Ongoing sequencing projects, such as the human genome project, have produced enormous amounts of biological sequences so far. The experimental techniques for determining protein structure, such as X-ray diffraction or Nuclear Magnetic Resonance methods, remain slow and laborious, and do not scale up to current sequencing speeds. Thus the gap between known sequences and known structure are increasing exponentially (Rost, 2001). There is need to develop computational methods for predicting tertiary structure of proteins whose sequence is known but structure is unknown. Broadly, structure prediction techniques can be divided in two categories; i) knowledge based methods (e.g. comparative or homology modeling; fold recognition) and ii) ab initio methods. The comparative modeling is the most reliable technique, when a protein has high sequence similarity with protein whose structure is known. This technique fails in absence of sequence similarity. Fold recognition methods allow searching of similar structures even when there is remote homology between proteins (Bowie et al., 1991; Jones et al., 1992; Lemer et al., 1995). The ab initio methods are based on principles of energy minimization. These methods are computer and resource intensive; it is nearly impossible to predict tertiary structure of protein from scratch using these methods. By far the most commonly used ab initio prediction methods are aimed at the prediction of secondary structural elements in proteins (Lim, 1974; Chou and Fasman, 1974; Garnier et al., 1978; Zvelebil et al., 1987; Rost and Sander, 1993a; Geourjon and Deleage, 1995; Salamov and Solovyev, 1995; Frishman and Argos, 1996; King and Sternberg, 1996). Since secondary structure prediction is a one-dimensional problem, it is considerable less complicated than full prediction.
Review of Literature

Protein secondary structure prediction has been around for almost a quarter of a century. Various algorithms have been used for secondary structure prediction from sequence alone. These include simple linear statistics (Garnier et al., 1978; Chou and Fasman, 1974; Garnier et al., 1996; Gibrat et al., 1987; Nishikawa and Nogughi, 1995), physicochemical properties (Lim, 1974), neural networks (Qian and Sejnowski, 1988; Rost and Sander, 1993a; Holley and Karplus, 1989; Kneller et al., 1990; Riis and Krogh, 1996; Chandonia and Karplus, 1996; Chandonia and Karplus, 1999), k-way nearest neighbors (Salamov and Solovyev, 1995; Rychlewski and Godzik, 1997; Frishman and Argos, 1996; Yi and Lander, 1993; Levin, 1997), evolutionary trees (Goldman et al., 1996; Jones, 1999), simple residue substitution matrices (Mehta et al., 1995) and combination of different methods with consensus approaches (Zimmerman and Gibrat, 1998; Cuff and Barton, 1999; Biou et al., 1995; Guermeur et al., 1999). The performance of all these methods is assessed time to time at CASP (Critical Assessment of Structure Prediction, http://predictioncenter.llnl.gov/), which provides the benchmarking for the different secondary structure prediction methods (Moult et al., 1997; 1998).

All the existing secondary structure prediction methods predict only two states - α-helix (H) and β-sheet (E) and rest as coil (C) as shown in Fig. 2.1. They do not provide any information about irregular secondary structures, such as tight turns despite the fact that turns play important roles both structurally and functionally and constitute a considerable percentage of protein residues.

Thus, there is a need to develop prediction methods for tight turns, which can aid in further understanding of protein folding and stability and will be an intermediate step in tertiary structure prediction. Before developing methods for prediction of tight turns, a detailed knowledge of their structures and geometries is required.
2.1 Tight turns

The meaning of the word “fold” describes a fundamental property without which proteins would not assume compact, globular shapes and would therefore remain extended. The ability to reverse direction abruptly is a critical element in this regard. A limited repertoire of structures has the capability to change the chain direction. These structures are named as “tight turns”, which are one of the fundamental classes of protein secondary structure (Chou, 2000). Tight turns play an important role in proteins and have been implicated in protein folding and molecular recognition (Rose et al., 1985; Takano et al., 2000). Unlike α-helices and β-sheets, turns are not as conspicuous because their backbone torsion angles are non-repeating. Their frequent occurrence is responsible for the globularity of proteins and connects regular secondary structure elements. Turns are usually situated at the protein surface (Kuntz, 1972). This topographical tendency is a consequence of the hydrogen-bonding requirements of backbone polar groups. Residues found in turns are predominantly the polar (Lewis et al., 1971; Kuntz, 1972; Rose, 1978). The surface localization of turns in proteins and the predominance in turns of amino acids bearing potentially reactive functional groups in their side chains (e.g., Asn, Ser, Pro, Thr and Lys) have led to the suggestion that turns function as recognition sites for complex immunological, metabolic, genomic and endocrinology regulatory mechanisms (Smith and Pease, 1980). They also provide very useful information for defining template structures for the design of new molecules such as drugs, pesticides, and antigens (Chou, 2000).

2.1.1 Types of tight turns

Depending on the number of residues forming the turn, tight turns are classified as following (Chou, 2000):

δ-turn: It is the smallest tight turns, which involve only two amino acid residues, and is also called 1→2 type, 2→3 type, or C₈ form. The intraturn hydrogen bond for a δ-turn is formed between the backbone NH of i residue and the backbone CO of i+1 residue.


Review of Literature

**γ-turn:** The next smallest tight turn is γ-turn, which involves three amino acid residues. The intraturn hydrogen bond for a γ-turn is formed between the backbone CO of \(i\) residue and the backbone NH of \(i+2\) residue.

**β-turn:** A β-turn involves four amino acids residues where the distance between \(C^\alpha(i)\) and \(C^\alpha(i+3)\) is less than 7Å and the tetrapeptide chain is not in a helical conformation (Richardson, 1981; Rose et al., 1985). It may have an intraturn hydrogen bond between the backbone CO of \(i\) residue and the backbone NH of \(i+3\) residue. It is the most predominant tight turn.

**α-turn:** An α-turn involves five amino acid residues where the distance between \(C^\alpha(i)\) and \(C^\alpha(i+4)\) is less than 7Å and the pentapeptide chain is not in a helical conformation.

**π-turn:** The largest tight turn is a π-turn, which involves six amino acid residues.

In general, a tight turn can be formulated as n-turn hydrogen bond, i.e., the hydrogen bond between the backbone CO of \(i\) residue and the backbone NH of \(i+n\) residue, where \(n\) varies between 1 to 5, i.e.,

\[
H\text{-bond } (i, i+n) \begin{cases} 
\delta\text{-turn, if } n = 1 \\
\gamma\text{-turn, if } n = 2 \\
\beta\text{-turn, if } n = 3 \\
\alpha\text{-turn, if } n = 4 \\
\pi\text{-turn, if } n = 5 
\end{cases}
\]

**β-turns and their types**

The β-turn constitutes a well-studied and predominant subset of tight turns and is a common feature in biologically active peptides and globular proteins. There is an increasing interest in exploring the roles of β-turns in proteins, which usually serve as linkers between the secondary structure elements of the proteins. Studies of the effects of turns on the stability and folding kinetics by modifying the loop length or sequence in the turns of α-helix bundle proteins (Brunet et al., 1993; Predki et al., 1996; Nagi et al., 1999), β-barrel proteins (Ybe and Hecht, 1996; Martinez et al., 1998; Kim and Frieden, 1998), and small proteins (<100 amino acid residues) (Zhou et al., 1996; Gu et al., 1997) have been performed recently. These studies have demonstrated the
important contribution of turns to the folding and stability of the proteins. It has also been demonstrated that β-turns are very important in fibrous proteins, recently it has been postulated that they could play a role in the molecular resilience of an elastin protein tropoelastin (Debelle and Tamburro, 1999). Moreover, β-turn is an important component of β-hairpin structures (Sibanda and Thornton, 1985).

A variety of definitions for the β-turn have been suggested in the past quarter of a century, illustrating the evolution of the concept. Originally, Venkatachalam described β-turns with a hydrogen bond between the CO of the first residue and the NH of the last one (Venkatachalam, 1968). However, Lewis et al. (1973) found that 25% of β-turns are ‘open’, i.e., have no intraturn hydrogen bond as stipulated by Venkatachalam (1968). Therefore, Richardson (1981) has since reappraised the situation and suggested a new definition for β-turn: A β-turn comprises four consecutive residues, i, i+1, i+2 and i+3, where the distance between Cα(i) and Cα(i+3) is less than 7Å and the tetrapeptide chain is not in a helical conformation (Richardson, 1981; Rose et al., 1985) (Fig. 2.2).

Although the distance between the Cα atoms in the first and last residues of a tetrapeptide, i.e., Cα(i) and Cα(i+3) is a key criterion common to all β-turns, the backbone dihedral angles of central residues i+1 and i+2 define different types of β-turns.

Originally, Venkatachalam defined β-turn types I, II, III, and their mirror images types I’, II’, III’. Using the increased number of known protein structures, Lewis et al.
Review of Literature

(1973) broadened the definition of β-turns and increased the number of β-turn types to 10 (I, I', II, II', III, III', IV, V, VI, VII). Later Richardson (1981) reduced the number of β-turn types to 7 (I, I', II, II', VIA, VIb, IV) in which the first 6 categories are based on \((\phi, \Psi)\) angles, and the last one (type IV) is a miscellaneous category. In this classification, the β-turn types III and VII have been eliminated because the conformation of type III is helical and easily recognized to be a part of a \(3_{10}\)-helix-like structure. As for Type VII turn which is a kink in the protein chain created by \(\Psi_{i+1} \approx 180^\circ\) and \(|\phi_{i+2}| < 60^\circ\) or \(|\Psi_{i+1}| < 60^\circ\) and \(\phi_{i+2} \approx 180^\circ\) (Lewis et al., 1973), since it is defined by only two angles and can have two different values for those, they vary greatly in appearance. Therefore, type VII also seems unjustifiable as a distinct category. The rationale given by Richardson is now widely accepted. As claimed by him, type VIII is a new class of β-turn in which the central residues \((i+1, i+2)\) adopt an \(\alpha_8\beta\) conformation. The representative values of \(\phi\) and \(\Psi\) for each of the nine turn types are given in Table 2.1.

Table 2.1: Dihedral angles of central residues \((i+1, i+2)\) for β-turn types.

<table>
<thead>
<tr>
<th>Turn Type</th>
<th>(\phi_{i+1})</th>
<th>(\psi_{i+1})</th>
<th>(\phi_{i+2})</th>
<th>(\psi_{i+2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>-60</td>
<td>-30</td>
<td>-90</td>
<td>0</td>
</tr>
<tr>
<td>I'</td>
<td>60</td>
<td>30</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>-60</td>
<td>120</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>II'</td>
<td>60</td>
<td>-120</td>
<td>-80</td>
<td>0</td>
</tr>
<tr>
<td>IV</td>
<td>-61</td>
<td>10</td>
<td>-53</td>
<td>17</td>
</tr>
<tr>
<td>VIA1</td>
<td>-60</td>
<td>120</td>
<td>-90</td>
<td>0</td>
</tr>
<tr>
<td>VIA2</td>
<td>-120</td>
<td>120</td>
<td>-60</td>
<td>0</td>
</tr>
<tr>
<td>VIB</td>
<td>-135</td>
<td>135</td>
<td>-75</td>
<td>160</td>
</tr>
<tr>
<td>VIII</td>
<td>-60</td>
<td>-30</td>
<td>-120</td>
<td>120</td>
</tr>
</tbody>
</table>

The turn types I, I', II, II' and VIII perfectly meet the structural requirement of the β-turn definition. For the turn types VIA1, VIA2 and VIB, only when residue \(i+2\) is proline and \(\phi(i+1)\) is 0° (Dyson et al., 1988) rather than 180° as in most of the β-turn cases, are the distances between \(C^\alpha(i)\) and \(C^\alpha(i+1)\) smaller than 7Å as shown in Fig.
2.3. Furthermore, the distance between $C\alpha(i)$ and $C\alpha(i+1)$ for the type IV $\beta$-turn is 7.15Å, greater than 7Å, not rigorously satisfying the definition of a $\beta$-turn. In addition, the dihedral angles given in Table 2.1 for type IV are an average result derived from the tetrapeptides assigned by the authors (Hutchinson and Thornton, 1994) as a miscellaneous category.

Figure 2.3: Stereo-diagrams of $\beta$-turn types.
Review of Literature

The region occupied by β-turn in Ramachandran plot can be represented by a vector corresponding to the $\phi$, $\Psi$ angles of residues $i+1$ and $i+2$ (Fig. 2.4).

![Ramachandran plot showing β-turn types I and II represented by a vector. The vector starts at the $\phi$, $\Psi$ coordinates of residue $i+1$ and points to the $\phi$, $\Psi$ coordinates of residue $i+2$.](image)

Figure 2.4: Ramachandran plot showing β-turn types I and II represented by a vector. The vector starts at the $\phi$, $\Psi$ coordinates of residue $i+1$ and points to the $\phi$, $\Psi$ coordinates of residue $i+2$.

**γ-turns and their types**

The second most characterized turn class, after the β-turns, is the γ-turn. A γ-turn resembles a β-turn, but it has only three residues instead of four. A γ-turn is defined as three residues turn with a hydrogen bond between the carbonyl oxygen of residue $i$ and the hydrogen of the amide group of residue $i+2$ (Smith and Pease, 1980; Toniolo, 1980; Rose et al., 1985).

γ-turns are divided into two classes called inverse and classic (Bystrov et al., 1969; Rose et al., 1985) based on dihedral angle values corresponding to the $i+1$ residue (Table 2.2).
Table 2.2: Dihedral angles of central residue \((i+1)\) for \(\gamma\)-turn types.

<table>
<thead>
<tr>
<th>(\gamma)-turn</th>
<th>(\phi(i+1))</th>
<th>(\Psi(i+1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic</td>
<td>75.0 ± 40</td>
<td>-64.0 ± 40</td>
</tr>
<tr>
<td>Inverse</td>
<td>-79.0 ± 40</td>
<td>69.0 ± 40</td>
</tr>
</tbody>
</table>

They differ in that the main-chain atoms of the two forms are related by mirror symmetry (just as type I and I’ or type II and II’ \(\beta\)-turns). Of the two, classic ones are far less common, and those that do exist are frequently found at the loop ends of \(\beta\)-hairpins (Milner-White et al., 1988). On the other hand inverse \(\gamma\)-turns tend not to give rise to polypeptide chain reversal (Milner-White et al., 1988). Fig. 2.5 shows the two types of \(\gamma\)-turns.

![Figure 2.5: \(\gamma\)-turn types.](image)

Earlier studies have shown that the \(\gamma\)-turn structure is fairly common in proteins (Milner-White et al., 1988; Milner-White, 1990). The first example of a \(\gamma\)-turn in a protein was described by Matthews (1972) at the end of a \(\beta\)-hairpin in thermolysin. In 54 analyzed proteins, 12 classic \(\gamma\)-turns and approximately 120 inverse \(\gamma\)-turns were located (Milner-White et al., 1988). From these studies it was also reported that most of the classic \(\gamma\)-turns occur at the ends of \(\beta\)-hairpins, but very few of the inverse. Milner-White et al. infers the difference between the occurrences of the two \(\gamma\)-turn types in hairpins by the fact that the main chain in inverse turns is more divergent than
the main chain in classic turns. In addition, it was reported that the side chain of the middle residue of classic \( \gamma \)-turn turns fold back towards the seven-membered ring that appears through the hydrogen bond, but this is not the case for inverse \( \gamma \)-turns. This folding of the side chain results in an extra hindrance for all residues, except for glycine that accounts for a more frequent occurrence in classic \( \gamma \)-turns, compared to inverse \( \gamma \)-turns. On the other hand, the inverse \( \gamma \)-turns fall into two overlapping categories (Milner-White, 1990), those with strong hydrogen bonds (<-1.0 Kcal/mol) and those with weak bonds (>1.0 Kcal/mol). It has been shown that the stronger bonds often are located in certain characteristic positions within proteins. Some are situated at the N termini or at the C termini of \( \alpha \)-helices, for example at the \( \text{Ca}^{2+} \) binding EF motif, but they can also be found at both ends of \( \beta \)-strands or \( \beta \)-sheets. A number of the stronger inverse \( \gamma \)-turns occur at ligand binding sites or active sites, for example the loop where the catalytically important aspartate is located in serine proteases (Milner-White, 1990). Weakly hydrogen bonded inverse \( \gamma \)-turns have long hydrogen bonds, on average 3.5\( \text{Å} \) (Milner-White, 1990). These turns are very common and they often exist as clusters of bonds, joining successive amino acids along the polypeptide chain.

It has however been postulated that inverse \( \gamma \)-turns may function as intermediates in folding and thereby help stabilizing \( \beta \)-strands before they become \( \beta \)-sheets (Milner-White, 1990). Recently, \( \gamma \)-turns has also brought attention through studies that describe the incorporation of peptide secondary structure mimetics into small bioactive peptides in the development of stable, effective and selective receptor ligands (Alkorta et al., 1996).

\( \alpha \)-turns and their types

Compared with \( \beta \)- and \( \gamma \)-turns, \( \alpha \)-turns has been little investigated especially for their lower occurrence in proteins and peptides. The \( \alpha \)-turn corresponds to a chain reversal involving five amino acids and may be stabilized by a hydrogen bond between the CO group of the first residue and the NH group of the fifth residue (Toniolo, 1980).
In 1996, Pavone et al. had undertaken a systematic search of isolated α-turns in a data set of 193 proteins and compared structures of different types of α-turns using a clustering procedure. In fact, on average about two isolated α-turns were identified in each protein analyzed (Pavone et al., 1996). This study has revealed that these structures are mainly characterized by hydrophilic amino acids. It has also been shown that these structures are not only exposed to solvent, but also protrudes outward from the protein surface with a hook-like shape and therefore these structural motifs can function in interaction mechanism (Pavone et al., 1996). Moreover, these turns provide connection between extended peptide chains.

In has been shown in past that α-turns have functional role in molecular recognition and protein folding (Artymiuk and Blake, 1981; Stout, 1989; Baker, 1988; Wang et al., 1990). For instance, it was found that the residues in the α-turn in protein human lysozyme participate in a cluster of hydrogen bonds and they are located in the active site cleft suggesting the possibility of a functional role (Artymiuk and Blake, 1981). Cys residue in α-turn (residues 35-39) in protein Ferredoxin I and His and Met residues in α-turn (residues 117-121) in protein azurin are involved in the metal ion coordination (Stout, 1989; Baker, 1988). In T-cell surface glycoprotein, residues 51-55 forming α-turn are located in the putative binding region of HIV gp120 protein (Wang et al., 1990). Thus, these rarely occurring motifs whenever present on the protein surface might contain specific information about molecular recognition processes. Moreover, α-turns are also relevant structural domains in small peptides, particularly in cyclopeptides containing 7-9 residues in their sequence. For example, α-turns are important structural domains in cyclo-peptides such as Ilamicinb1 (Iitaka et al., 1974) and cyclolinopeptide A (Di Blasio et al., 1989, 1992) that have important biological functions. In several cases, these turns are also found to be located at the tip of type 3 and 4 β-hairpins (Pavone, 1988). Amatoxins (Zanotti et al., 1989), a family of toxic peptides isolated from Amanita phalloides, are examples of cyclo-octapeptides containing both types I-αRS and II-αRS turns.

α-turns can be classified into nine types depending on the φ, Ψ values of the residues $i+1$, $i+2$ and $i+3$ (Nemethy and Printz, 1972). The α-turns containing all trans peptide bonds are grouped in eight possible different clusters on the basis of the $φ(i+1)$,
Review of Literature

Ψ(i+1), φ(i+2), Ψ(i+2) and φ(i+3) sign, corresponding to eight possible permutations (Table 2.3) (Pavone et al., 1996).

Table 2.3: Naming scheme of α-turns according to the sign of the φ(i+1), φ(i+2) and φ(i+3) angles.

<table>
<thead>
<tr>
<th>Names</th>
<th>φ(i+1)</th>
<th>φ(i+2)</th>
<th>φ(i+3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-αRS</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>I-αLS</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>II-αRS</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>II-αLS</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>I-αRU</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>I-αLU</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>II-αRU</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>II-αLU</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

These clusters formed by three pairs of consecutive φ, Ψ angles were named, according to the φ(i+1), φ(i+2) and φ(i+3) sign, as “X-αyz turns”, where X = I, II depends on residue i+2 and i+3 φ angles [“I” if sign(φ(i+2)) = sign(φ(i+3)); “II” if sign(φ(i+2)) ≠ sign(φ(i+3))]; Y = R, L depends on the sign of residue i+3 φ angle [“R” if φ(i+3) < 0; “L” if φ(i+3) > 0]; Z = S, U depends on the residue i+1 and i+3 φ angles [“S” if sign(φ(i+1)) = sign(φ(i+3)); “U” if sign(φ(i+1)) ≠ sign(φ(i+3))]. In fact, “R” or “L” represents a clockwise (right-handed) or a counterclockwise (left-handed) chain reversal, when viewing the structure down the C’O bond, “U” or “S” represents a U-shaped or screw-like-shaped turn (Calascibetta, 1976), and “I” or “II” discriminate structures with similar topology but different conformations. The representative values of φ(i+1), Ψ(i+1), φ(i+2), Ψ(i+2) and φ(i+3), Ψ(i+3) for each of the nine α-turn types as classified by Pavone et al. (1996) are given in Table 2.4.
2.1.2 Prediction of tight turns

Numerous methods have been developed in the past for the secondary structure prediction (http://PredictionCenter.llnl.gov/casp5/). Except for a few of the earlier methods, all the secondary structure prediction methods predict only three states in a protein – helices, β-strand and coil. The coil region in a protein includes tight turns, bulges and random coil structures (Chou, 2000). The present secondary structure prediction methods do not provide any information about tight turns, despite the fact that tight turns are the most common type of non-repetitive structure comprising 50% of the residues. In contrast to a vast number of methods for regular secondary structure prediction, only few methods have been reported for prediction of tight turns. The development of an accurate method for identifying the location and type of turns within a protein sequence would be an important step in ultimate goal of predicting the detailed three-dimensional structure of a protein from its amino-acid sequence alone and will help in better understanding of protein folding and stability.
Review of Literature

2.1.3 β-turn prediction methods

A number of methods have been developed for predicting β-turns since the recognition of their existence in proteins. The various methods in use depend on different approaches including positional preference (Lewis et al., 1973; Chou and Fasman, 1974; Wilmot and Thornton, 1988, Hutchinson and Thornton, 1994), correlation of coupling effects (Chou and Blinn, 1997; Zhang and Chou, 1997), pattern matching (Cohen et al., 1986) and neural network (McGregor et al., 1989; Shepherd et al., 1999). Each is briefly described here.

Chou-Fasman algorithm: The Chou-Fasman algorithm for the prediction of protein secondary structure is one of the most widely used predictive schemes. The Chou-Fasman method of secondary structure prediction depends on assigning a set of prediction values to a residue and then applying a simple algorithm to the conformational parameters and positional frequencies (Chou and Fasman, 1974). The conformational parameters for each amino acid are calculated by considering the relative frequency of a given amino acid within a protein, its occurrence in a given type of secondary structure, and the fraction of residues occurring in that type of structure.

Chou and Fasman (1977) calculated the positional propensities of amino acids in β-turns using a small dataset of 29 protein structures. It consisted of following steps:

a) The absolute amino acid occurrences for each of the four positions in the β-turn were calculated and were normalized to calculate the positional frequencies f(i), f(i+1), f(i+2), f(i+3) as

\[
 f(i) = \frac{\text{Number of particular amino acid in position } i \text{ of turn}}{\text{Number of particular amino acid in proteins studied}}
\]  

(2.1)

b) The conformational parameter Pt for turns was calculated as

\[
 T_a = \text{number of amino acid } X \text{ in turns/number of amino acid } X \text{ in proteins studied}
\]

\[
 T_t = \frac{\text{total number of amino acids in turns}}{\text{total number of amino acids in proteins}}
\]

Number of particular amino acid in position i of turn
Number of particular amino acid in proteins studied

22
Similarly, the conformational parameters for helices and $\beta$-sheets $P_a$ and $P_b$ were calculated.

c) To identify a turn at residue $j$ at position $i$, the probability, ‘$Pr$’ was calculated as

$$ Pr = f(i)f(i+1)f(i+2)f(i+3) $$

(2.3)

d) A $\beta$-turn was predicted if ‘$Pr$’ was greater than 0.000075 and the average value of $P_i$ was greater than unity in the tetrapeptide, and the averages for the tetrapeptide obey the inequality, $P(\text{helix}) < P(\text{turn}) > P(\text{sheet})$

It was found that the most frequently occurring $\beta$-turn residues were Asn, Cys, Asp in the first position, Pro, Ser, Lys in the second position, Asn, Asp, Gly in the third position, and Trp, Gly, Tyr in the fourth position. Residues with the highest $\beta$-turn potential in all four positions were Pro, Gly, Asn, Cys, Asp, and Ser while the most hydrophobic residues (i.e. Val, Ile, and Leu) showed the lowest $\beta$-turn potential (Chou and Fasman, 1977).

**Thornton’s algorithm:** In 1988, a prediction program for $\beta$-turns was developed on a data set of 59 non-identical proteins (resolution 2Å) by Wilmot and Thornton based on the statistical method employed by Chou & Fasman (1977). Initially, the absolute amino acid occurrences for each of the four positions in the $\beta$-turn and its types were calculated. These were then normalized to give positional frequencies $f(i)$, $f(i+1)$, $f(i+2)$ and $f(i+3)$, which were used by the predictive algorithm. Conformational parameters $P_i$ were also calculated for $\beta$-turn types I and II. The conformational parameters for helix, $P_a$ and $\beta$-sheet, $P_b$ were taken from Chou & Fasman (1977).

The positional potential for residue $j$ (in which $j$ is one of the twenty amino acids) at position $i$ in turns, $P_{\beta}(j)$ was calculated as

$$ P_{\beta}(j) = \frac{f_{\beta}(j)}{<f_{\beta}>} $$

(2.4)
Review of Literature

\[ f(j) = \frac{\text{Number of residue } j \text{ at position } i \text{ of turns}}{\text{Number of residue } j \text{ in proteins}} \] (2.5)

\[ <f_j> = \frac{\text{total number of residues at position } i \text{ of turns}}{\text{total number of residues in proteins}} \] (2.6)

For each turn type \((k)\) type-dependent positional preference was calculated as

\[ P_{nk}(j) = \frac{f_a(j)}{<f_{nk}>} \] (2.7)

\[ f_a(j) = \frac{\text{Number of residue } j \text{ at position } i \text{ of turn type } k}{\text{Number of residue } j \text{ in proteins}} \] (2.8)

\[ <f_a> = \frac{\text{total number of residues at position } i \text{ of turn type } k}{\text{total number of residues in proteins}} \] (2.9)

The algorithm used the following criteria to test for a particular \(\beta\)-turn type:

a) \(P_t = f(i)f(i+1)f(i+2)f(i+3)\) was greater than a calculated cut-off value.

- cut-off value used for Type I \(\beta\)-turn = 4.00
- cut-off value used for Type II \(\beta\)-turn = 2.70

b) \(<P_t>\) was greater than unity.

c) \(<P_t>\) was greater than either \(<P_a>\) or \(<P_b>\).

d) The cut-off values were obtained by the “fine tuning” of the average probability of occurrence of that particular \(\beta\)-turn type. The average probability was the average of the \(P_t\) values for every possible four residues sequence. The fine-tuning was in the form of a multiplication factor, which when applied to the average probability, maximized the number of database-located turns of a particular type.

It was found that \(\beta\)-turn types I and II have different sequence preferences. Type I turns favor Asp, Asn, Ser and Cys at position \(i\); Asp, Ser, Thr and Pro at position \(i+1\); Asp, Ser, Asn and Arg at position \(i+2\); Gly, Trp and Met at position \(i+3\), whilst type II turns prefer Pro at position \(i+1\); Gly and Asn at position \(i+2\); Gly, Trp and Met at position \(i+3\) (Wilmot and Thornton, 1988).
1-4 & 2-3 Correlation Model: The correlation between different amino acids plays an important role in the β-turn prediction. Based on this, a model called 1-4 & 2-3 correlation model was proposed by Zhang and Chou (Zhang and Chou, 1997). In this model, the coupling effect between the 1st and 4th residue and that between the 2nd and 3rd was given a special consideration. When a tetrapeptide folds into a β-turn, the interaction between its 1st and 4th residue and between its 2nd and 3rd residue play an important role. Particularly, a hydrogen bond may be formed between the backbone C=O of the 1st residue and the backbone NH of the 4th residue. Based on this concept, a model of first-order Markov chain involving conditional probabilities \( P_{i}(X_i | X_{i-1}) \) and \( P_{i}(X_i | X_{i+1}) \) was developed. On the basis of these probabilities, an attribute function \( \phi \) was calculated as

\[
\phi(R_{i-1}, R_{i-2}, R_{i}, R_{i+1}) = g P_i(R_i)P_{i-2}(R_{i-2} | R_{i-1})P_{i+1}(R_{i+1} | R_{i})
\]

where \( g = 10^4 \) was the amplifying factor used for making the data in a range easier to handle. \( P_i(R_i) \) was the probability of amino acid \( R_i \) occurring at \( i^{th} \) position in the β-turn set and \( P_{i-1}(R_{i-1}) \) was the probability of amino acid \( R_{i-1} \) occurring at \( (i-1)^{th} \) position in the β-turn set. \( P_{i-2}(R_{i-2} | R_{i-1}) \) was the probability of amino acid \( R_{i-2} \) occurring at \( (i-2)^{th} \) position given that \( R_{i-1} \) had occurred at \( (i-1)^{th} \) position. \( P_{i+1}(R_{i+1} | R_i) \) was the probability of amino acid \( R_{i+1} \) occurring at \( (i+1)^{th} \) position given that \( R_i \) had occurred at \( i^{th} \) position. Thus, the 1-4 and 2-3 coupling was incorporated via the conditional probabilities \( P_{i-2}(R_{i-2} | R_{i-1}) \) and \( P_{i+1}(R_{i+1} | R_i) \). The larger the \( \phi \) of tetrapeptide, the closer its attribute to the β-turn formation. A β-turn is predicted if the discriminant function, \( \Lambda \) is positive where, \( \Lambda = \phi - \lambda \) and \( \lambda \) is the threshold value which was determined by an optimization procedure.

Sequence Coupled Model: Chou (1997b) proposed a residue-coupled model based on first order Markov chain to predict β-turns in proteins. Given a tetrapeptide, its attribute to the β-turn set, \( S^+ \) or the non β-turn set, \( S^- \) was expressed, respectively by an attribute function \( \Psi \) defined as:

\[
\Psi^+(R_{i-1}, R_{i-2}, R_{i}, R_{i+1}) = g P_i(R_i)P_{i-1}(R_{i-1} | R_i)P_{i-2}(R_{i-2} | R_{i-1})P_{i+1}(R_{i+1} | R_i)
\]

(2.11)
where \( g = 10^4 \) was the amplifying factor used for making the data in a range easier to handle. \( P^+(R_i) \) was the probability of amino acid \( R_i \) occurring at sub site \( i \) in the \( \beta \)-turn tetrapeptide set \( S' \), and was independent of the other subsites because \( R_i \) was located at the first position of the four sub site sequence. \( P^+_{i+1}(R_{i+1}|R_i) \) was the probability of amino acid \( R_{i+1} \) occurring at the sub site \( i+1 \) given that \( R_i \) had occurred at position \( i \) and so forth.

For non \( \beta \)-turn set, the attribute function \( \Psi' \) was:

\[
\Psi'(R_iR_{i+1}R_{i+2}R_{i+3}) = gP^+(R_i)P^+_{i+1}(R_{i+1}|R_i)P^+_{i+2}(R_{i+2}|R_{i+1})P^+_{i+3}(R_{i+3}|R_{i+2})
\]  

(2.12)

The discrimination function \( A \) was

\[
A(R_iR_{i+1}R_{i+2}R_{i+3}) = w^+\Psi'(R_iR_{i+1}R_{i+2}R_{i+3}) - w^-\Psi'(R_iR_{i+1}R_{i+2}R_{i+3})
\]  

(2.13)

where \( w^+ \) and \( w^- \) were the weight factors for the probabilities derived from the \( \beta \)-turn and non-\( \beta \)-turn training datasets, respectively. Thus, a \( \beta \)-turn was predicted if \( A > 0 \) (Chou 1997b).

Later, the conditional probabilities have been calculated for different types of \( \beta \)-turns to enable the residue-coupled model to predict different \( \beta \)-turn types as well (Chou and Blinn, 1997). They have used \( S^1, S^1, S^2, S^2, S^6, S^8, \) and \( S' \) to represent the tetrapeptide sets of type I \( \beta \)-turn, type I' \( \beta \)-turn, type II \( \beta \)-turn, type II' \( \beta \)-turn, type VI\( \beta \)-turn, type VIII \( \beta \)-turn and non-\( \beta \)-turn. Given a tetrapeptide, its attribution to the sets \( S^1, S^1, S^2, S^2, S^6, S^8, \) and \( S' \) was expressed by \( \phi^1, \phi'^1, \phi^2, \phi'^2, \phi^6, \phi^8, \) and \( \phi' \), respectively. These attribute functions were calculated for each turn type in the same way as they were calculated for turns and non-turns sets (Eqs. 2.11 and 2.12).

For a given tetrapeptide \( R_iR_{i+1}R_{i+2}R_{i+3} \), if its attribute function to the \( \beta \)-turn set \( S^1 \) is greater than that to the \( \beta \)-turn set \( S^2 \), i.e., \( \phi^1 > \phi^2 \), then the tetrapeptide would have a greater propensity for the \( \beta \)-turn type I than for the \( \beta \)-turn type II. Based on such a rationale, the tetrapeptide \( R_iR_{i+1}R_{i+2}R_{i+3} \) can be logically predicted to be the structural type for which \( \phi \) has the maximum value, as can be formulated as follows:

\[
\phi^* = \max \{\phi^1, \phi'^1, \phi^2, \phi'^2, \phi^6, \phi^8, \phi'\}
\]  

(2.14)
where the operator \texttt{max} means taking the maximum of the quantities in the braces; then the superscript will give the corresponding $S' (\times = 1, 1', 2, 2', 6, 8, \text{or} -)$ to which the predicted tetrapeptide $R_iR_{i+1}R_{i+2}R_{i+3}$ should belong.

**GORBTURN(v3.0):** The program GORBTURN (v3.0), a new version of BTURNPRED (Wilmot and Thornton, 1990) is a user-friendly piece of software written in Fortran77. The original program, known as BTURNPRED consist of two different programs, BTURN (v2.0) and GORBTURN (v1.0). In 1994, a revised set of potentials for $\beta$-turn formation in proteins was published by Gail Hutchinson and Janet Thornton (Hutchinson and Thornton, 1994) based on the exponential growth in the number of atomic resolution protein structures solved by X-ray crystallography and NMR. These potentials have been incorporated into a new version of GORBTURN (v3.0), which allows the differential prediction of five $\beta$-turn types (I, II, I', II', VIII). The program uses the positional frequencies and the directional parameters in combination with equivalent parameters produced from work by Gibrat (Gibrat \textit{et al.} 1987) to eliminate potential helix and strand forming residues from the $\beta$-turn prediction. It uses a 17 residues window in the algorithm, which makes it less suitable for the prediction of $\beta$-turns in short peptide sequences than in proteins. The new version of GORBTURN (v3.0) has been tested on 16 protein chains, which were non-homologous to those used in the derivation of the $\beta$-turn potentials, and whose structures were determined to 2Å resolution or better. The positional accuracy of $\beta$-turn prediction was only marginally improved by 1% over GORBTURN (v1.0) (Wilmot and Thornton, 1988), but the program carries out a 9 state prediction compared to the 6 state prediction of GORBTURN (v1.0).

**BTPRED:** It is a neural network based method developed on a set of 300 non-homologous protein domains with resolution 2.0Å or better (Shepherd \textit{et al.}, 1999). A neural network was used to predict whether a given residue is part of a $\beta$-turn or not. A filtering network was used to improve the accuracy and the individual turn type is predicted using a separate neural network for each turn type to be predicted. Conflicting predictions arising were resolved on a ‘winner-takes-all’ basis; the turn
type is set to be the same as that of the network with the largest output. The method uses secondary structure information obtained from PHDsec program (Rost and Sander, 1993a; Rost and Sander, 1994) about each amino acid rather than just amino acid type. A web server based on this method is available at http://www.biochem.ucl.ac.uk/bsm/btred/. Fig. 2.6 shows a sample output of BTPRED server.

Residue hydrophobicities: The β-turn prediction approach based on residue hydrophobicities was based on the hypothesis that turns occur at those sites in the peptide chain where the hydrophobicity is at a local minimum. A method for locating β-turns using this hypothesis was presented by Rose (1978).

Local clusters of large hydrophobic residues were identified by scanning the amino acid sequence. The clusters can consist of one to a few nearly consecutive residues in the sequence. Rose used the Nozaki and Tanford free energy of transfer (Nozaki and Tanford, 1971) from water to an organic solvent to construct a graph of residue

Figure 2.6: A sample of the prediction output of BTPRED. AA is the amino acid sequence as submitted by the user. SS is the secondary structure. Predicted turn residues are denoted by t. The fourth row is the reliability index (0-9) for each turn/non-turn prediction. Type I and II β-turns are denoted by 1 and 2 respectively.
number against hydrophobicity. After smoothing the graph, the profile was differentiated to find local minima, which correspond to peptide chain turns (Fig. 2.7).

![Graph showing hydrophobicity profile](image)

Figure 2.7: A hydrophobicity profile plot showing the free energy of transfer against the residue number. Local minima in these curves correspond to the peptide chain turns.

### 2.1.4 γ-turn prediction methods

Presently, there does not exist any method for predicting γ-turns in a given amino acid sequence.

### 2.1.5 α-turn prediction methods

**Sequence Coupled Model:** Chou (1997a) proposed a residue-coupled model based on first order Markov chain to predict α-turns in proteins. Given a pentapeptide, its attribute to the α-turn set $S^+$ or the non-α-turn set $S^-$ is expressed, respectively by an attribute function $\psi$ ($\psi^+$ for α-turn & $\psi^-$ for non-α-turn), which was defined as

\[
\psi^+(R,R_1,R_2,R_3,R_4) = gP_+(R_4)P_+(R_1|R_4)P_+(R_2|R_1)P_+(R_3|R_2)P_+(R_4|R_3)
\]

(2.15)

\[
\psi^-(R,R_1,R_2,R_3,R_4) = gP^-(R_4)P^-(R_1|R_4)P^-(R_2|R_1)P^-(R_3|R_2)P^-(R_4|R_3)
\]

(2.16)
where, $g=10^4$ was the amplifying factor used for making the data in a range easier to handle. $P_i(R_i)$ was the probability of amino acid $R_i$ occurring at sub site $i$ in the $\alpha$-turn pentapeptide set $S^+$, and it was independent of the other subsites because $R_i$ is located at the first position of the five subsite sequence. $P_{i+1}(R_{i+1}|R_i)$ was the probability of amino acid $R_{i+1}$ occurring at the subsite $i+1$ given that $R_i$ had occurred at position $i$ and so forth. The discriminant function was calculated from following equation

$$
\Delta(R,R_{i+1}R_{i+2}R_{i+3}R_{i+4}) = w'\psi'(R,R_{i+1}R_{i+2}R_{i+3}R_{i+4}) - w\psi'(R,R_{i+1}R_{i+2}R_{i+3}R_{i+4})
$$

(2.17)

where $w'$ and $w$ are the weight factors for the probabilities derived from the $\alpha$-turn and non-$\alpha$-turn training datasets respectively. Thus, a $\alpha$-turn was predicted if $\Delta > 0$.

### 2.2 Non-conventional interactions in proteins

Due to restrictive rotations about certain bonds and various inter-atomic interactions, the protein chain folds in a definite manner. One of the main interactions in proteins is hydrogen bonding. Fundamentally, the hydrogen bonding can be classified as “conventional hydrogen bonding” and “non-conventional hydrogen bonding”. Conventional hydrogen bonds are a major feature of the structure and function of macromolecules (Baker and Hubbard, 1984; Jeffrey and Saenger, 1991). They are the key to many phenomenons, including the formation and stabilization of secondary structure, protein folding and stability, molecular recognition and enzymatic reactions, which utilize general acid and general base catalysis. The rapidly accumulating data from crystallographic and theoretical studies of organic compounds have provided detailed insights into the nature of hydrogen bond. Classically, O-H and N-H groups (with the proton attached to an electronegative atom) were considered to be hydrogen bond donor. But in recent years, the ‘non-conventional’ hydrogen bonds are emerging to be of great importance, in particular the N-H⋯\pi, C-H⋯\pi and C-H⋯O interactions in stabilizing protein structures (Weiss et al., 2001).
2.2.1 Aromatic-backbone NH interactions

Benzene rings and other aromatic systems can interact with hydrogen bond donors such as NH, CH giving rise to hydrogen bonding. Hydrogen bonds between donors X-H and the π-electron cloud of an aromatic moiety were discovered by Wulf et al. (1936) and are today well documented in protein structures. In this hydrogen bond type, the donor group XH is placed roughly above the center of an aromatic ring, and the X-H vector points at it. Probably the first example of an X-H-π hydrogen bond in a peptide crystal structure was reported by McPhail and Sim (1965), but it had little impact in structural science at that time. Much later, N-H-π hydrogen bonds in proteins attracted greater attention following the observation of such interactions in bovine pancreatic trypsin inhibitor (BPTI) by Huber and co-workers (Wlodawer et al., 1984) and in hemoglobin protein-ligand interactions by Perutz et al. (1986). The prevalence of N-H-π hydrogen bonding interactions and its undoubted structural functional importance has recently been documented in a variety of biological systems.

Attractive interactions between aromatic groups of amino acids and nearby amides, in polypeptides and proteins (Ar-NH interaction) are weakly polar and have a quadrupole-dipole nature (Levitt and Perutz, 1988). The interaction can be effectively modeled by the electrostatic interaction between the partial negative charge of an aromatic ring and the partial positive charge of amide hydrogen (Levitt and Perutz, 1988). The strength of this interaction in vacuum (1-4 Kcal/mol) is comparable with that of a conventional hydrogen bond (2-7 Kcal/mol) (Levitt and Perutz, 1988; Cheney et al., 1988; Rodham et al., 1993).

Figure 2.8: Geometry of an Ar-NH interaction.
The geometry of the Ar-NH interaction can be described using the angle $\alpha$ between the vector of the NH bond and the plane of the aromatic ring (Fig. 2.8). When $\alpha$ is larger than $30^\circ$, the Ar-NH interaction is regarded as perpendicular, and if smaller than $30^\circ$, parallel.

In solution, $^1$H-NMR spectroscopy has been used to identify Ar-NH interactions in peptides and proteins (Kemmink et al., 1993; Kemmink and Creighton, 1993; Kemmink and Creighton, 1995; Nardi et al., 1997; Toth et al., 1998; Worth and Wade, 1995). In Ar-NH interactions, the aromatic ring is in close proximity of the amide proton; thus, the delocalized electrons of the aromatic ring change the local magnetic field of the amide proton, thereby causing an anomalous shift called ring shift ($\delta_{\text{ring}}$). The value of the ring shift can be related to the interaction geometry of the Ar-NH interaction (Worth and Wade, 1995).

There are many reports where the Ar-NH interaction has been shown to constrain side-chain conformation and stabilize local structures (Worth and Wade, 1995; Kemmink and Creighton, 1995; Shimohigashi et al., 1999). For instance, in the deltorfin 2 analogue, Hat-D-Ala-Phe-Glu-Ile-Ile-Gly-NH$_2$, in which an Ar($i$)-NH($i+2$) interaction is between the aromatic ring of Phe3 and the backbone amide of Ile5. Another example is in crystalline Tyr-Tyr-Leu monohydrate where the amide group of Tyr2 interacts with the aromatic ring of Tyr1 (Steiner, 1999). NMR measurements of reduced bovine pancreatic trypsin inhibitor (BPTI) revealed local nonrandom conformations (van Mierlo et al., 1993). These local structures are formed by Ar($i$)-NH($i+2$) interactions with the aromatic ring of Tyr10 interacts with the backbone amide of Gly12 and the aromatic ring of Tyr35 interacts with the backbone of Gly37 (Kemmink et al., 1993; Nardi et al., 1997). Further studies by Kemmink and Creighton revealed that the interaction between Tyr35 and Gly37 assists in the folding of BPTI by stabilizing an intermediate structure along the folding pathway (Kemmink and Creighton, 1993; 1995).

Several statistical surveys of sequence distribution and geometry of the Ar-NH interactions in proteins have been reported. Studies by Burley and Petsko suggested involvement of Ar-NH(side-chain) interactions in the stabilization of protein tertiary structures (Burley and Petsko, 1986; 1988). Flocco and Mowbray examined the occurrence and geometry of the interactions between the guanidinium group of Arg,
the amide group of Asn and Gln, and the aromatic side chain of Phe, Tyr, and Trp in 83 non-homologous protein structures (Flocco and Mowbray, 1994). They found that the parallel arrangement of these groups is the most frequent geometry, because the parallel arrangement is sterically more favorable because the nitrogen is able to engage in hydrogen bonding with another residue. Mitchell and coworkers found that approximately 10% of the identified aromatic-amide, -amino, -guanido pairs of 55 non-homologous protein structures are in either side chain-side chain or side chain backbone interactions, mostly in parallel geometry (Mitchell et al., 1994). Worth and Wade searched for Ar(i)-NH(i+1 and i+2) interactions in 297 non-homologous structures and found that in Ar(i)-NH(i+2) interaction the propensity for Gly to be in position i+2 is by far the highest than for all other amino acids (Worth and Wade, 1995). Toth et al. (2001a) searched for Ar-NH interactions between the aromatic ring of the residues Phe, Tyr, and Trp at position i and the backbone amide group of any residue except Pro, at the positions i, i-1, i-2, i-3, i+1, i+2 and i+3 in a dataset of 560 non-homologous protein structures. The study showed the presence of Ar-NH interactions in different secondary structures. It has been found that Ar(i)-NH(i+2, i+3, i-2, i-3) interactions are predominantly in turns. After β-sheets, Ar(i)-NH(i+1) interactions are found more frequently in turns where they have a stabilizing role. The geometry of the Ar-NH interactions was mostly parallel except in Ar(i)-NH(i+2) interactions. In the N-terminal of helices where side-chain backbone amide interaction becomes important for stability, several Ar-NH interactions were found. All these studies have suggested that Ar-NH interactions might have a stabilizing effect on all types of secondary structures.

N-H⋯π interactions are also recognized in peptides containing dehydrophenylalanine residues. It was found that in the most stable conformational state of peptide Ac-(Δ’Phe)₆-NHMe, the amino group of each and every Δ’Phe is involved in N-H⋯π interaction with one of the nearest edge (C₇-C₁₀) of its own phenyl ring (Nandel et al., 2001). The peptides containing (Δ’Phe)₆ with achiral (ΔAla, Gly) and chiral (Ala, Leu) residues at both the N- and C-terminal positions have the lowest energy conformational state corresponding to φ = 0° and Ψ = ±90° where N-H⋯π interactions have been found between the amino group of Δ’Phe and its own aromatic moiety (Nandel and Kaur, 2003).
2.2.2 C-H⋯O and C-H⋯π interactions

In recent years it has been established that a C-H group can be a donor and can form hydrogen bond with hydrogen acceptors such as oxygen atom (C-H⋯O interaction) and π ring system (C-H⋯π interaction) (Fig. 2.9).

A large body of evidence points to the existence of the C-H⋯O and C-H⋯π hydrogen bonds in proteins (Desiraju, 1991; Desiraju and Steiner, 1999). Although considered weak in nature, these non-conventional interactions are gaining wide acceptance as genuine hydrogen bond and suggested to play pivotal role in macromolecular recognitions, drug binding, protein-nucleic acid interactions and stabilization of folding motifs (Derewenda et al., 1994; Derewenda et al., 1995; Brandl et al., 2001; Steiner and Saenger, 1993; Chamberlain and Bowie, 2002). C-H⋯O interactions form 20-25% of the total number of hydrogen bonds constituting the second most important group (Weiss et al., 2001). Brandl et al. had shown that proteins could contain C-H⋯π hydrogen bond between 0.0 and 22.6 per 100 amino acids (Brandl et al., 2001).

Derewenda et al., 1994; 1995 has explored in detail the C-H⋯O hydrogen-bonding capacity of proteins. This work indicated a role for C-H⋯O interactions in protein

![Figure 2.9: Figures showing (a) C-H⋯π and (b) C-H⋯O interactions.](image)
structures. In both parallel and antiparallel β-sheets, mainchain C-H atoms were found to be involved in a bifurcated hydrogen-bonding arrangement to backbone carbonyl groups. In α-helices, the C\(^5\) atom of proline can form a C-H...O bond with the carbonyl oxygen atom to replace the lost N-H...O bond. C-H...O bonds have been identified in collagen, myoglobin, serine hydrolases, dehalogenases, BPTI, ribonuclease A as well as in DNA, protein-DNA interactions and protein-ligand interactions. Derewenda et al., 1994 presented evidence of functional significance in serine hydrolases of a hydrogen bond involving the C-H group of the active site His and a neighboring main-chain carbonyl oxygen. Desiraju and Steiner (1999) later studied in detail the occurrence of C\(^\alpha\)-H...O hydrogen bonds. C-H...O bonds have been identified in membrane protein structures and are hypothesized to stabilize the interhelical interactions. Steiner and Saenger (1993) reported the presence of C-H...O bonds completing the coordination of the buried water molecules in papain. The past studies have also revealed that C\(^\alpha\)-H...O hydrogen bonds occur almost ubiquitously in β-sheets (Fabiola et al., 1997), but they also appear to occur frequently in α-helical proteins (Chakrabarti and Chakrabarti, 1998).

Hydrogen bonds, C-H...π with π-acceptor constitute yet another considerable fraction. Steiner and Koellner described hydrogen bonds in proteins involving aromatic acceptors (Steiner and Koellner, 2001), and Brandl et al. (2001) exhaustively surveyed the occurrence of interactions involving all possible C-H groups (C\(^\alpha\)-H, Caliphatic-H and Caromatic-H) as donors and all possible side chain π-systems as acceptors. The cases in which C-H...π interactions have been described in proteins include the formation of complexes of proteins with ligands or cofactors such as the heme group (Nishio et al. 1998) and design of serine proteases inhibitors (Shimohigashi et al., 1996, 1999). Previous studies have also shown that C-H...π interactions are even responsible for the stabilization of structural elements such as α or 3\(_{10}\)-helices or non-proline cis peptide bonds (Brandl et al., 2001). C-H...π interactions are stable in both polar and non-polar solvents. The overall stabilization energy of about 0.5-1.0Kcal/mol per interaction is enough to make it a potentially important contributor to the overall protein stability (Brandl et al., 2001).
C-H⋯π interactions have also been found in peptides containing (Δ²Phe)₆ with chiral (Ala, Leu) residues at both the N- and C-terminal positions. In these peptides, the Ala/Leu residues at the N-terminus further stabilized the structure through C-H⋯π interactions with the farthest edge of the aromatic ring of i plus 3 Δ²Phe residue (Nandel and Kaur, 2003).

2.3 Prediction of non-conventional interactions in proteins

There is a large number of published literature, which describes the importance of non-conventional hydrogen bonding interactions, their role in stabilization of protein structure and analysis of these interactions in proteins. Best to our knowledge there is no literature that describes the method for predicting these interactions in proteins from their amino acid sequence. The preliminary knowledge of these interactions or at least some limited information about them could be useful in protein tertiary structure prediction.

2.4 Role of β-turns in bioactive peptides

Peptides have the capability to control important functions of the organism, such as cell reproduction, appetite, euphoria, sleep, learning, immune response etc. There is a plethora of bioactive peptides, which act as hormones, neurotransmitters, antioxidants, toxins and antibiotics. Due to the importance of bioactive peptides, extensive studies have been carried out directed at their structure determination with a goal to understand function and to design clinically and diagnostically useful compounds.

Each role assumed by a bioactive peptide typically corresponds to a unique three-dimensional structure. Moreover, to design biologically active peptide requires detailed knowledge of the three-dimensional structure and is generally focused towards the modification of secondary structure elements. Also, the secondary
structure rather than the tertiary structure is the dominant factor affecting the binding characteristics of the peptides (Kaiser, 1987).

In peptides, turns are the conformations of choice for simultaneously optimizing both backbone-chain compactness (intramolecular nonbonded contacts) and side-chain clustering (to facilitate intermolecular recognition) (Rose et al., 1985). Presence of turns in bioactive conformations may in fact reflect the lack of alternative conformational possibilities. Although β-turns in peptides and in proteins is homologous structures, the difference in molecular sizes leads to distinct geometric and energetic characteristics.

A small peptide that elicits a particular biological response must maintain a suitable orientation of binding groups for productive interaction with a receptor. A turn is a means of stabilizing a folded conformation in a small molecule via short-range interactions, which results in proper arrangement of groups essential for receptor binding. A large body of evidence points to the significant occurrence of β-turns in bioactive peptides in addition to regular secondary structure elements (please see Smith and Pease, 1980). It has also been observed that β-turns occur frequently among the conformationally active forms of the various linear and cyclic peptides (Kessler, 1982; Smith and Pease, 1980). For instance, in many antimicrobial peptides the structure that is responsible for bactericidal activity contains β-turn (Kawano et al., 1990). Researchers have concentrated significant efforts at attempting to duplicate β-turns, in order to create novel therapeutics or deduce additional information about biological interactions. Examples include, peptides from the immunoglobulin family (specifically from CD4 protein) for autoimmune diseases; a somatostatin analog (octreotide), which is used for the treatment of severe diarrhea in patients with intestinal tumors (Bajetta et al., 2002). Moreover, the introduction of non-peptide bond mimics of the β-turn motif provides greater potential therapeutic value (Kahn et al., 1988; Pavone et al., 1994) also.

Due to the frequent occurrence of β-turns in small bioactive peptides, the prediction knowledge of turns can be applied to the overall tertiary structure prediction of such peptides.