J Biomol NMR (2014) 59:175–184
DOI 10.1007/s10858-014-9837-z

ARTICLE

Type I and II β-turns prediction using NMR chemical shifts

Ching-Cheng Wang · Wen-Chung Lai · Woei-Jer Chuang

Received: 26 February 2014 / Accepted: 2 May 2014 / Published online: 17 May 2014

Abstract A method for predicting type I and II β-turns using nuclear magnetic resonance (NMR) chemical shifts is proposed. Isolated β-turn chemical-shift data were collected from 1,798 protein chains. One-dimensional statistical analyses on chemical-shift data of three classes β-turn (type I, II, and VIII) showed different distributions at four positions, (i) to ($i + 3$). Considering the central two residues of type I β-turns, the mean values of C_o, C_α, H_N, and N_H chemical shifts were generally $(i + 1) > (i + 2)$. The mean values of C_β and H_α chemical shifts were $(i + 1) < (i + 2)$. The distributions of the central two residues in type II and VIII β-turns were also distinguishable by trends of chemical shift values. Two-dimensional cluster analyses on chemical-shift data show positional distributions more clearly. Based on these propensities of chemical shift classified as a function of position, rules were derived using scoring matrices for four consecutive residues to predict type I and II β-turns. The proposed method achieves an overall prediction accuracy of 83.2 and 84.2 % with the Matthews correlation coefficient values of 0.317 and 0.632 for type I and II β-turns, indicating that its higher accuracy for type II turn prediction. The results show that it is feasible to use NMR chemical shifts to predict the β-turn types in proteins. The proposed method can be incorporated into other chemical-shift based protein secondary structure prediction methods.

Keywords β-Turn · NMR · Chemical shift · Secondary structure · Prediction

Abbreviations

BMRB	BioMagResBank
NMR	Nuclear magnetic resonance
PDB	Protein data bank

Introduction

β-Turns were first recognized in proteins by Venkatachalam (1968). They are the most common non-repetitive structure and are on average about 25 % of all residues in proteins (Kabsch and Sander 1983). Turns play an important role in reversing the direction of the main chain and folding structure of proteins (Richardson 1981; Rose et al. 1985). Unlike regular secondary structures of α-helices and β-strands, β-turns are four consecutive residues denoted by (i) to ($i + 3$) in peptides, and typically involve a hydrogen bond between the main chain C=O(i) and N–H($i + 3$) atoms. However, after Lewis et al. (1973) found that hydrogen bonding was not a sufficiently precise condition, the β-turn definition was broadened to require that the distance between the main chain $C_\alpha(i)$ and $C_\alpha(i + 3)$ is <7 Å. With increasing number of known protein structures, various types of β-turn data are sufficiently populated to allow meaningful analysis (Lewis et al. 1971, 1973; Chou and Fasman 1974; Richardson 1981; Wilmot and Thornton 1988, 1990; Hutchinson and Thornton 1994).

Electronic supplementary material The online version of this article (doi:10.1007/s10858-014-9837-z) contains supplementary material, which is available to authorized users.

C.-C. Wang · W.-C. Lai
Institute of Manufacturing Information and Systems, National Cheng Kung University College of Electrical Engineering and Computer Science, Tainan 701, Taiwan

W.-J. Chuang (✉)
Department of Biochemistry and Molecular Biology, National Cheng Kung University College of Medicine, Tainan 701, Taiwan
e-mail: wjcnmr@mail.ncku.edu.tw

β-Turns have been classified into eight categories of turns, namely I, I′, II, II′, VIa, VIb, VIII and a miscellaneous category IV according to the φ, ψ angles of the central two residues ($i + 1$, $i + 2$). β-Turns play many biological roles in proteins and peptides, and therefore it is necessary to develop an accurate and quick prediction method.

Most β-turn prediction methods are based on amino acid contents of proteins and can be divided into those based on statistical methods and those based on machine-learning methods. The first empirical β-turn prediction method was introduced by Lewis et al. (1971), which evaluated the priori probability of occurrence for each amino acid residue in a β-turn observed in three proteins. A series of methods, which uses patterns information, has been proposed concerning the positional frequencies of each amino acid residue (Chou and Fasman 1974, 1979; Cohen et al. 1986; Wilmot and Thornton 1988; Chou 1997; Chou and Blinn 1997; Zhang and Chou 1997; Fuchs and Alix 2005). Machine-learning methods have been proposed for predicting β-turns and β-turn type (McGregor et al. 1989; Shepherd et al. 1999; Kaur and Raghava 2003, 2004; Kim 2004; Asgary et al. 2007; Kirschner and Frishman 2008; Zheng and Kurgan 2008; Petersen et al. 2010; Tang et al. 2011; Song et al. 2012). Current prediction methods perform well, but they rely on amino acid sequence alone.

Chemical shifts of amino acids in proteins may be the most sensitive and easily obtainable nuclear magnetic resonance (NMR) parameters (Wishart 2011). They are related to φ, ψ backbone dihedral angles and can reflect the primary, secondary, and tertiary structures of the proteins (Osapay and Case 1994; Beger and Bolton 1997; Santiveri et al. 2001). Several approaches use chemical shifts to accurately predict the secondary structures of α-helices and β-strands (Wishart et al. 1992; Wang and Jardetzky 2002; Hung and Samudrala 2003; Eghbalnia et al. 2005; Wang et al. 2007; Zhao et al. 2010), tertiary structures (Shen et al. 2008, 2009; Wishart et al. 2008; Shen and Bax 2013; Cheung et al. 2010; Berjanskii et al. 2006), torsion angles (Wishart et al. 1991), and the redox state of cysteine residues (Sharma and Rajarathnam 2000; Wang et al. 2006). The shifts are easily obtainable parameters and have become a key in the process of traditional NMR structure determination (Wishart 2011; Guerry and Herrmann 2011). However, the prediction of non-regular secondary structures in a protein via chemical shifts is still a challenge. The chemical shift values of amino acids involved in β-turns remarkably overlap between the distributions of helical and extended structures, hence most current NMR-based protein structure prediction methods classify turn structures as random coils. Recently, Shen and Bax (2012) proposed the first β-turn identification method that applies artificial neural network algorithms to chemical shift data. Their method can also identify small structural elements, including N-caps, C-caps and various types of β-turn motifs.

In the present paper, $^{1}H_{\alpha}$, $^{1}H_{N}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C_{o}$, and $^{15}N_{H}$ chemical shift data of various types of β-turn in proteins were collected and classified into four positional categories, (i) to ($i + 3$). One- (1D) and two-dimensional (2D) analyses were performed on these examples. The 1D frequency plots of chemical-shift data reveal trends from (i) to ($i + 3$) curves, and 2D cluster analyses are helpful in distinguishing four positional clusters. Based on the chemical shift propensities of β-turn residues, rules are derived using a scoring matrix for four consecutive residues to predict β-turns. The approach is the first statistical method applied chemical shifts for predicting β-turns so far. In addition, a 2DCSi(t) web server for prediction was established. It enables users to submit NMR chemical shift files of protein in BMRB (NMR-STAR) format and predict not only the redox states of cysteine residues, but also the secondary structures of α-helices, β-strands, and type I and II β-turns.

Materials and methods

Data extraction

Two data sets of NMR chemical shifts and β-turn patterns were generated for the analysis. The first data set of six nuclei, namely $^{1}H_{\alpha}$, $^{1}H_{N}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C_{o}$, and $^{15}N_{H}$, NMR chemical shifts of 20 amino acids was extracted from files from the re-referenced protein chemical shift database (Zhang et al. 2003). The set contains 1,798 re-referenced BMRB (Ulrich et al. 2008) recently deposited entries. The second data set is the DSSP algorithm (Kabsch and Sander 1983) assigned secondary structure data set, which is used to define structures of H (Helix: H or G), E (Extended strand: E or B), and C (Coil) from the related proteins.

β-turns are defined using the standard criterion that the distance between the main chain $C_{\alpha}(i)$ and $C_{\alpha}(i + 3)$ is <7 Å (1 Å = 0.1 nm). β-Turn patterns denoted by four consecutive residues, (i) to ($i + 3$), and the central two residues ($i + 1$) and ($i + 2$) are not in helical or extended conformation. The types of β-turn are determined according to the dihedral angle constraints; the ideal angles of the central two residues are (−60°, −30°) and (−90°, 0°) for type I turns; (−60°, 120°) and (80°, 0°) for type II; (60°, 30°) and (90°, 0°) for type I′; (60°, −120°) and (−80°, 0°) for type II′; (−60°, −30°) and (−120°, 120°) for type VIII. The φ, ψ angles are allowed to vary by ± 30° from the ideal angles. In each protein, all residues not assigned as H or E structures are used to explore the sequence and structure-related pattern of the various types of β-turn. β-Turns that occur in isolation only were collected because

Table 1 Numbers of β-turns of various types observed in data set

Turn type	No. of turns	Ideal dihedral angles			
		$\varphi(i+1)$ (°)	$\psi(i+1)$ (°)	$\varphi(i+2)$ (°)	$\psi(i+2)$ (°)
I	3,564	−60	−30	−90	0
II	1,058	−60	120	80	0
I′	559	60	30	90	0
II′	246	60	−120	−80	0
VIII	757	−60	−30	−120	120

the existence of other hydrogen bonds might cause chemical shift deviation in multiple turns. The numbers of isolated β-turns for five types are listed in Table 1, and the occurrence of 20 amino acids for each position is shown in Supplementary Material I Table S1.

Cluster analysis of chemical shifts

1D and 2D cluster plots of chemical-shift data for residues in β-turns as a function of position were drawn using Microsoft Excel spreadsheet program and the MATLAB program. The two most numerous and well-defined types of β-turn, type I and II, are analyzed in the present work. For example, 1D frequency plots of ASP chemical shifts in type I and II turns are shown in Figs. 1a–f and 2a–f. The ASP residues at positions (i), $(i+1)$, $(i+2)$, and $(i+3)$ are shown in blue, red, green, and purple, respectively. 2D cluster plots depict the joint or disjoint areas of positional clusters more clearly (Fig. 3a, b). Due to the joints between clusters, multi-dimensional outlier checking was performed with a tolerance of 0.1. Four ellipses mark cluster boundaries and each ellipse contain 90 % of the paired chemical shift data points.

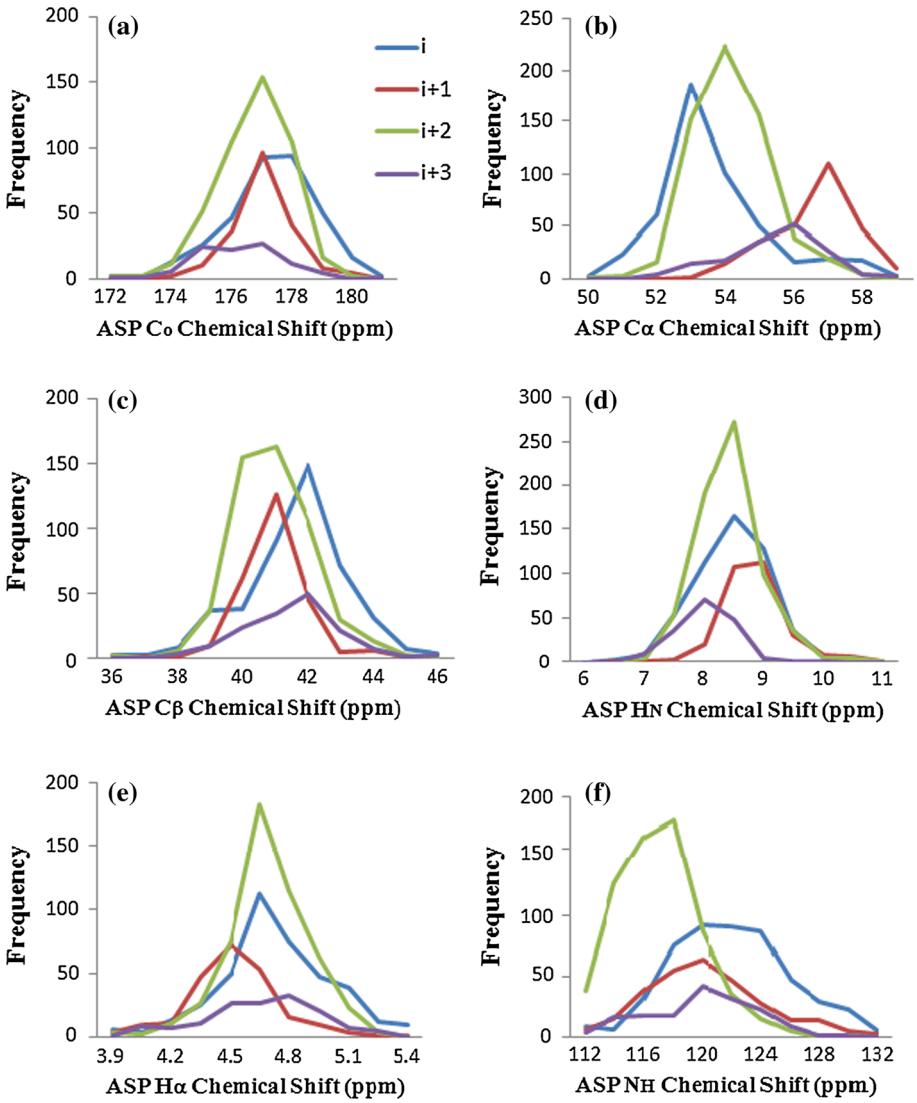


Fig. 1 Distribution of ASP for **a** C_o, **b** C_α, **c** C_β, **d** H_N, **e** H_α, and **f** N_H chemical shifts as a function of position in type I β-turns. Positions (i), $(i+1)$, $(i+2)$, and $(i+3)$ are colored in *blue*, *red*, *green*, and *purple*, respectively

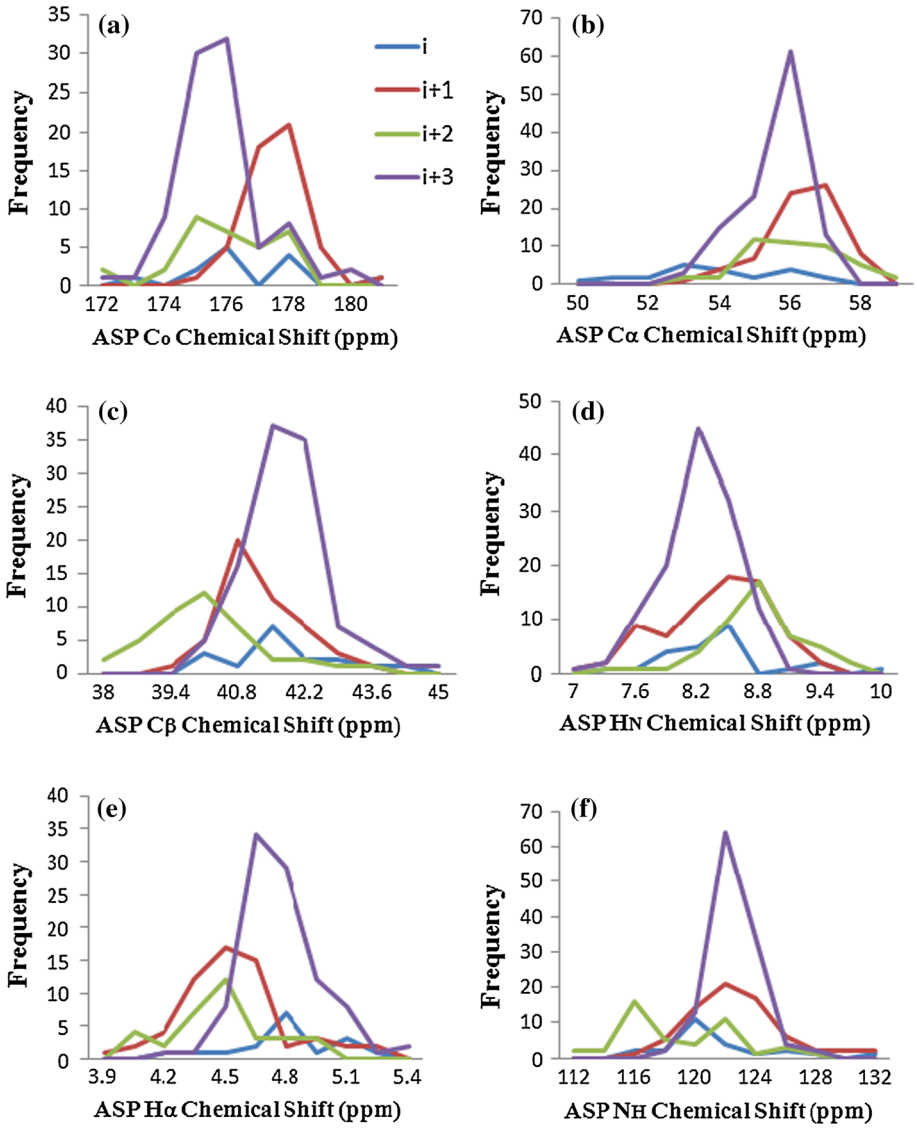


Fig. 2 Distribution of ASP for **a** C_o, **b** C_α, **c** C_β, **d** H_N, **e** H_α, and **f** N_H chemical shifts as a function of position in type II β-turns

Prediction rules and scoring matrix

We first used a scoring matrix derived from the results of 2D cluster analyses of $^{1}H_\alpha$, $^{1}H_N$, $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C_o$, and $^{15}N_H$ NMR chemical shifts to predict the contents of α-helix and β-strands (Wang et al. 2007). We then used the following rules to predict turn structures. A given chemical-shift data point, $\vec{y} = (y_{Co}, y_{C\alpha}, y_{C\beta}, y_H, y_{H\alpha}, y_N)$, can be classified into specific structural type s by calculating the distance $\vec{y}$ from the mean vector of the structure type $\vec{\mu}_s$, which is known as the squared Mahalanobis distance D_s^2:

$$D_s^2 = \left(\vec{y} - \vec{\mu}_s\right)' \vec{S}_s^{-1} \left(\vec{y} - \vec{\mu}_s\right)$$

where S_s represents the covariance matrix of secondary structure type s. The hypothesis of observation $\vec{y}$ whether a sample belongs to structural type s could be checked by an F-test:

$$F = \frac{(N - p - 1)ND_s^2}{p(N-1)^2 - NpD_s^2}$$

The quantity has the F distribution with degrees of freedom p and $N - p - 1$, where N is the number of samples at a p-dimensional vector. A critical squared Mahalanobis distance is found with the formula. The null hypothesis that the observation $\vec{y}$ belongs to structural type s can be accepted if the D_s^2 value does not exceed the critical value $D_{s\,crit.}^2$.

Based on the conclusion of the hypothesis test derived above, the given chemical-shift data can either be or not be classified into specific structural type s. However, a given sample may lie in the joint area of two or more structural

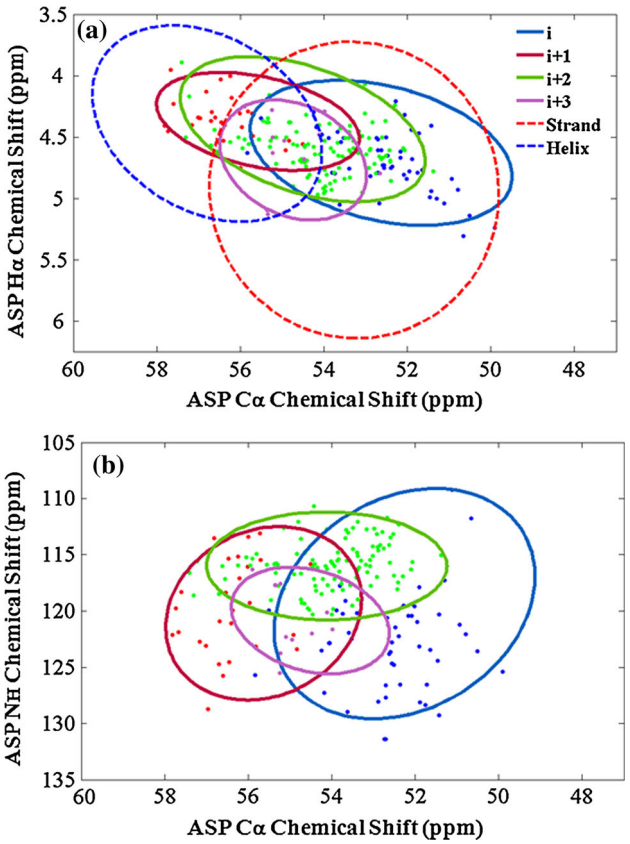


Fig. 3 **a** Two-dimensional H_α/C_α and **b** N_H/C_α chemical shift plots of ASP in type I β-turns. Positions (i), $(i + 1)$, $(i + 2)$, and $(i + 3)$ are colored in *blue*, *red*, *green* and *purple*, respectively. Plot **a** also shows the clusters of helical and extended structures in *blue* and *red dashed-line* boundaries to reveal the joints between the clusters of turn residues and the regular secondary structures

clusters simultaneously in a multi-dimensional coordinate system. To decide whether an observation falls inside a cluster area, the probability score is estimated according to the ratio of the sample size in Table S1 and two scoring matrices of type I and II for four consecutive residues ($AA_{j\sim j+3}$) is established (see Table S2). Here, $Pr(AA_j|i)$ represents the probability of an amino acid (AA_j) located at position (i), $Pr(AA_j|i + 1)$ represents the probability of (AA_j) located at position $(i + 1)$, and so on. Thus, the $\tau(AA_{j\sim j+3})$ value represents the total probability of each four consecutive residues which is not part of helical or extended conformation in a protein:

$$\tau(AA_{j\sim j+3}) = Pr(AA_j|i) + Pr(AA_{j+1}|i+1) + Pr(AA_{j+2}|i+2) + Pr(AA_{j+3}|i+3)$$

The four consecutive residues are candidates for a specific β-turn type when the $\tau(AA_{j\sim j+3})$ value is greater than the minimum threshold. The selection of an appropriate minimum threshold depends on the type of β-turn and, empirically, the $\tau(AA_{j\sim j+3})$ values of type II β-turns are observed to be significantly higher than those of type I β-turns. Therefore, the $Pr(AA_j)_{type\ I} \geqq 0.30$ and $Pr(AA_j)_{type\ II} \geqq 0.50$ are used as minimum thresholds. Furthermore, while the central two residues play an important role in defining β-turn type, a preliminary strategy was tested: the probabilities of $Pr(AA_{j+1}|i + 1)$ and $Pr(AA_{j+2}|i + 2)$ must be not both zero. Four rules are proposed to predict type I and II β-turns using the total probabilities.

Rule 1 A tetra-peptide is predicted to be a type I β-turn if $\tau(AA_{j\sim j+3})_{type\ I} \geqq 1.20$; otherwise apply Rule 2

Rule 2 A tetra-peptide is predicted to be a type II β-turn if $\tau(AA_{j\sim j+3})_{type\ II} \geqq 2.00$; otherwise apply Rule 3

Rule 3 If the predicted state of a tetra-peptide is a type I and II β-turn simultaneously, the predicted state is in any state of type I or II β-turn with the maximum τ value, otherwise apply Rule 4

Rule 4 If the probability of $Pr(AA_j|i)$ or $Pr(AA_{j+3}|i + 3)$ equals zero due to the chemical shift data lost, the tetra-peptide is a type I or II β-turn if and only if the probabilities of the central two residues are both $\geqq 0.50$, that is, $Pr(AA_{j+1}|i + 1) \geqq 0.50$ and $Pr(AA_{j+2}|i + 2) \geqq 0.50$; otherwise no prediction is made

Evaluation of the prediction method

Four measures, namely Q_{total}, $Q_{predicted}$, $Q_{observed}$, and Matthews correlation coefficient (MCC), are used to evaluate the predictive performance of the proposed method. These measures have been applied to many β-turn prediction methods (Shepherd et al. 1999; Kaur and Raghava 2003, 2004; Kim 2004; Asgary et al. 2007; Kirschner and Frishman 2008; Zheng and Kurgan 2008; Petersen et al. 2010; Tang et al. 2011; Song et al. 2012; Shen and Bax 2012). To evaluate these measures, four scalar quantities must be given: true positive (TP), the number of correctly classified β-turn residues, true negative (TN), the number of correctly classified non-β-turn residues, false positive (FP), the number of non-β-turn incorrectly classified as β-turn residues, and false negative (FN), the number of β-turn incorrectly classified as non-β-turn residues.

1. Q_{total} is the percentage of correctly classified β-turns

$$Q_{total} = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

2. $Q_{predicted}$ is the percentage of correctly predicted β-turns among the predicted β-turns

$$Q_{predicted} = \frac{TP}{TP + FP} \times 100$$

3. $Q_{observed}$ is the percentage of correctly predicted β-turns among the observed β-turns

$$Q_{observed} = \frac{TP}{TP + FN} \times 100$$

4. MCC is:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

The MCC value takes into account of both FP (over-predictions) and FN (under-predictions) and is confined to the interval $\langle -1, 1 \rangle$, where a value of 1 indicates a perfect prediction, and a value of -1 indicates an inverse prediction. A value of 0 indicates a random prediction.

Results and discussion

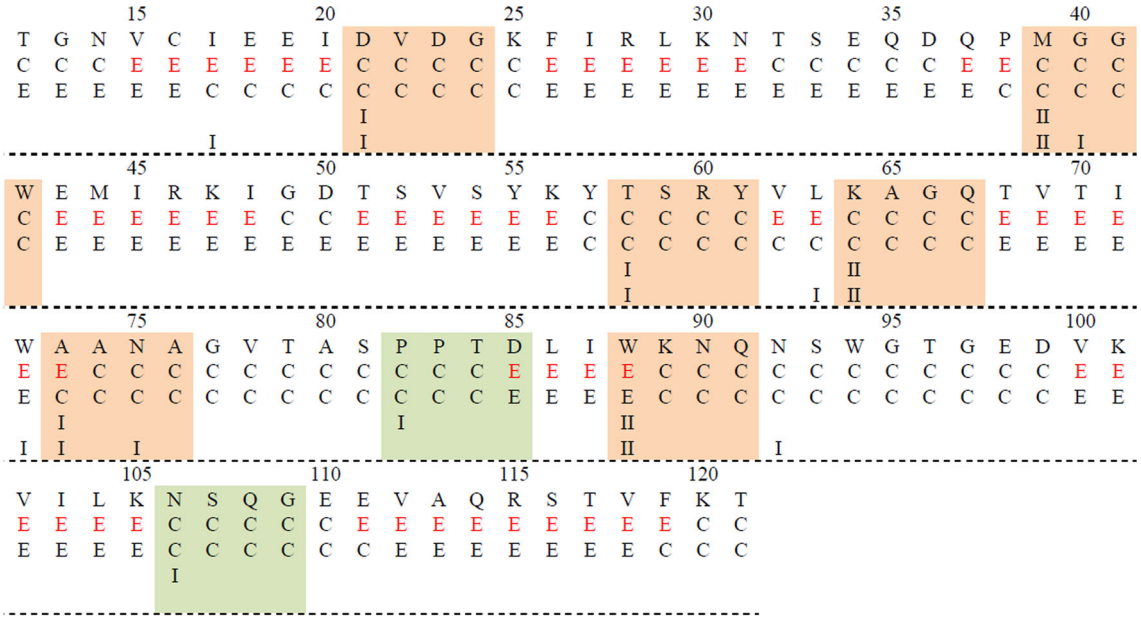
Statistical and cluster analyses of chemical shifts

Nuclear magnetic resonance chemical-shift data of isolated β-turns were extracted from 1,798 re-referenced BMRB entries for our data set. In total, 3,564, 1,058, 559, 246, and 757 β-turns with types I, II, I′, II′, and VIII, respectively, were observed in the data set. The set was analyzed for each residue; however, type I′, II′ and VIII turns had very small numbers while being classified by 20 amino acids. Only type I and II turns are discussed here. In type I β-turns, the most common residue is aspartic acid (ASP), which is strongly favored at positions (i) and ($i + 2$). The second commonest residue is glycine (GLY), which is significantly favored at position ($i + 3$). Proline (PRO) is the most common at position ($i + 1$) among the 20 amino acids. The preference at each position is consistent with previous studies (Hutchinson and Thornton 1994). In type II β-turns, PRO is strongly favored at position ($i + 1$) and GLY is favored at position ($i + 2$). Glutamic acid (GLU) and lysine (LYS) are common in type II turns, but have relatively smaller numbers at ($i + 2$) compared to those at other positions. A detail discussion of other common residues in types I, II, and VIII turns, like asparagines (ASN) and serine (SER), is given in Hutchinson's studies.

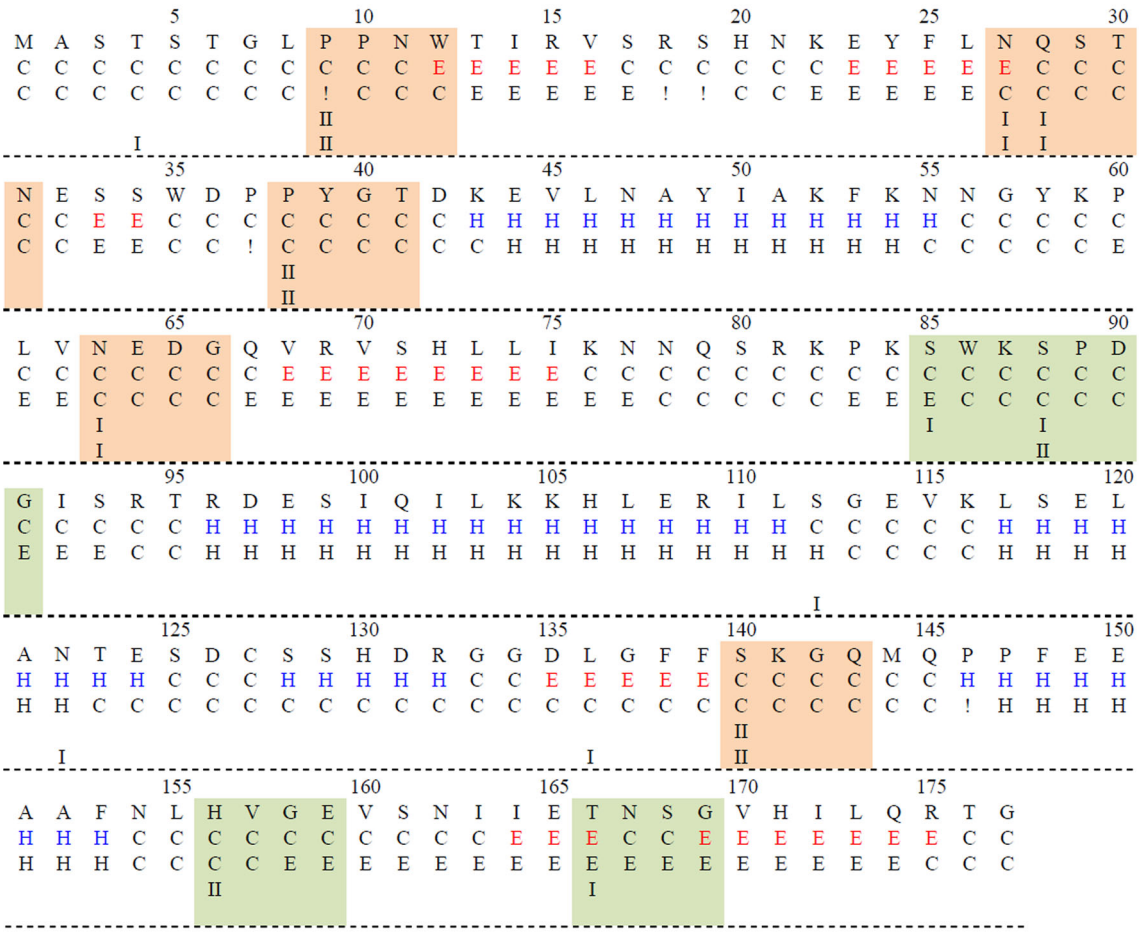
A 1D frequency plot indicates that there are slight shifts between positional distributions. Figure 1a–f shows individually the frequency curves of C_o, C_α, C_β, H_N, H_α, and N_H chemical shifts of ASP in type I β-turns. Four positional distributions, (i) to ($i + 3$), are colored in blue, red, green and purple, respectively. In Fig. 1a, the C_o frequency curve of ASP at position (i) shifted downfield with higher shift value than those at other positions; in contrast, the curve of position ($i + 3$) shifted upfield and the samples appeared spread. The mean ppm values of C_o chemical shifts as a function of position are (175.79 ppm; $i + 3$) < (176.26 ppm; $i + 2$) < (176.59 ppm; $i + 1$) < (176.79 ppm; i). Similar outcomes of the C_o chemical shift mean ppm value trends ($i + 3$) < ($i + 2$) < ($i + 1$) < (i) are observed from 1D frequency plots for the other 19 amino acids. In Fig. 1b, the distinguishing peaks of the central two residues ($i + 1$ and $i + 2$) can be seen. The mean ppm values of C_α chemical shifts as a function of position are (53.19 ppm; i) < (53.71 ppm; $i + 2$) < (54.92 ppm; $i + 3$) < (56.09ppm; $i + 1$). The average C_α chemical shift had a higher chemical shift value at position ($i + 1$) than those at other positions. In contrast, Fig. 1d shows that the ($i + 3$) curve of H_N chemical shift shifted upfield. The H_N chemical shift mean ppm values as a function of position are (7.79 ppm; $i + 3$) < (8.20 ppm; $i + 2$) < (8.24 ppm; i) < (8.58 ppm; $i + 1$). Furthermore, Fig. 1e shows that the H_α chemical shift appears upfield with the lowest average value (4.48 ppm) at position ($i + 1$); Fig. 1f shows that the N_H chemical shift appears upfield with the lowest average value (115.99 ppm) at position ($i + 2$). The C_β chemical shift showed no clear trend.

In Fig. 2a–f for ASP in type II β-turns, the frequency curves of six chemical shifts reveal distinguishable distributions. Figure 2a, b shows respectively that C_o and C_α chemical shift frequency curves of position ($i + 1$) were shifted downfield; whereas the C_β (Fig. 2c) and the H_N chemical shifts (Fig. 2d) had lower shift values at positions ($i + 2$) and ($i + 3$), respectively. In Fig. 2e, all samples were divided into two groups: one was upfield of positions ($i + 1$) and ($i + 2$) with lower shift values; the other was downfield of positions (i) and ($i + 3$) with higher shift values. The H_α chemical shift mean ppm values as a function of position are (4.41 ppm; $i + 2$) < (4.46 ppm; $i + 1$) < (4.72 ppm; i) < (4.73 ppm; $i + 3$). It was worth noting that there was an order, ($i + 1$) ≈ ($i + 2$) < (i) ≈ ($i + 3$), for most of the other amino acids, except that GLY lacked H_α chemical shift information. The last plot (Fig. 2f) shows that some samples shifted upfield at position ($i + 2$), which lowered the mean ppm value of the N_H chemical shift. The numerical descriptive measures including the means and the standard deviations, of six nuclei chemical shifts of 20 amino

Fig. 4 Two examples for predicting β-turn types from **a** BMRB 16572, **b** BMRB 16690. For each *row*, *row 1* is the amino acid sequence; *row 2* is the observed secondary structure (*H* helix, *E* strand, *C* coil); *row 3* is the predicting secondary structure using 2DCSi method; *row 4* is the observed location of β-turn, which first residue is labeled *I* or *II*; and *row 5* is the predicted location of β-turn. In **a** five type I and three type II β-turns were predicted. In **b** six type I and four type II β-turns were predicted. The true positive and false negative predictions are colored in *orange* and *green* blocks, respectively



(a) BMRB16572/ PDB3HN9



(b) BMRB16690/ PDB1YW5

acids are summarized in Supplementary Material I Table S3, which are consistent with the chemical shift patterns for type I and II β-turns reported by previous studies (Shen and Bax 2012). A total of 366 1D frequency plots for 20 amino acids are plotted and placed on 2DCSi(t) web page.

Figure 3a, b shows 2D plots of H_α/C_α and N_H/C_α paired chemical shifts for ASP. The four ellipses in blue, red, green, and purple mark the boundaries of (i) to ($i + 3$) clusters, respectively, and contain 90 % of the data points. Figure 3a also shows the boundaries (dashed lines) of helical and extended structures in blue and red, respectively, revealing that the chemical shift distributions of turn residues overlap with regular secondary structures. 2D plots can be used to distinguish four positional clusters. For example, the (i) cluster shifted upfield with the lowest C_α mean value, but the ($i + 1$) cluster shifted downfield with the highest mean value in Fig. 3a. Considering the C_α shift only, the ($i + 2$) cluster overlapped with others and was difficult to distinguish. However, in Fig. 3b, the ($i + 2$) cluster significantly shifted upfield in the y-coordinate with the lowest N_H mean value. Fifteen paired chemical shifts plot of ASP residues in types I and II turns are given in Supplementary Material II Fig. S1.

Application of scoring matrix to predict β-turns

To demonstrate the feasibility of applying rules of the scoring matrix to predict type I and II β-turns, two entries of BMRB16572 (the related PDB code 3HN9, chain A) and BMRB16690 (PDB code 1YW5, chain A) were tested for example, which are not included in the data set. Figure 4a, b shows the results of the two proteins by applying the four scoring matrix prediction rules. Five type I and three type II β-turns are observed in chain 3HN9 (110 residues) as shown in Fig. 4a. Rule 1 is correctly applied to the first three of five sequences: DVDG, TSRY, AANA, PPTD, and NSQG (residues 21–24, 58–61, 73–76, 63–66, and 106–109, respectively). The $\tau(AA_{j\sim j+3})_{type\ I}$ values of the sequences DVDG, TSRY, and AANA are greater than or equal to 1.2. Thus, the three sequences are predicted as type I β-turns by applying Rule 1. However, no predictions are made for the sequences PPTD and NSQG ($\tau(AA_{j\sim j+3})_{type\ I} = 0.58$ and 1.17, respectively). Two sequences: MGGW and KAGQ (residues 39–42 and 64–67, respectively) are correctly predicted to be type II β-turns by Rule 2. Rule 4 is applied to the third type II β-turn, the sequence WKNQ, while the probabilities of the central two residues are both equal to 0.52 ($\geqq 0.50$).

In Fig. 4b, six type I and four type II β-turns are observed in chain 1YW5 (177 residues). Three type I β-turn sequences NQST, QSTN, and NEDG (residues 27–30, 28–31, and 63–66, respectively) are correctly predicted by Rule 1: the $\tau(AA_{j\sim j+3})_{type\ I}$ values are all greater than 1.20. However, the sequences SWKS, SPDG, and TNSG (residues 85–88, 88–91, and 166–169, respectively) are incorrectly classified as non-type I β-turns. Among four type II β-turn sequences PPNW, PYGT, SKGQ, and HVGE (residues 9–12, 38–41, 140–143, and 156–159, respectively), Rule 4 is applied for the first sequence PPNW, because the probabilities of the central two residues are both greater than 0.50. The two sequences of PYGT and SKGQ are predicted to be type II turns by applying Rule 3. The $\tau(AA_{j\sim j+3})_{type\ II}$ of PYGT is 3.70, which is greater than $\tau(AA_{j\sim j+3})_{type\ I}$ (=1.69). The $\tau(AA_{j\sim j+3})_{type\ II}$ of SKGQ is 2.75, which is greater than $\tau(AA_{j\sim j+3})_{type\ I}$ (=1.37). However, no predictions are made for the sequence HVGE with the given secondary structure. The prediction results of 15 testing proteins, including the above two examples, are given in Supplementary Material III.

Description of 2DCSi(t) web server

In our previous study (Wang et al. 2007), a 2DCSi web server was established to allow users to submit chemical shift data in BMRB format and to predict protein secondary structures and the redox states of cysteine residues. Now the approach using a scoring matrix for four consecutive residues to predict type I and II β-turns is incorporated into the 2DCSi method, a new chemical shift prediction program 2DCSi(t). It is available at http://www.2dsci.idv.tw.

In other sequence-based prediction methods, test databases of BT426, BT547, and BT823 are frequently used; however, they contain a few related BMRB entries available. For evaluating the predictive performance of the proposed method, 15 proteins were collected as an independent test set (see Supplementary Material III). 2DCSi(t) achieves accuracy of $Q_{total} = 83.2$ %, $Q_{predicted} = 32.4$ %, $Q_{observed} = 51.4$ %, and a MCC of 0.32 for type I β-turn; $Q_{total} = 84.2$ %, $Q_{predicted} = 69.2$ %, $Q_{observed} = 61.5$ %, and a MCC of 0.63 for type II β-turn. The comparison of type I and II β-turns prediction accuracy with other prediction methods is reported in Table 2. However, it should be noted that the test sets are not the same.

Conclusion and future developments

The NMR chemical shift trends of various types of β-turns were determined using 1D and 2D cluster analyses. According to the chemical shift propensities at different positions, rules were derived using a scoring matrix to predict type I and II turns. The distributions of chemical shift data are significantly different between two β-turn types. For example, in type I β-turns, the backbone φ, ψ angles of the central two residues are (−60°, −30°) and (−90°, 0°), respectively. The dihedral angles of residue ($i + 1$) are

Table 2 Comparison of type I and II β-turns prediction accuracy with other methods

Method	Q_{total} (%)	$Q_{predicted}$ (%)	$Q_{observed}$ (%)	MCC
Type I				
2DCSi(t)	83.2	32.4	51.4	0.317
BTPRED	91.2	13.9	46.6	0.219
BETATURNS	74.5	22.1	74.1	0.290
COUDES	84.5	30.8	50.0	0.309
Asgray et al.	63.9	51.8	53.6	0.235
MOLEBRNN	82.5	28.9	56.9	0.317
DEBT	78.6	N/A	75.2	0.360
NetTurnP	N/A	N/A	N/A	0.360
Shi et al.	89.1	71.1	27.6	0.398
MICS	N/A	76.0	86.0	0.800
Type II				
2DCSi(t)	84.2	69.2	61.5	0.632
BTPRED	95.5	12.2	58.4	0.253
BETATURNS	93.5	25.5	52.8	0.290
COUDES	91.0	22.2	52.8	0.302
Asgray et al.	89.1	54.1	52.9	0.473
MOLEBRNN	96.2	50.2	25.2	0.339
DEBT	87.4	N/A	64.3	0.290
NetTurnP	N/A	N/A	N/A	0.310
Shi et al.	96.0	66.6	34.3	0.460
MICS	N/A	100.0	54.0	0.730

For BTPRED, values taken from Shepherd et al. (1999); for BETATURNS, values taken from Kaur and Raghava (2004); for COUDES, values taken from Fuchs and Alix (2005); for Asgary et al.'s method, values taken from Asgary et al. (2007); for MOLEBRNN, values taken from Kirschner and Frishman (2008); for DEBT, values taken from Kountouris and Hirst (2010); for NetTurnP, values taken from Petersen et al. (2010); for Shi et al.'s method, values taken from Shi et al. (2011); for MICS method, values taken from Shen and Bax (2013)

located near the α-helical region [(−60°, −60°) for idealized values]. In contrast, the angles of residue $(i + 2)$ leave the α-helical region and approach to the β-strand region [(−120°, 120°) for idealized values]. As a consequence, the C_α chemical shift mean values showed the trend: β-strand < $(i + 2)$ residue of type I β-turn < $(i + 1)$ residue of type I β-turn < α-helix. It is expected that other types of β-turn can be predicted because the dihedral angles of residues $(i + 1)$ and $(i + 2)$ are in different regions on the Ramachandran plot between types.

The identification of the small secondary structure elements is quite important (Shen and Bax 2012). The proposed method gives a significant level of accuracy (MCC = 0.632) for type II β-turns, much better than type I. This is expected because the chemical shift distributions of the central two residues of type II turns are more distinguishable than those of type I turns. It suggests that the type II turn/not-type II turn prediction should be more accurate than type I turn. Compared with other type specific β-turn prediction methods, our approach can achieve better accuracy in terms of MCC for type II β-turn prediction. As for type I β-turn, the predicting accuracy may be improved by giving weighting on the probabilities of four consecutive residues. Moreover, β-turn prediction is highly dependent on the accuracy of the secondary structure prediction (Shepherd et al. 1999). We believe that as more protein chemical shift data becomes available, the type I′, II′ and VIII β-turns prediction using the proposed method will be feasible. It will make the 2DCSi(t) web server a more comprehensive tool for protein secondary structures prediction using NMR chemical shifts.

Acknowledgments The work has been partially financed by National Science Council of ROC, NSC-97-2221-E-006-079-MY3 and NSC-101-2311-B-006-009-MY3. We are grateful for their supports.

References

Asgary MP, Jahandideh S, Abdolmaleki P, Kazemnejad A (2007) Analysis and identification of β-turn types using multinomial logistic regression and artificial neural network. Bioinformatics 23:3125–3130

Beger RD, Bolton PH (1997) Protein φ and ψ dihedral restraints determined from multidimensional hypersurface correlations of backbone chemical shifts and their use in the determination of protein tertiary structures. J Biomol NMR 10:129–142

Berjanskii MV, Neal S, Wishart DS (2006) PREDITOR: a web server for predicting protein torsion angle restraints. Nucleic Acids Res 34:W63–W69

Cheung MS, Maguire ML, Stevens TJ, Broadhurst RW (2010) DANGLE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. J Magn Reson 202:223–233

Chou KC (1997) Prediction of β-turns. J Pept Res 49:120–144

Chou KC, Blinn JR (1997) Classification and prediction of β-turn types. J Protein Chem 16:575–595

Chou PY, Fasman GD (1974) Prediction of protein conformation. Biochemistry 13:222–245

Chou PY, Fasman GD (1979) Prediction of beta-turns. Biophys J 26:367–384

Cohen FE, Abarbanel RM, Kuntz ID, Fletterick RJ (1986) Turn prediction in proteins using a pattern-matching approach. Biochemistry 25:266–275

Eghbalnia HR, Wang L, Bahrami A, Assadi A, Markley JL (2005) Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements. J Biomol NMR 32:71–81

Fuchs PF, Alix AJ (2005) High accuracy prediction of β-turns and their types using propensities and multiple alignments. Proteins Struct Funct Bioinf 59:828–839

Guerry P, Herrmann T (2011) Advances in automated NMR protein structure determination. Q Rev Biophys 44:257–309

Hung LH, Samudrala R (2003) Accurate and automated classification of protein secondary structure with PsiCSI. Protein Sci 12:288–295

Hutchinson EG, Thornton JM (1994) A revised set of potentials for β-turn formation in proteins. Protein Sci 3:2207–2216

Kabsch W, Sander CA (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637

Kaur H, Raghava GPS (2003) Prediction of beta-turns in proteins from multiple alignment using neural network. Protein Sci 12:627–634

Kaur H, Raghava GPS (2004) A neural network method for prediction of β-turn types in proteins using evolutionary information. Bioinformatics 20:2751–2758

Kim S (2004) Protein β-turn prediction using nearest-neighbor method. Bioinformatics 20:40–44

Kirschner A, Frishman D (2008) Prediction of β-turns and β-turn types by a novel bidirectional Elman-type recurrent neural network with multiple output layers (MOLEBRNN). Gene 422(1–2):22–29

Kountouris P, Hirst JD (2010) Predicting β-turns and their types using predicted backbone dihedral angles and secondary structures. BMC Bioinformatics 11:407

Lewis PN, Momany FA, Scheraga HA (1971) Folding of polypeptide chains in proteins: a proposed mechanism for folding. Proc Natl Acad Sci USA 68:2293–2297

Lewis PN, Momany FA, Scheraga HA (1973) Chain reversals in proteins. Biochim Biophys Acta 303:211–229

McGregor MJ, Flores TP, Sternberg MJE (1989) Prediction of beta-turns in proteins using neural networks. Protein Eng 2(7):521–526

Osapay K, Case DA (1994) Analysis of proton chemical shifts in regular secondary structure of proteins. J Biomol NMR 4:215–230

Petersen B, Lundegaard C, Petersen TN (2010) NetTurnP—neural network prediction of beat-turns by use of evolutionary information and predicted protein sequence features. PLoS One 5(11):e15079

Richardson JS (1981) The anatomy and taxonomy of protein structure. Adv Protein Chem 34:167–339

Rose GD, Gierasch LM, Smith JA (1985) Turns in peptides and proteins. Adv Protein Chem 37:1–109

Santiveri CM, Rico M, Jimemez MA (2001) $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts as a tool to delineate β-hairpin structures in peptides. J Biomol NMR 19:331–345

Sharma D, Rajarathnam K (2000) ^{13}C NMR chemical shifts can predict disulfide bond formation. J Biomol NMR 18:165–171

Shen Y, Bax A (2012) Identification of helix capping and β-turns motifs from NMR chemical shifts. J Biomol NMR 52:211–232

Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J Biomol NMR 56:227–241

Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu GH, Eletsky A, Wu YB, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci USA 105:4685–4690

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223

Shepherd AJ, Gorse D, Thornton JM (1999) Prediction of the location and type of β-turns in proteins using neural networks. Protein Sci 8:1045–1055

Shi X, Hu X, Li S, Liu X (2011) Prediction of β-turn types in protein by using composite vector. J Theor Biol 286:24–30

Song Q, Li T, Cong P, Sun J, Li D, Tang S (2012) Predicting turns in proteins with a unified model. PLoS One 7(11):e48389

Tang Z, Li T, Liu R, Xiong W, Sun J, Zhu Y, Chen G (2011) Improving the performance of β-turn prediction using predicted shape strings and a two-layer support vector machine model. BMC Bioinformatics 12:283

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao H, Markley JL (2008) BioMagResBank. Nucleic Acids Res 36:402–408

Venkatachalam CM (1968) Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. Biopolymers 6:1425–1436

Wang Y, Jardetzky O (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. Protein Sci 11:852–861

Wang CC, Chen JH, Yin SH, Chuang WJ (2006) Predicting the redox state and secondary structure of cysteine residues in proteins using NMR chemical shifts. Proteins Struct Funct Bioinf 63:219–226

Wang CC, Chen JH, Lai WC, Chuang WJ (2007) 2DCSi: identification of protein secondary structure and redox state using 2D cluster analysis of NMR chemical shifts. J Biomol NMR 38:57–63

Wilmot CM, Thornton JM (1988) Analysis and prediction of the different types of β-turn in proteins. J Mol Biol 203:221–232

Wilmot CM, Thornton JM (1990) β-Turns and their distortions: a proposed new nomenclature. Protein Eng Des Sel 3:479–493

Wishart DS (2011) Interpreting protein chemical shift data. Prog Nucl Magn Reson Spectrosc 58:62–87

Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol 222:311–333

Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry 31:1647–1651

Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G (2008) CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. Nucleic Acids Res 36:496–502

Zhang CT, Chou KC (1997) Prediction of β-turns in proteins by 1–4 and 2–3 correlation model. Biopolymers 41:673–702

Zhang H, Neal S, Wishart DS (2003) RefDB: a database of uniformly referenced protein chemical shifts. J Biomol NMR 25:173–195

Zhao Y, Alipanahi B, Li SC, Li M (2010) Protein secondary structure prediction using NMR chemical shift data. J Bioinform Comput Biol 8:867–884

Zheng C, Kurgan L (2008) Prediction of beta-turns at over 80% accuracy based on an ensemble of predicted secondary structures and multiple alignments. BMC Bioinformatics 9:430