	1.97	1	0.66	0	0.00		3	1.97	
THR	17	2.05	0	0.00	>	10	7.23	2	1.45	3	2.17	0	0.00		2	1.45
TRP	1	0.44	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		1	2.66	
TYR	3	0.56	1	1.12	1	1.12	0	0.00	1	1.12	0	0.00		0	0.00	
GLY	>	30	2.90	1	0.58	3	1.74	0	0.00	1	0.58	>	24	13.90	1	0.58
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00	
ALL	150	1.00	25	1.00	25	1.00	25	1.00	25	1.00	25	1.00		25	1.00	

cluster 14: turn-type XI

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	< 135	0.44	< 27	0.53	< 29	0.57	< 22	0.43	< 4	0.08	< 30	0.59	< 23	0.45
ILE	239	1.01	> 53	1.34	> 80	2.03	< 0	0.00	< 0	0.00	44	1.12	> 62	1.57
LEU	< 210	0.57	47	0.77	49	0.81	< 9	0.15	< 2	0.03	< 29	0.48	74	1.22
MET	62	0.79	18	1.37	< 4	0.31	< 6	0.46	< 2	0.15	14	1.07	18	1.37
PHE	140	0.82	36	1.26	> 47	1.64	< 13	0.45	< 4	0.14	< 12	0.42	28	0.98
PRO	< 91	0.50	< 8	0.26	< 1	0.03	< 0	0.00	< 0	0.00	< 0	0.00	> 82	2.70
VAL	319	1.12	> 67	1.41	> 111	2.34	< 2	0.04	< 2	0.04	> 69	1.46	> 68	1.44
ARG	177	0.87	30	0.89	36	1.07	24	0.71	< 7	0.21	> 55	1.63	25	0.74
ASP	270	1.15	< 23	0.59	< 21	0.54	> 150	3.83	41	1.05	< 11	0.28	< 24	0.61
GLU	261	0.96	55	1.22	44	0.97	41	0.91	< 12	0.27	> 69	1.53	40	0.89
LYS	> 358	1.47	45	1.11	> 62	1.53	40	0.98	< 12	0.30	> 154	3.79	45	1.11
ASN	> 346	1.92	29	0.97	< 17	0.57	> 208	6.93	> 54	1.80	< 19	0.63	< 19	0.63
CYS	53	0.80	14	1.27	9	0.82	5	0.45	< 1	0.09	6	0.54	> 18	1.63
GLN	157	1.00	21	0.80	26	0.99	29	1.11	< 9	0.34	> 58	2.22	< 14	0.54
HIS	96	0.97	12	0.73	> 26	1.58	21	1.28	< 4	0.24	23	1.40	10	0.61
SER	< 166	0.67	48	1.17	< 27	0.66	< 22	0.54	< 20	0.49	< 22	0.54	< 27	0.66
THR	< 148	0.66	> 64	1.71	< 16	0.43	< 0	0.00	< 3	0.08	30	0.80	35	0.94
TRP	53	0.87	12	1.18	10	0.99	< 1	0.10	< 0	0.00	4	0.39	> 26	2.56
TYR	165	1.14	> 51	2.11	> 56	2.32	< 9	0.37	< 3	0.12	20	0.83	26	1.08
GLY	> 600	2.15	< 13	0.28	< 4	0.09	> 71	1.52	> 495	10.6	< 6	0.13	< 11	0.24
OTH	4	0.27	2	0.82	0	0.00	2	0.82	0	0.00	0	0.00	0	0.00
ALL	4050	1.00	675	1.00	675	1.00	675	1.00	675	1.00	675	1.00	675	1.00

cluster 15: turn-type XII

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	< 70	0.45	< 11	0.43	18	0.70	< 2	0.08	< 12	0.47	< 14	0.55	< 13	0.51
ILE	84	0.71	19	0.96	22	1.11	< 1	0.05	< 3	0.15	14	0.71	25	1.26
LEU	130	0.71	< 20	0.65	22	0.72	< 2	0.07	< 8	0.26	< 17	0.56	> 61	1.99
MET	35	0.88	11	1.67	5	0.76	2	0.30	< 0	0.00	7	1.06	10	1.51
PHE	70	0.81	12	0.83	21	1.46	< 1	0.07	< 2	0.14	17	1.18	17	1.18
PRO	< 24	0.26	< 5	0.33	< 1	0.07	< 0	0.00	12	0.78	< 0	0.00	< 6	0.39
VAL	152	1.06	> 37	1.55	> 35	1.47	< 4	0.17	< 2	0.08	29	1.22	> 45	1.89
ARG	118	1.16	> 36	2.12	> 25	1.47	< 4	0.24	23	1.35	21	1.24	< 9	0.53
ASP	> 176	1.49	< 6	0.30	28	1.42	26	1.32	> 88	4.46	22	1.11	< 6	0.30
GLU	155	1.14	< 13	0.57	> 33	1.45	< 10	0.44	> 41	1.80	> 34	1.49	24	1.06
LYS	146	1.19	> 37	1.81	26	1.27	13	0.64	20	0.98	> 36	1.76	14	0.68
ASN	109	1.20	11	0.73	17	1.13	13	0.86	> 42	2.78	19	1.26	< 7	0.46
CYS	24	0.72	7	1.26	7	1.26	1	0.18	1	0.18	3	0.54	5	0.90
GLN	89	1.13	20	1.52	19	1.44	< 5	0.38	12	0.91	> 21	1.59	12	0.91
HIS	33	0.66	6	0.72	12	1.45	< 1	0.12	< 2	0.24	9	1.09	3	0.36
SER	111	0.90	19	0.92	16	0.77	< 4	0.19	> 32	1.55	21	1.02	19	0.92
THR	110	0.97	25	1.33	< 9	0.48	< 0	0.00	22	1.17	> 36	1.91	18	0.96
TRP	26	0.85	7	1.37	2	0.39	1	0.20	< 0	0.00	4	0.78	> 12	2.35
TYR	73	1.00	17	1.40	> 20	1.64	< 2	0.16	< 3	0.25	11	0.90	> 20	1.64
GLY	> 302	2.14	21	0.90	< 2	0.09	> 248	10.57	15	0.64	< 3	0.13	< 13	0.55
OTH	3	0.41	0	0.00	0	0.00	0	0.00	0	0.00	2	1.62	1	0.81

ALL	2040	1.00	340	1.00	340	1.00	340	1.00	340	1.00	340	1.00	340	1.00
-----	------	------	-----	------	-----	------	-----	------	-----	------	-----	------	-----	------

cluster 16: turn-type XIIIa

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6						
	f	P	f	P	f	P	f	f	P	f	P	f	P	f					
ALA	41	1.10	10	1.62	<	1	0.16	<	1	0.16	7	1.13	8	1.29	>	14	2.26		
ILE	17	0.59	5	1.04		2	0.42	<	0	0.00	3	0.63	2	0.42		5	1.04		
LEU	34	0.77	11	1.49		3	0.41	<	0	0.00	8	1.08	3	0.41		9	1.22		
MET	8	0.84	0	0.00		2	1.26		0	0.00	3	1.88	3	1.88		0	0.00		
PHE	13	0.62	2	0.57		4	1.15		0	0.00	3	0.86	3	0.86		1	0.29		
PRO	12	0.54	3	0.81	<	0	0.00	<	0	0.00	4	1.08	1	0.27		4	1.08		
VAL	15	0.43	4	0.70		7	1.22	<	0	0.00	2	0.35	<	0	0.00	2	0.35		
ARG	36	1.46	7	1.71	>	9	2.19		2	0.49	3	0.73	3	0.73	>	12	2.93		
ASP	28	0.98	3	0.63		1	0.21		2	0.42	>	11	2.31	>	11	2.31	<	0	0.00
GLU	23	0.70	<	0	0.00	8	1.46	<	0	0.00	7	1.28	6	1.09		2	0.36		
LYS	22	0.74	2	0.41		4	0.81		1	0.20	2	0.41	5	1.01		8	1.62		
ASN	18	0.82	4	1.10		2	0.55		0	0.00	4	1.10	7	1.92		1	0.27		
CYS	7	0.87	1	0.75		2	1.49		1	0.75	1	0.75	0	0.00		2	1.49		
GLN	13	0.68	2	0.63		5	1.57		0	0.00	3	0.94	3	0.94		0	0.00		
HIS	15	1.25	3	1.50	>	5	2.50		2	1.00	1	0.50	3	1.50		1	0.50		
SER	36	1.20	5	1.00		9	1.81		6	1.20	>	10	2.01		4	0.80	2	0.40	
THR	27	0.99	4	0.88	>	13	2.86	<	0	0.00	3	0.66	4	0.88		3	0.66		
TRP	8	1.08	0	0.00		1	0.81		0	0.00	3	2.43	2	1.62		2	1.62		
TYR	12	0.68	3	1.02		0	0.00		0	0.00	2	0.68	3	1.02		4	1.36		
GLY	>	104	3.06	>	13	2.30	3	0.53	>	66	11.66	<	1	0.18	>	11	1.94	10	1.77
OTH	3	1.68	0	0.00		1	3.36		1	3.36	1	3.36	0	0.00		0	0.00		
ALL	492	1.00	82	1.00		82	1.00		82	1.00	82	1.00	82	1.00		82	1.00		

cluster 17: turn-type XIIIb

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6				
	f	P	f	P	f	P	f	f	P	f	P	f	P	f			
ALA	29	1.31	7	1.89	<	0	0.00	<	0	0.00	7	1.89	5	1.35	>	10	2.70
ILE	12	0.70	4	1.40		4	1.40		0	0.00	2	0.70	0	0.00		2	0.70
LEU	9	0.34	4	0.91	<	0	0.00	<	0	0.00	1	0.23	2	0.45		2	0.45
MET	4	0.70	1	1.05		2	2.10		0	0.00	0	0.00	0	0.00		1	1.05
PHE	5	0.40	2	0.96		2	0.96		0	0.00	1	0.48	0	0.00		0	0.00
PRO	9	0.68	2	0.91		2	0.91		0	0.00	3	1.36	2	0.91		0	0.00
VAL	18	0.87	4	1.16		6	1.75		0	0.00	3	0.87	3	0.87		2	0.58
ARG	11	0.75	0	0.00		3	1.23		0	0.00	4	1.63	3	1.23		1	0.41
ASP	10	0.59	0	0.00		0	0.00		0	0.00	2	0.70	4	1.41		4	1.41
GLU	21	1.07	0	0.00		2	0.61		0	0.00	5	1.53	>	8	2.44	6	1.83
LYS	12	0.68	0	0.00		5	1.69		0	0.00	4	1.36	2	0.68		1	0.34
ASN	11	0.84	2	0.92		3	1.38		1	0.46	2	0.92	1	0.46		2	0.92
CYS	4	0.83	1	1.25		1	1.25		0	0.00	0	0.00	2	2.50		0	0.00
GLN	8	0.70	0	0.00		3	1.58		0	0.00	2	1.05	1	0.53		2	1.05
HIS	9	1.26	1	0.84		2	1.68		0	0.00	1	0.84	>	4	3.35	1	0.84
SER	22	1.23	5	1.68		2	0.67		4	1.34	3	1.01	2	0.67		6	2.01
THR	19	1.17	2	0.74		5	1.84		2	0.74	3	1.11	>	6	2.21	1	0.37
TRP	5	1.13	1	1.36		1	1.36		0	0.00	1	1.36	0	0.00		2	2.71

TYR	10	0.95	1	0.57	>	5	2.85	0	0.00	2	1.14	1	0.57	1	0.57	
GLY	>	65	3.20	>	11	3.25	1	0.30	>	42	12.42	3	0.89	3	0.89	
OTH	1	0.94	1	5.62	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	294	1.00	49	1.00	49	1.00	49	1.00	49	1.00	49	1.00	49	1.00	49	1.00

cluster 18: turn-type XIV

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6							
	f	P	f	P	f	P	f	f	P	f	P	f	P	f						
ALA	1011	1.00	>	211	1.26	156	0.93	<	57	0.34	166	0.99	>	213	1.27	>	208	1.24		
ILE	<	410	0.53	130	1.00	<	50	0.39	<	18	0.14	<	82	0.63	<	51	0.39	<	79	0.61
LEU	<	777	0.65	199	0.99	<	128	0.64	<	49	0.25	<	123	0.61	<	116	0.58	<	162	0.81
MET	191	0.74	31	0.72	43	1.00	<	20	0.46	<	23	0.53	<	30	0.70	44	1.02			
PHE	522	0.92	>	143	1.52	<	69	0.73	<	41	0.44	<	57	0.61	<	66	0.70	>	146	1.55
PRO	>	888	1.48	>	187	1.87	97	0.97	<	6	0.06	>	556	5.55	<	33	0.33	<	9	0.09
VAL	<	588	0.63	161	1.03	<	67	0.43	<	24	0.15	<	104	0.67	<	78	0.50	154	0.99	
ARG	572	0.86	<	75	0.68	>	147	1.32	<	75	0.68	95	0.86	108	0.97	<	72	0.65		
ASP	>	1263	1.63	<	86	0.67	>	185	1.43	>	440	3.41	113	0.88	>	221	1.71	>	218	1.69
GLU	>	1109	1.24	<	91	0.61	167	1.12	<	84	0.56	>	216	1.45	>	349	2.35	>	202	1.36
LYS	<	669	0.83	<	86	0.64	>	185	1.38	<	64	0.48	137	1.02	132	0.99	<	65	0.49	
ASN	>	762	1.29	<	65	0.66	>	153	1.55	>	257	2.60	<	44	0.45	>	139	1.41	104	1.05
CYS	249	1.14	>	80	2.20	<	19	0.52	>	68	1.87	<	10	0.28	<	18	0.50	>	54	1.49
GLN	501	0.97	<	53	0.62	103	1.20	<	51	0.59	89	1.03	101	1.17	>	104	1.21			
HIS	368	1.13	54	1.00	57	1.05	>	92	1.70	<	34	0.63	68	1.26	63	1.16				
SER	>	1337	1.65	>	173	1.28	>	206	1.53	>	508	3.76	131	0.97	>	177	1.31	142	1.05	
THR	>	865	1.17	114	0.93	>	194	1.58	>	242	1.97	<	81	0.66	132	1.07	102	0.83		
TRP	191	0.95	38	1.14	24	0.72	<	14	0.42	34	1.02	39	1.17	42	1.26					
TYR	432	0.91	>	104	1.31	<	56	0.70	<	37	0.47	<	58	0.73	75	0.94	>	102	1.28	
GLY	<	603	0.66	131	0.85	<	113	0.74	<	72	0.47	<	66	0.43	<	75	0.49	146	0.95	
OTH	18	0.37	9	1.12	<	2	0.25	<	2	0.25	<	2	0.25	<	0	0.00	3	0.37		
ALL	13326	1.00	2221	1.00	2221	1.00	2221	1.00	2221	1.00	2221	1.00	2221	1.00	2221	1.00	2221	1.00		

cluster 19: turn-type XV

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6							
	f	P	f	P	f	P	f	f	P	f	P	f	P	f						
ALA	64	1.05	12	1.19	9	0.89	>	24	2.37	6	0.59	<	4	0.40	9	0.89				
ILE	23	0.49	6	0.77	4	0.51	<	0	0.00	3	0.38	5	0.64	5	0.64					
LEU	55	0.76	9	0.75	8	0.66	12	0.99	<	5	0.41	7	0.58	14	1.16					
MET	16	1.02	5	1.92	5	1.92	3	1.15	1	0.38	1	0.38	1	0.38						
PHE	20	0.59	6	1.06	4	0.70	3	0.53	<	1	0.18	5	0.88	<	1	0.18				
PRO	>	73	2.01	>	11	1.82	3	0.50	<	1	0.17	>	26	4.30	<	1	0.17	>	31	5.13
VAL	31	0.55	4	0.43	<	3	0.32	<	0	0.00	5	0.53	12	1.28	7	0.74				
ARG	44	1.10	4	0.60	6	0.90	4	0.60	7	1.05	10	1.49	>	13	1.94					
ASP	50	1.07	5	0.64	9	1.16	4	0.51	>	17	2.18	12	1.54	3	0.39					
GLU	67	1.25	7	0.78	12	1.34	12	1.34	>	19	2.12	10	1.12	7	0.78					
LYS	64	1.32	8	0.99	12	1.49	6	0.74	>	14	1.73	>	14	1.73	10	1.24				
ASN	53	1.48	>	13	2.18	>	14	2.35	6	1.01	8	1.34	9	1.51	3	0.50				
CYS	10	0.76	>	6	2.74	1	0.46	1	0.46	0	0.00	1	0.46	1	0.46					
GLN	34	1.09	1	0.19	9	1.73	8	1.54	3	0.58	>	10	1.93	3	0.58					
HIS	16	0.82	3	0.92	4	1.23	3	0.92	1	0.31	3	0.92	2	0.61						

SER	76	1.56	>	15	1.84	8	0.98	>	26	3.19	9	1.11	9	1.11	9	1.11	
THR	48	1.08		7	0.94	>	15	2.02	8	1.08	4	0.54	11	1.48	3	0.40	
TRP	10	0.83		3	1.49		2	0.99	0	0.00	1	0.50	1	0.50	3	1.49	
TYR	19	0.66		1	0.21		5	1.04	4	0.83	2	0.42	6	1.25	1	0.21	
GLY	30	0.54		8	0.87	<	1	0.11	9	0.97	<	2	0.22	<	2	0.22	
OTH	1	0.34		0	0.00		0	0.00	0	0.00		0	0.00	1	2.06	0	0.00
ALL	804	1.00		134	1.00		134	1.00	134	1.00		134	1.00	134	1.00	134	1.00

cluster 20: turn-type XVI

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6					
	f	P	f	P	f	P	f	f	P	f	P	f	P	f				
ALA	58	0.94	10	0.97	8	0.78	>	19	1.85	9	0.88	<	2	0.20	10	0.97		
ILE	24	0.50	5	0.63	<	2	0.25	10	1.26	3	0.38	<	0	0.00	4	0.50		
LEU	61	0.83	10	0.82	13	1.06	>	24	1.96	<	1	0.08	<	3	0.25	10	0.82	
MET	16	1.01	2	0.76	4	1.51		5	1.89	2	0.76		1	0.38	2	0.76		
PHE	28	0.81	5	0.87	4	0.69		3	0.52	<	1	0.17	<	1	0.17	>	14	2.43
PRO	64	1.74	9	1.47	3	0.49	<	1	0.16	>	50	8.15	<	0	0.00	<	1	0.16
VAL	35	0.61	4	0.42	<	3	0.31	10	1.05	6	0.63	<	0	0.00	12	1.26		
ARG	30	0.74	7	1.03	4	0.59		7	1.03	6	0.88		3	0.44	3	0.44		
ASP	44	0.93	11	1.39	12	1.52	<	2	0.25	10	1.27		7	0.89	<	2	0.25	
GLU	42	0.77	4	0.44	8	0.88		7	0.77	12	1.32		5	0.55	6	0.66		
LYS	45	0.92	4	0.49	13	1.59		12	1.47	8	0.98		3	0.37	5	0.61		
ASN	42	1.16	3	0.50	8	1.32		5	0.83	2	0.33	>	15	2.48	9	1.49		
CYS	17	1.27	>	7	3.15	0	0.00	2	0.90	2	0.90		0	0.00	>	6	2.70	
GLN	29	0.92	4	0.76	6	1.14	>	10	1.90	5	0.95		1	0.19	3	0.57		
HIS	15	0.76	2	0.60	5	1.51		2	0.60	3	0.91		2	0.60	1	0.30		
SER	68	1.37	>	25	3.02	>	16	1.94	7	0.85	4	0.48		6	0.73	10	1.21	
THR	40	0.89	10	1.33	>	16	2.13	4	0.53	4	0.53	<	0	0.00	6	0.80		
TRP	9	0.73	1	0.49	1	0.49		2	0.98	0	0.00		0	0.00	>	5	2.44	
TYR	17	0.58	<	0	0.00	3	0.62	1	0.21	3	0.62		1	0.21	9	1.85		
GLY	>	132	2.34	13	1.38	7	0.75	<	3	0.32	5	0.53	>	86	9.16	>	18	1.92
OTH	0	0.00	0	0.00	0	0.00		0	0.00	0	0.00		0	0.00	0	0.00		
ALL	816	1.00	136	1.00	136	1.00		136	1.00	136	1.00		136	1.00	136	1.00		

cluster 21: turn-type XVII

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6					
	f	P	f	P	f	P	f	f	P	f	P	f	P	f				
ALA	39	1.25	>	11	2.11	4	0.77	4	0.77	>	10	1.92	2	0.38	8	1.54		
ILE	13	0.54		3	0.74	1	0.25	5	1.24	1	0.25	<	0	0.00	3	0.74		
LEU	28	0.75		8	1.29	6	0.96	7	1.13	3	0.48	<	0	0.00	4	0.64		
MET	3	0.37		1	0.75	1	0.75	0	0.00	0	0.00		1	0.75	0	0.00		
PHE	9	0.51		2	0.68	0	0.00	2	0.68	0	0.00		2	0.68	3	1.03		
PRO	>	39	2.09	>	7	2.25	>	7	2.25	0	0.00	>	25	8.04	0	0.00		
VAL	14	0.48		4	0.83	3	0.62	5	1.03	<	0	0.00	<	0	0.00	2	0.41	
ARG	13	0.63		1	0.29	5	1.45	1	0.29	2	0.58		1	0.29	3	0.87		
ASP	24	1.00		4	1.00	5	1.25	1	0.25	3	0.75		7	1.75	4	1.00		
GLU	29	1.05		5	1.08	4	0.87	7	1.52	4	0.87		3	0.65	6	1.30		
LYS	26	1.04		3	0.72	>	9	2.17	4	0.96	8	1.92		2	0.48	<	0	0.00
ASN	17	0.92		1	0.33	3	0.98	0	0.00	2	0.65	>	9	2.94	2	0.65		

CYS	7	1.03	>	4	3.55	1	0.89	0	0.00	1	0.89	0	0.00	1	0.89			
GLN	19	1.18		2	0.75	3	1.12	>	6	2.24	2	0.75	1	0.37	5	1.87		
HIS	5	0.50		0	0.00	0	0.00		1	0.60	2	1.19	0	0.00	2	1.19		
SER	23	0.91		7	1.67	5	1.19		5	1.19	1	0.24	1	0.24	4	0.95		
THR	23	1.00		4	1.05	5	1.31		7	1.83	1	0.26	<	0	0.00	6	1.57	
TRP	3	0.48		0	0.00	1	0.96		1	0.96	1	0.96		0	0.00	0	0.00	
TYR	20	1.35		0	0.00	2	0.81	>	11	4.45	2	0.81		1	0.41	4	1.62	
GLY	>	59	2.06	2	0.42	4	0.84		1	0.21	1	0.21	>	39	8.19	>	12	2.52
OTH	1	0.67		0	0.00	0	0.00		1	3.99	0	0.00		0	0.00	0	0.00	
ALL	414	1.00		69	1.00	69	1.00		69	1.00	69	1.00		69	1.00	69	1.00	

cluster 22: turn-type XVIII

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6					
	f	P	f	P	f	P	f	f	P	f	P	f	P	f				
ALA	49	1.57	4	0.77	8	1.54	4	0.77	6	1.15	>	13	2.49	>	14	2.69		
ILE	10	0.41	6	1.49	3	0.74	<	0	0.00	1	0.25	<	0	0.00	<	0	0.00	
LEU	26	0.70	6	0.96	>	13	2.09	<	0	0.00	4	0.64		2	0.32	<	1	0.16
MET	13	1.62	3	2.24	1	0.75		0	0.00	3	2.24	2	1.49	>	4	2.99		
PHE	19	1.08	4	1.37	3	1.03		0	0.00	2	0.68	2	0.68	>	8	2.74		
PRO	8	0.43	2	0.64	3	0.96		0	0.00	2	0.64	1	0.32		0	0.00		
VAL	13	0.45	6	1.24	1	0.21	<	0	0.00	1	0.21	3	0.62		2	0.41		
ARG	11	0.53	3	0.87	1	0.29		0	0.00	2	0.58	2	0.58		3	0.87		
ASP	15	0.62	1	0.25	2	0.50	<	0	0.00	7	1.75	5	1.25	<	0	0.00		
GLU	29	1.05	6	1.30	5	1.08		1	0.22	>	9	1.95	6	1.30		2	0.43	
LYS	24	0.96	6	1.44	4	0.96	<	0	0.00	5	1.20	8	1.92		1	0.24		
ASN	11	0.60	3	0.98	1	0.33		0	0.00	3	0.98	2	0.65		2	0.65		
CYS	6	0.89	1	0.89	1	0.89		0	0.00	2	1.77	1	0.89		1	0.89		
GLN	18	1.12	3	1.12	5	1.87		1	0.37	2	0.75	>	6	2.24	1	0.37		
HIS	8	0.79	4	2.38	0	0.00		0	0.00	1	0.60	2	1.19		1	0.60		
SER	26	1.03	3	0.72	2	0.48		3	0.72	>	10	2.38	1	0.24	7	1.67		
THR	12	0.52	1	0.26	3	0.79		1	0.26	<	0	0.00	5	1.31	2	0.52		
TRP	7	1.12	0	0.00	3	2.89		0	0.00	2	1.93	1	0.96		1	0.96		
TYR	14	0.95	1	0.41	5	2.03		1	0.41	1	0.41	5	2.03		1	0.41		
GLY	>	94	3.29	6	1.26	4	0.84	>	58	12.18	6	1.26	2	0.42	>	18	3.78	
OTH	1	0.67	0	0.00	1	3.99		0	0.00	0	0.00	0	0.00		0	0.00		
ALL	414	1.00	69	1.00	69	1.00		69	1.00	69	1.00	69	1.00		69	1.00		

cluster 23: hook-1

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6								
	f	P	f	P	f	P	f	f	P	f	P	f	P	f							
ALA	<	866	0.78	<	138	0.75	<	137	0.74	<	61	0.33	>	211	1.14	192	1.04	<	127	0.69	
ILE	<	503	0.59		127	0.89	>	213	1.49	<	9	0.06	<	81	0.57	<	35	0.25	<	38	0.27
LEU	<	990	0.75	<	188	0.85	>	409	1.85	<	73	0.33	<	146	0.66	<	88	0.40	<	86	0.39
MET		257	0.90	<	32	0.67	>	102	2.15	<	17	0.36	<	31	0.65	<	21	0.44		54	1.14
PHE		524	0.84		121	1.17	>	186	1.79	<	38	0.37	<	64	0.62	<	54	0.52	<	61	0.59
PRO	>	808	1.22	>	136	1.23	>	132	1.20	<	41	0.37	>	460	4.17	<	34	0.31	<	5	0.05
VAL	<	701	0.68		192	1.12	>	282	1.64	<	18	0.11	<	106	0.62	<	39	0.23	<	64	0.37
ARG	<	535	0.73		103	0.84	<	81	0.66	<	65	0.53		120	0.98		102	0.83	<	64	0.52
ASP	>	1397	1.64	<	118	0.83	<	72	0.51	>	268	1.89	>	165	1.16	>	432	3.04	>	342	2.41

GLU	>1276	1.30	173	1.06	< 73	0.45	< 65	0.40	> 260	1.59	> 421	2.57	> 284	1.74
LYS	< 743	0.84	141	0.96	135	0.92	< 86	0.58	> 177	1.20	138	0.94	< 66	0.45
ASN	764	1.17	117	1.08	< 47	0.43	> 200	1.84	< 69	0.64	> 200	1.84	> 131	1.21
CYS	182	0.76	29	0.73	> 65	1.63	28	0.70	< 25	0.63	< 11	0.28	< 24	0.60
GLN	585	1.03	98	1.03	< 75	0.79	< 48	0.51	85	0.90	113	1.19	> 166	1.75
HIS	390	1.09	> 85	1.43	55	0.92	72	1.21	< 39	0.66	57	0.96	> 82	1.38
SER	>1176	1.32	< 113	0.76	< 67	0.45	> 496	3.34	155	1.04	> 207	1.39	138	0.93
THR	>1301	1.60	> 169	1.25	< 94	0.69	> 721	5.32	< 95	0.70	127	0.94	< 95	0.70
TRP	177	0.80	33	0.90	> 54	1.47	< 8	0.22	34	0.92	31	0.84	< 17	0.46
TYR	< 378	0.72	93	1.06	> 107	1.22	< 28	0.32	< 50	0.57	< 43	0.49	< 57	0.65
GLY	1061	1.05	> 231	1.37	< 32	0.19	< 102	0.60	< 67	0.40	< 92	0.55	> 537	3.18
OTH	60	1.13	9	1.01	> 28	3.15	< 2	0.23	6	0.68	8	0.90	7	0.79
ALL	14674	1.00	2446	1.00	2446	1.00	2446	1.00	2446	1.00	2445	1.00	2445	1.00

cluster 24: hook-2

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	37	1.15	> 10	1.87	5	0.93	9	1.68	6	1.12	2	0.37	5	0.93
ILE	22	0.88	5	1.21	7	1.69	< 0	0.00	2	0.48	4	0.97	4	0.97
LEU	33	0.86	10	1.56	6	0.94	3	0.47	3	0.47	< 0	0.00	11	1.72
MET	8	0.97	3	2.17	2	1.45	0	0.00	0	0.00	2	1.45	1	0.73
PHE	16	0.89	5	1.66	4	1.33	0	0.00	1	0.33	2	0.67	4	1.33
PRO	26	1.35	1	0.31	0	0.00	0	0.00	> 7	2.19	0	0.00	> 18	5.62
VAL	35	1.17	8	1.61	8	1.61	8	1.61	3	0.60	1	0.20	7	1.41
ARG	17	0.80	4	1.13	3	0.85	0	0.00	3	0.85	4	1.13	3	0.85
ASP	38	1.54	1	0.24	7	1.70	5	1.21	> 13	3.15	> 12	2.91	< 0	0.00
GLU	28	0.98	2	0.42	7	1.47	2	0.42	> 9	1.89	4	0.84	4	0.84
LYS	12	0.47	2	0.47	2	0.47	< 0	0.00	2	0.47	4	0.94	2	0.47
ASN	26	1.37	1	0.32	3	0.95	2	0.63	4	1.27	> 13	4.12	3	0.95
CYS	9	1.29	0	0.00	1	0.86	> 4	3.45	1	0.86	3	2.59	0	0.00
GLN	6	0.36	2	0.73	0	0.00	1	0.36	0	0.00	2	0.73	1	0.36
HIS	13	1.25	1	0.58	2	1.16	3	1.73	2	1.16	3	1.73	2	1.16
SER	32	1.24	1	0.23	7	1.62	> 14	3.24	6	1.39	4	0.93	< 0	0.00
THR	25	1.06	3	0.76	1	0.25	> 10	2.54	3	0.76	5	1.27	3	0.76
TRP	5	0.78	3	2.81	1	0.94	1	0.94	0	0.00	0	0.00	0	0.00
TYR	21	1.38	> 7	2.76	4	1.57	3	1.18	2	0.79	4	1.57	1	0.39
GLY	17	0.58	2	0.41	1	0.20	6	1.22	4	0.82	2	0.41	2	0.41
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	426	1.00	71	1.00	71	1.00	71	1.00	71	1.00	71	1.00	71	1.00

cluster 25: hook-3

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	< 146	0.48	< 32	0.63	< 31	0.61	< 31	0.61	< 19	0.37	< 3	0.06	< 30	0.59
ILE	243	1.03	> 73	1.86	> 52	1.33	> 77	1.96	< 0	0.00	< 0	0.00	41	1.05
LEU	< 205	0.56	70	1.16	50	0.83	49	0.81	< 7	0.12	< 3	0.05	< 26	0.43
MET	59	0.75	15	1.15	17	1.30	8	0.61	< 5	0.38	< 1	0.08	13	1.00
PHE	158	0.93	> 48	1.69	37	1.30	> 45	1.58	< 14	0.49	< 4	0.14	< 10	0.35
PRO	< 26	0.14	< 14	0.46	< 12	0.40	< 0	0.00	< 0	0.00	< 0	0.00	< 0	0.00

VAL	341	1.21	>	94	1.99	>	63	1.34	>	108	2.29	<	1	0.02	<	3	0.06	>	72	1.53	
ARG	179	0.89		30	0.89		28	0.83		34	1.01		25	0.74	<	8	0.24	>	54	1.61	
ASP	271	1.16	<	13	0.33	<	22	0.56	<	19	0.49	>	155	3.97		49	1.25	<	13	0.33	
GLU	246	0.91	<	23	0.51	>	58	1.29		44	0.98		43	0.96	<	11	0.25	>	67	1.49	
LYS	>	320	1.32	<	28	0.69		42	1.04	>	56	1.38		35	0.86	<	14	0.35	>	145	3.58
ASN	>	341	1.90	<	11	0.37		30	1.00	<	13	0.44	>	210	7.03	>	56	1.88		21	0.70
CYS		42	0.64		9	0.82		13	1.18		11	1.00	<	3	0.27	<	1	0.09		5	0.46
GLN		153	0.98	<	13	0.50		21	0.81		27	1.04		29	1.11	<	9	0.35	>	54	2.07
HIS		103	1.05		20	1.22		10	0.61	>	26	1.59		20	1.22	<	4	0.24		23	1.41
SER		176	0.72		32	0.78		48	1.18	<	28	0.69	<	21	0.51	<	18	0.44		29	0.71
THR	<	152	0.68		36	0.97	>	60	1.61	<	17	0.46	<	0	0.00	<	4	0.11		35	0.94
TRP		44	0.73	>	21	2.08		9	0.89		11	1.09	<	1	0.10	<	0	0.00	<	2	0.20
TYR		189	1.31	>	42	1.75	>	53	2.20	>	63	2.62	<	11	0.46	<	4	0.17		16	0.67
GLY	>	630	2.26		44	0.95	<	14	0.30	<	5	0.11	>	71	1.53	>	480	10.35	<	16	0.35
OTH		8	0.55		4	1.64		2	0.82		0	0.00		2	0.82		0	0.00		0	0.00
ALL		4032	1.00		672	1.00		672	1.00		672	1.00		672	1.00		672	1.00		672	1.00

cluster 26: hook-4

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6								
	f	P	f	P	f	P	f	f	P	f	P	f	P	f							
ALA	222	0.77	44	0.91	39	0.81	<	2	0.04	>	70	1.45	42	0.87	<	25	0.52				
ILE	<	154	0.69	48	1.29	>	68	1.82	<	1	0.03	<	14	0.38	<	8	0.21	<	15	0.40	
LEU	<	212	0.61	65	1.13	>	80	1.39	<	5	0.09	<	25	0.43	<	19	0.33	<	18	0.31	
MET		57	0.77	10	0.81	>	22	1.77	<	1	0.08	<	5	0.40	<	5	0.40		14	1.13	
PHE		135	0.83	27	1.00	>	56	2.07	<	0	0.00	<	14	0.52		18	0.67		20	0.74	
PRO	<	86	0.50	19	0.66		31	1.08	<	3	0.10		22	0.76	<	8	0.28	<	3	0.10	
VAL	<	193	0.72	>	66	1.47	>	73	1.63	<	2	0.05	<	20	0.45	<	19	0.42	<	13	0.29
ARG	<	126	0.66	<	20	0.63		25	0.78	<	5	0.16		29	0.91		37	1.16	<	10	0.31
ASP	>	444	1.99		30	0.81	<	13	0.35	>	216	5.82	>	76	2.05	>	61	1.64		48	1.29
GLU		301	1.17		46	1.08	<	22	0.51	<	7	0.16	>	101	2.36	>	86	2.01		39	0.91
LYS		234	1.01		35	0.91		44	1.14	<	6	0.16	>	60	1.56	>	71	1.84	<	18	0.47
ASN	>	283	1.66	<	14	0.49	<	13	0.46	>	129	4.54		28	0.99	>	67	2.36		32	1.13
CYS		50	0.80		10	0.96		14	1.34		6	0.57	<	3	0.29		5	0.48		12	1.15
GLN		117	0.79		17	0.69		17	0.69	<	3	0.12		26	1.05		33	1.33		21	0.85
HIS		102	1.09		15	0.96		10	0.64	<	5	0.32		18	1.16	>	30	1.93	>	24	1.54
SER		267	1.15		35	0.90	<	26	0.67	>	59	1.52	>	62	1.60		50	1.29		35	0.90
THR		231	1.09		28	0.79		31	0.88	>	70	1.98	<	21	0.59		36	1.02		45	1.27
TRP		27	0.47		6	0.62		7	0.73	<	1	0.10		5	0.52		4	0.42		4	0.42
TYR		113	0.82	>	35	1.53		22	0.96	<	4	0.18	<	10	0.44		25	1.09		17	0.74
GLY	>	461	1.74	>	66	1.50	<	19	0.43	>	110	2.49	<	29	0.66	<	15	0.34	>	222	5.03
OTH		19	1.37		3	1.29	>	7	3.02		4	1.73		1	0.43		0	0.00		4	1.73
ALL		3834	1.00		639	1.00		639	1.00		639	1.00		639	1.00		639	1.00		639	1.00

cluster 27: hook-5

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6							
	f	P	f	P	f	P	f	f	P	f	P	f	P	f						
ALA	110	0.67	29	1.05	<	16	0.58	26	0.94	<	3	0.11	<	16	0.58	20	0.73			
ILE	97	0.76	>	35	1.64		19	0.89	24	1.13	<	1	0.05	<	3	0.14	15	0.70		
LEU	<	125	0.63	>	45	1.37		24	0.73	27	0.82	<	3	0.09	<	6	0.18	<	20	0.61

TYR	9	0.65	1	0.43	1	0.43	0	0.00	3	1.29	0	0.00	4	1.72
GLY	40	1.49	2	0.45	3	0.67	1	0.22	> 32	7.13	1	0.22	1	0.22
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	390	1.00	65	1.00	65	1.00	65	1.00	65	1.00	65	1.00	65	1.00

cluster 3: turn-type IIb

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	1	0.32	0	0.00	0	0.00	0	0.00	0	0.00	1	1.89	0	0.00
ILE	1	0.41	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	2.45
LEU	1	0.26	0	0.00	1	1.58	0	0.00	0	0.00	0	0.00	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
VAL	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ARG	2	0.95	0	0.00	0	0.00	0	0.00	1	2.86	1	2.86	0	0.00
ASP	7	2.87	1	2.46	1	2.46	> 5	12.30	0	0.00	0	0.00	0	0.00
GLU	1	0.36	0	0.00	0	0.00	0	0.00	0	0.00	1	2.13	0	0.00
LYS	6	2.37	0	0.00	> 2	4.74	0	0.00	1	2.37	> 2	4.74	1	2.37
ASN	3	1.61	> 3	9.64	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
CYS	1	1.46	0	0.00	0	0.00	> 1	8.74	0	0.00	0	0.00	0	0.00
GLN	2	1.23	0	0.00	0	0.00	0	0.00	0	0.00	1	3.69	1	3.69
HIS	1	0.98	0	0.00	0	0.00	0	0.00	> 1	5.86	0	0.00	0	0.00
SER	4	1.57	0	0.00	1	2.35	1	2.35	1	2.35	0	0.00	1	2.35
THR	4	1.72	0	0.00	1	2.58	0	0.00	0	0.00	1	2.58	> 2	5.16
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	1	0.67	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	3.99
GLY	7	2.41	> 3	6.21	1	2.07	0	0.00	> 3	6.21	0	0.00	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	42	1.00	7	1.00	7	1.00	7	1.00	7	1.00	7	1.00	7	1.00

cluster 4: turn-type IIc

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	5	1.58	1	1.89	> 2	3.78	1	1.89	0	0.00	1	1.89	0	0.00
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LEU	1	0.26	0	0.00	1	1.58	0	0.00	0	0.00	0	0.00	0	0.00
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PRO	1	0.53	0	0.00	1	3.17	0	0.00	0	0.00	0	0.00	0	0.00
VAL	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ARG	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ASP	4	1.64	1	2.46	0	0.00	> 3	7.38	0	0.00	0	0.00	0	0.00
GLU	1	0.36	0	0.00	0	0.00	0	0.00	0	0.00	1	2.13	0	0.00
LYS	2	0.79	0	0.00	0	0.00	1	2.37	0	0.00	1	2.37	0	0.00
ASN	4	2.14	> 2	6.43	0	0.00	0	0.00	0	0.00	> 2	6.43	0	0.00
CYS	1	1.46	0	0.00	0	0.00	0	0.00	> 1	8.74	0	0.00	0	0.00
GLN	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
HIS	1	0.98	0	0.00	> 1	5.86	0	0.00	0	0.00	0	0.00	0	0.00

SER	6	2.35	>	2	4.70	1	2.35	>	2	4.70	0	0.00	0	0.00	1	2.35				
THR	4	1.72		1	2.58	1	2.58		0	0.00	0	0.00	>	2	5.16	0	0.00			
TRP	0	0.00		0	0.00	0	0.00		0	0.00	0	0.00		0	0.00	0	0.00			
TYR	1	0.67		0	0.00	0	0.00		0	0.00	0	0.00		0	0.00	1	3.99			
GLY	>	11	3.79		0	0.00	0	0.00		0	0.00	>	6	12.42	0	0.00	>	5	10.35	
OTH	0	0.00		0	0.00	0	0.00		0	0.00	0	0.00		0	0.00	0	0.00		0	0.00
ALL	42	1.00		7	1.00	7	1.00		7	1.00	7	1.00		7	1.00	7	1.00		7	1.00

cluster 5: turn-type III

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6			
	f	P	f	P	f	P	f	f	P	f	P	f	P	f		
ALA	8	0.77	0	0.00	1	0.58	3	1.73	1	0.58	1	0.58	2	1.15		
ILE	7	0.87	1	0.74	1	0.74	1	0.74	1	0.74	0	0.00	3	2.23		
LEU	4	0.32	0	0.00	0	0.00	1	0.48	2	0.96	1	0.48	0	0.00		
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		
PHE	4	0.68	1	1.03	1	1.03	1	1.03	0	0.00	1	1.03	0	0.00		
PRO	7	1.13	0	0.00	>	5	4.82	0	0.00	0	0.00	2	1.93	0	0.00	
VAL	6	0.62	1	0.62	0	0.00	2	1.24	2	1.24	0	0.00	1	0.62		
ARG	7	1.01	0	0.00	2	1.74	1	0.87	2	1.74	2	1.74	0	0.00		
ASP	14	1.75	>	5	3.74	1	0.75	2	1.50	2	1.50	2	1.50	2	1.50	
GLU	9	0.98	2	1.30	2	1.30	1	0.65	1	0.65	2	1.30	1	0.65		
LYS	13	1.56	0	0.00	>	4	2.89	3	2.17	1	0.72	2	1.44	3	2.17	
ASN	10	1.63	>	3	2.94	1	0.98	0	0.00	1	0.98	>	4	3.91	1	0.98
CYS	3	1.33	1	2.66	0	0.00	1	2.66	0	0.00	0	0.00	1	2.66		
GLN	6	1.12	1	1.12	0	0.00	0	0.00	1	1.12	2	2.24	2	2.24		
HIS	4	1.19	2	3.57	1	1.79	1	1.79	0	0.00	0	0.00	0	0.00		
SER	16	1.91	2	1.43	>	4	2.86	1	0.72	>	5	3.58	1	0.72	3	2.15
THR	5	0.65	1	0.79	0	0.00	0	0.00	3	2.36	0	0.00	1	0.79		
TRP	3	1.45	1	2.89	0	0.00	0	0.00	0	0.00	1	2.89	1	2.89		
TYR	7	1.42	1	1.22	0	0.00	>	4	4.86	1	1.22	0	0.00	1	1.22	
GLY	5	0.53	1	0.63	0	0.00	1	0.63	0	0.00	2	1.26	1	0.63		
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		
ALL	138	1.00	23	1.00	23	1.00	23	1.00	23	1.00	23	1.00	23	1.00		

cluster 6: turn-type IV

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6			
	f	P	f	P	f	P	f	f	P	f	P	f	P	f		
ALA	4	0.49	0	0.00	1	0.74	0	0.00	0	0.00	1	0.74	2	1.47		
ILE	5	0.79	>	4	3.81	0	0.00	0	0.00	1	0.95	0	0.00	0	0.00	
LEU	6	0.62	2	1.23	1	0.62	0	0.00	1	0.62	0	0.00	2	1.23		
MET	1	0.48	0	0.00	0	0.00	1	2.86	0	0.00	0	0.00	0	0.00		
PHE	5	1.09	1	1.31	0	0.00	1	1.31	0	0.00	0	0.00	>	3	3.93	
PRO	9	1.85	0	0.00	>	8	9.86	0	0.00	0	0.00	1	1.23	0	0.00	
VAL	4	0.53	3	2.37	0	0.00	0	0.00	0	0.00	0	0.00	1	0.79		
ARG	8	1.48	1	1.11	1	1.11	0	0.00	>	3	3.33	1	1.11	2	2.22	
ASP	7	1.12	0	0.00	0	0.00	1	0.96	0	0.00	>	6	5.74	0	0.00	
GLU	11	1.52	2	1.66	2	1.66	0	0.00	1	0.83	3	2.49	3	2.49		
LYS	9	1.38	3	2.77	3	2.77	0	0.00	1	0.92	1	0.92	1	0.92		
ASN	7	1.46	0	0.00	0	0.00	>	3	3.75	0	0.00	>	3	3.75	1	1.25

CYS	2	1.13	1	3.40	1	3.40	0	0.00	0	0.00	0	0.00	0	0.00
GLN	1	0.24	0	0.00	0	0.00	1	1.43	0	0.00	0	0.00	0	0.00
HIS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
SER	11	1.68	1	0.91	0	0.00	1	0.91	> 7	6.40	1	0.91	1	0.91
THR	5	0.84	0	0.00	0	0.00	0	0.00	> 4	4.01	0	0.00	1	1.00
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TYR	1	0.26	0	0.00	1	1.55	0	0.00	0	0.00	0	0.00	0	0.00
GLY	12	1.61	0	0.00	0	0.00	> 10	8.05	0	0.00	1	0.81	1	0.81
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	108	1.00	18	1.00	18	1.00	18	1.00	18	1.00	18	1.00	18	1.00

cluster 7: turn-type V

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	11	1.28	> 5	3.48	2	1.39	1	0.70	0	0.00	2	1.39	1	0.70
ILE	6	0.90	1	0.90	0	0.00	0	0.00	0	0.00	1	0.90	> 4	3.61
LEU	8	0.78	4	2.34	2	1.17	1	0.58	0	0.00	1	0.58	0	0.00
MET	2	0.90	1	2.71	0	0.00	0	0.00	0	0.00	1	2.71	0	0.00
PHE	1	0.21	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	1.24
PRO	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
VAL	8	1.00	0	0.00	1	0.75	1	0.75	0	0.00	1	0.75	> 5	3.75
ARG	5	0.88	1	1.05	0	0.00	0	0.00	0	0.00	> 3	3.16	1	1.05
ASP	7	1.06	0	0.00	> 4	3.62	2	1.81	0	0.00	0	0.00	1	0.91
GLU	7	0.92	0	0.00	0	0.00	0	0.00	1	0.79	2	1.57	> 4	3.15
LYS	5	0.73	0	0.00	0	0.00	2	1.75	0	0.00	3	2.62	0	0.00
ASN	7	1.38	0	0.00	2	2.37	2	2.37	2	2.37	1	1.18	0	0.00
CYS	1	0.54	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	3.22
GLN	4	0.91	1	1.36	0	0.00	0	0.00	1	1.36	2	2.72	0	0.00
HIS	1	0.36	0	0.00	1	2.16	0	0.00	0	0.00	0	0.00	0	0.00
SER	4	0.58	1	0.87	1	0.87	0	0.00	1	0.87	1	0.87	0	0.00
THR	5	0.79	3	2.85	1	0.95	0	0.00	0	0.00	0	0.00	1	0.95
TRP	1	0.58	0	0.00	0	0.00	1	3.50	0	0.00	0	0.00	0	0.00
TYR	5	1.23	0	0.00	2	2.94	1	1.47	1	1.47	1	1.47	0	0.00
GLY	> 26	3.30	2	1.53	3	2.29	> 8	6.10	> 13	9.91	0	0.00	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ALL	114	1.00	19	1.00	19	1.00	19	1.00	19	1.00	19	1.00	19	1.00

cluster 8: turn-type VI

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6	
	f	P	f	P	f	P	f	f	P	f	P	f	P	f
ALA	5	1.00	0	0.00	0	0.00	1	1.20	2	2.41	2	2.41	0	0.00
ILE	1	0.26	0	0.00	0	0.00	0	0.00	1	1.56	0	0.00	0	0.00
LEU	1	0.17	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	1.01
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PRO	4	1.34	0	0.00	> 4	8.07	0	0.00	0	0.00	0	0.00	0	0.00
VAL	4	0.86	0	0.00	0	0.00	> 3	3.89	1	1.30	0	0.00	0	0.00
ARG	3	0.91	0	0.00	0	0.00	1	1.82	0	0.00	> 2	3.64	0	0.00
ASP	4	1.04	> 4	6.26	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00

GLU	3	0.68	0	0.00	1	1.36	2	2.72	0	0.00	0	0.00	0	0.00			
LYS	7	1.76	1	1.51	2	3.02	1	1.51	2	3.02	1	1.51	0	0.00			
ASN	10	3.41	>	2	4.09	1	2.05	1	2.05	0	0.00	>	4	8.18	>	2	4.09
CYS	3	2.78	1	5.56	0	0.00	0	0.00	>	2	11.12	0	0.00	0	0.00		
GLN	3	1.17	0	0.00	0	0.00	>	2	4.69	0	0.00	1	2.35	0	0.00		
HIS	2	1.24	1	3.73	1	3.73	0	0.00	0	0.00	0	0.00	0	0.00			
SER	4	1.00	0	0.00	1	1.50	0	0.00	1	1.50	0	0.00	2	2.99			
THR	5	1.37	1	1.64	1	1.64	0	0.00	2	3.28	0	0.00	1	1.64			
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00			
TYR	1	0.42	1	2.54	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00			
GLY	6	1.32	0	0.00	0	0.00	0	0.00	0	0.00	1	1.32	>	5	6.58		
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00			
ALL	66	1.00	11	1.00	11	1.00	11	1.00	11	1.00	11	1.00	11	1.00			

cluster 9: turn-type VII

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6		
	f	P	f	P	f	P	f	f	P	f	P	f	P	f	
ALA	6	1.32	1	1.32	1	1.32	0	0.00	0	0.00	>	3	3.97	1	1.32
ILE	1	0.28	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	1.71	
LEU	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	
MET	1	0.86	0	0.00	1	5.15	0	0.00	0	0.00	0	0.00	0	0.00	
PHE	1	0.39	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	2.36	
PRO	7	2.59	0	0.00	0	0.00	0	0.00	>	7	15.53	0	0.00	0	0.00
VAL	1	0.24	0	0.00	0	0.00	1	1.43	0	0.00	0	0.00	0	0.00	
ARG	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	
ASP	6	1.72	0	0.00	2	3.44	2	3.44	1	1.72	0	0.00	1	1.72	
GLU	4	1.00	0	0.00	1	1.49	0	0.00	0	0.00	2	2.99	1	1.49	
LYS	5	1.38	1	1.66	1	1.66	1	1.66	1	1.66	0	0.00	1	1.66	
ASN	5	1.88	0	0.00	1	2.25	>	2	4.50	0	0.00	1	2.25	1	2.25
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	
GLN	1	0.43	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	2.58	
HIS	3	2.05	0	0.00	0	0.00	0	0.00	0	0.00	>	2	8.21	1	4.11
SER	7	1.92	1	1.65	2	3.29	>	3	4.93	0	0.00	0	0.00	1	1.65
THR	4	1.20	>	3	5.42	0	0.00	0	0.00	0	0.00	1	1.81	0	0.00
TRP	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	
TYR	2	0.93	0	0.00	1	2.79	0	0.00	0	0.00	1	2.79	0	0.00	
GLY	6	1.45	>	4	5.79	0	0.00	1	1.45	1	1.45	0	0.00	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	
ALL	60	1.00	10	1.00	10	1.00	10	1.00	10	1.00	10	1.00	10	1.00	

cluster 10: turn-type VII

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6		
	f	P	f	P	f	P	f	f	P	f	P	f	P	f	
ALA	2	0.63	>	2	3.78	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
ILE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	
LEU	2	0.53	2	3.17	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	
MET	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	
PHE	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	
PRO	1	0.53	0	0.00	1	3.17	0	0.00	0	0.00	0	0.00	0	0.00	

VAL	5	1.70	0	0.00	1	2.04	1	2.04	0	0.00	0	0.00	>	3	6.11	
ARG	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00	
ASP	1	0.41	0	0.00	0	0.00	0	0.00	0	0.00	1	2.46		0	0.00	
GLU	2	0.71	1	2.13	1	2.13	0	0.00	0	0.00	0	0.00		0	0.00	
LYS	1	0.40	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		1	2.37	
ASN	4	2.14	>	2	6.43	0	0.00	0	0.00	0	0.00	1	3.22	1	3.22	
CYS	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00	
GLN	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00	
HIS	4	3.91	0	0.00	>	3	17.59	0	0.00	0	0.00	0	0.00	>	1	5.86
SER	4	1.57	0	0.00	0	0.00	0	0.00	>	3	7.05	1	2.35	0	0.00	
THR	2	0.86	0	0.00	0	0.00	0	0.00	>	2	5.16	0	0.00	0	0.00	
TRP	3	4.75	0	0.00	>	1	9.50	0	0.00	>	2	19.00	0	0.00	0	0.00
TYR	2	1.33	0	0.00	0	0.00	0	0.00	0	0.00	1	3.99	1	3.99		
GLY	9	3.10	0	0.00	0	0.00	>	6	12.42	0	0.00	>	3	6.21	0	0.00
OTH	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00		0	0.00	
ALL	42	1.00	7	1.00	7	1.00	7	1.00	7	1.00	7	1.00		7	1.00	

cluster 11: turn-type IX

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6					
	f	P	f	P	f	P	f	f	P	f	P	f	P	f				
ALA	7	2.21	1	1.89	1	1.89	>	2	3.78	>	2	3.78	0	0.00	1	1.89		
ILE	1	0.41	0	0.00	0	0.00		1	2.45	0	0.00	0	0.00	0	0.00			
LEU	3	0.79	2	3.17	0	0.00		0	0.00	0	0.00	0	0.00	1	1.58			
MET	0	0.00	0	0.00	0	0.00		0	0.00	0	0.00	0	0.00	0	0.00			
PHE	1	0.56	0	0.00	0	0.00		1	3.37	0	0.00	0	0.00	0	0.00			
PRO	3	1.58	0	0.00	1	3.17		0	0.00	>	2	6.34	0	0.00	0	0.00		
VAL	3	1.02	0	0.00	0	0.00		1	2.04	0	0.00	0	0.00	>	2	4.07		
ARG	0	0.00	0	0.00	0	0.00		0	0.00	0	0.00	0	0.00	0	0.00			
ASP	1	0.41	0	0.00	0	0.00		0	0.00	1	2.46	0	0.00	0	0.00			
GLU	3	1.07	0	0.00	>	2	4.27		0	0.00	0	0.00	0	0.00	1	2.13		
LYS	2	0.79	0	0.00	1	2.37		0	0.00	1	2.37	0	0.00	0	0.00			
ASN	0	0.00	0	0.00	0	0.00		0	0.00	0	0.00	0	0.00	0	0.00			
CYS	0	0.00	0	0.00	0	0.00		0	0.00	0	0.00	0	0.00	0	0.00			
GLN	1	0.61	0	0.00	0	0.00		0	0.00	0	0.00	1	3.69	0	0.00			
HIS	3	2.93	>	1	5.86	>	1	5.86		0	0.00	>	1	5.86	0	0.00		
SER	4	1.57	1	2.35	1	2.35	>	2	4.70	0	0.00	0	0.00	0	0.00			
THR	3	1.29	1	2.58	0	0.00		0	0.00	1	2.58	0	0.00	1	2.58			
TRP	0	0.00	0	0.00	0	0.00		0	0.00	0	0.00	0	0.00	0	0.00			
TYR	1	0.67	0	0.00	0	0.00		0	0.00	0	0.00	0	0.00	1	3.99			
GLY	6	2.07	1	2.07	0	0.00		0	0.00	0	0.00	>	5	10.35	0	0.00		
OTH	0	0.00	0	0.00	0	0.00		0	0.00	0	0.00	0	0.00	0	0.00			
ALL	42	1.00	7	1.00	7	1.00		7	1.00	7	1.00	7	1.00	7	1.00			

cluster 12: turn-type X

aa	turn		position 1		position 2		position 3		position 4		position 5		position 6			
	f	P	f	P	f	P	f	f	P	f	P	f	P	f		
ALA	2	0.74	0	0.00	0	0.00	0	0.00	0	0.00	>	2	4.41	0	0.00	
ILE	4	1.90	1	2.85	0	0.00	1	2.85	0	0.00	0	0.00	>	2	5.71	
LEU	1	0.31	0	0.00	1	1.85	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00

Lebenslauf

Persönliche Daten

Name: Oliver Koch
Geburtsdatum und -ort: 10.01.1977 in Herford, Deutschland
Nationalität: Deutsch
Familienstand: ledig

Schulische und Universitäre Ausbildung

04/2003 – 09/2007 Doktorand
Arbeitskreis Prof. Dr. G. Klebe
Pharmazeutisches Institut, Phillips-Universität Marburg
Forschungsgruppe für Wirkstoffdesign, Biocomputing und Röntgenkristallographie

04/2003 – 03/2006 Vertiefung der Informatikkenntnisse
im Rahmen eines Aufbaustudiums Informatik

01/2004 Vorlage einer Diplomarbeit am Fachbereich Pharmazie
der Martin-Luther-Universität in Halle-Wittenberg
mit Erhalt des Akademischen Grad „Diplom-Pharmazeut“

01/2003 Abschluss der Pharmazeutischen Prüfung
mit anschließender Approbation als Apotheker

01/2002 – 06/2002 Praktikum im Rahmen der Pharmazeutischen Ausbildung
Arbeitskreis Prof. Dr. G. Klebe
Pharmazeutisches Institut, Phillips-Universität Marburg
Untersuchungen zur Anfertigung einer Diplomarbeit:
„Ein Werkzeug zur Korrelation von sekundären Strukturmustern
mit der Ligandenbindung: Erweiterung der Seabase“

1997 - 2001 Studium der Pharmazie
Philipps-Universität Marburg, Marburg (Lahn)

1987 - 1996 Gymnasium im Schulzentrum Aspe, Bad Salzuflen
Abschluß: Abitur

Berufliche Erfahrung

08/2008 – heute Post-Doktorarbeit im Bereich Drug Discovery /
BioChemInformatics
Intervet Innovation GmbH, Schwabenheim, Deutschland

10/2007 – 07/2008	Scientific Software Engineer The Cambridge Crystallographic Data Centre, Cambridge, UK
04/2003 – 09/2007	Wissenschaftlicher Mitarbeiter Arbeitskreis Prof. Dr. G. Klebe Pharmazeutisches Institut, Phillips-Universität Marburg
02/2003 – 03/2003	Tätigkeit als Apotheker Brandes'sche Apotheke, Bad Salzufen
07/2002 – 12/2002	Praktikum im Rahmen der Pharmazeutischen Ausbildung Neustädter Apotheke, Herford
07/1996 – 07/1997	Zivildienst in der Klinik am Kurpark, Bad Salzufen Betreuung von Krebs- und Herzinfarktpatienten im Rahmen ihrer Rehabilitationsmaßnahmen

Forschungsaufenthalt im Rahmen dieser Arbeit

07/2006 – 08/2006	The Cambridge Crystallographic Data Centre, Cambridge, UK Implementierung einer Erweiterung der Relibase um Zugriff auf die Sekundär-Strukturdaten über ein integriertes Web-Interface zu erhalten
-------------------	---

ERKLÄRUNG

Ich versichere, daß ich meine Dissertation

„Secbase: A novel tool to correlate secondary structure elements with ligand binding“

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den

.....
(Oliver Koch)