PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

In silico platform for predicting and initiating β-turns in a protein at desired locations

Harinder Singh, Sandeep Singh, and Gajendra P. S. Raghava*

Bioinformatics Center, Institute of Microbial Technology, Chandigarh, India

ABSTRACT

Numerous studies have been performed for analysis and prediction of β-turns in a protein. This study focuses on analyzing, predicting, and designing of β-turns to understand the preference of amino acids in β-turn formation. We analyzed around 20,000 PDB chains to understand the preference of residues or pair of residues at different positions in β-turns. Based on the results, a propensity-based method has been developed for predicting β-turns with an accuracy of 82%. We introduced a new approach entitled "Turn level prediction method," which predicts the complete β-turn rather than focusing on the residues in a β-turn. Finally, we developed BetaTPred3, a Random forest based method for predicting β-turns by utilizing various features of four residues present in β-turns. The BetaTPred3 achieved an accuracy of 79% with 0.51 MCC that is comparable or better than existing methods on BT426 dataset. Additionally, models were developed to predict β-turn types with better performance than other methods available in the literature. In order to improve the quality of prediction of turns, we developed prediction models on a large and latest dataset of 6376 nonredundant protein chains. Based on this study, a web server has been developed for prediction of β-turns and their types in proteins. This web server also predicts minimum number of mutations required to initiate or break a β-turn in a protein at specified location of a protein.

Proteins 2015; 83:910–921.

Key words: beta turn prediction; analysis of beta turn residue; designing of beta turn; beta turn type prediction; statistical based beta turn prediction.

INTRODUCTION

Proteins play a vital role in living organisms and hence it is essential to understand the function of a protein. The function of a protein depends upon its tertiary structure, which in turn depends upon its secondary structure. The secondary structure is classified mainly into three broad categories: helix, sheet, and coil.[1] The coil region further splits into tight turns (α-turns, γ-turns, δ-turns, π-turns, β-turns), bulges and random coil structures.[2,3] Among these structures, β-turns are the most abundant type of turns; they constitute on an average of 25% of amino acids in proteins.[1,4,5] They are present in disproportionally large number in B-cell epitopes.[6,7] The β-turns are commonly involved in mediating interaction between peptide ligands and their receptors.[8] In protein engineering, loop segments/hairpins are designed by introducing β-turns in proteins/peptides.[9] The structural stability of these peptides/hairpins is mostly determined by β-turns.[10,11] Thus, understanding the formation of β-turns is helpful in understanding various processes, interactions and its contributions to the overall prediction of protein/peptide tertiary structure.[12,13]

In the past four decades, several methods have been developed to predict β-turns. Initially, statistical methods were developed to predict β-turns.[5,14,15] Chou-Fasman used precalculated positional frequencies of residues. Thornton improved Chou-Fasman method by calculating the normalized amino acid positional frequencies. Using these positional frequencies, BTURN and GORBTURN were developed; GORBTURN was further improved in 1994 using revised set of frequencies. Chou improved prediction using amino acid pair frequencies instead of

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Council of Scientific and Industrial Research (project OSDD and GENESIS); Grant number: BSC0121; Grant sponsor: Department of Biotechnology (project BTISNET), Government of India.

Harinder Singh and Sandeep Singh contributed equally to this work

*Correspondence to: G. P. S. Raghava, Scientist & Head Bioinformatics Centre, CSIR-Institute of Microbial Technology, Sector-39A, Chandigarh, India.
E-mail: raghava@imtech.res.in

Received 9 November 2014; Revised 9 February 2015; Accepted 14 February 2015
Published online 27 February 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24783

single amino acid frequencies. Chou observed that interaction between first to fourth and second to third amino acids plays a significant role in β-turn formation. Based on this observation, Zhang and Chou proposed the 1–4 and 2–3 correlation model for the prediction of β-turns.[16] Chou's group further improved their model using sequence coupled approach that is based on the first-order Markov chain.[17] Most of the statistically based methods were implemented in a web server (BetaTPred) developed by Kaur and Raghava.[18] These statistical methods achieved a maximum Q_{total} of 65.2%, $Q_{predicted}$ of 37.6%, $Q_{observed}$ of 63.5%, and MCC of 0.26.

The first machine learning method was developed using neural networks, which achieved an MCC of 0.20.[19] Later, Shepherd *et al.* developed a method BTpred that enhanced the MCC to the 0.34, using secondary structure information.[20] Kim used k-NN to improve the MCC to 0.40,[21] which was further improved to 0.42 (COUDES) by adding propensities, secondary structure, and position specific scoring matrix (PSSM).[22] Kaur and Raghava developed a two-layer neural network, which improved the MCC up to 0.43 (BetaTPred2, BetaTurns).[23,24] The MCC was further improved to 0.45 by MOLEBRNN.[25] Hu and Li used increment of diversity, position conservation scoring function and secondary structure, which raised the MCC up to 0.47.[26] Zheng and Kurgan combined the predicted secondary structure from PSIPRED, JNET, TRANSSEC, and PROTEUS2 to improve the performance.[27] Kountouris and Hirst used predicted dihedral angles apart from PSSM and secondary structure and obtained an MCC of 0.49.[28] Petersen *et al.* developed independent four models for predicting four positions in β-turns and combined these models with standard PSSM and secondary structure model, and achieved the MCC of 0.50 (NetTurnP).[4]

These methods were developed for predicting residues in β-turn, instead of predicting the complete β-turn. Further, the available statistical details of β-turn forming residues are outdated. Hence, the analysis has to be updated with the latest PDB structures for better understanding of β-turn forming residues and different pairs of residues. Biologists are more interested in understanding and initiating or breaking β-turns at a given position in a protein. In this study an attempt has been made to address following issues, (i) analysis of β-turns to understand positional preference of residues, (ii) propensity of a complete turn and contribution of each residue, (iii) models for predicting a complete β-turn, (iv) prediction of all nine types of β-turns, (v) possible minimum mutations required to initiate or break β-turns in a protein.

MATERIALS AND METHODS

Datasets

We used three types of datasets in this study, "Unique," "Standard nonredundant," and "Nonredundant updated" datasets. "Unique" dataset was used for analyzing the preference of residues and residue pairs in β-turns, and for calculation of different propensity scores. To compare our prediction method with other methods, we used "Standard nonredundant" dataset. Finally, a new and updated "nonredundant" dataset (with a large number of protein chains) was used for development of a new prediction method. The model developed with this dataset was implemented in BetaTPred3 web server.

Unique dataset BT20142

This dataset contains a total number of 20,142 high-resolution (<2.0 Å) PDB chains, extracted from the ccPDB server.[29] We also ensured that each protein chain has a minimum of one β-turn. In this dataset, all the protein chains are unique that is, no two protein chains are identical to each other. We used this dataset for analyzing the preference of different type of residues in β-turn at different positions and for calculation of propensity scores.

Standard nonredundant dataset BT426

The standard nonredundant dataset BT426 is a golden dataset that is commonly used for benchmarking β-turn prediction methods.[30] BT426 dataset contains 426 protein chains with resolution better than 2.0 Å and the sequence identity is <25%. Kaur and Raghava used this dataset for the first time for benchmarking β-turn prediction methods.[31] This dataset has been used to compare our method with existing methods as most of these methods (BTpred,[20] KNN,[21] COUDES,[22] BetaTPred2,[32] BTSVM,[33] MOLEBRNN,[25] SVM,[26] BTNpred,[27] E-SSpred,[34] DEBT,[28] NetTurnP[4]) have been evaluated on this dataset. We performed sevenfold and fivefold cross-validation of our method on BT426 dataset and compared the results with existing methods.

Nonredundant updated dataset BT6376

We also created the latest nonredundant dataset (BT6376) from ccPDB server.[29] The minimum resolution of each PDB chain is better than 2 Å and the sequence length varies from 50 to 1000 residues. This dataset contains 6376 protein chains in which no two protein chains have >30% sequence identity. In other words, BT6376 dataset is a subset of BT20142 culled with the sequence identity of <30%. We performed fivefold cross-validation of our method to report the results on the updated nonredundant dataset. The dataset BT6376 created from PDB released in year 2014, whereas dataset BT426 created from PDB released in year 2000. Most of the protein chains or similar chains in BT426 are subsets of protein chains in BT6376. The percentage

of β-turns in proteins is shown in Supporting Information Figure F1 by histogram.

Assignment of β-turns and β-turn types

Promotif software package has been used to assign β-turns in proteins.[35] This is standard software, commonly used to assign β-turns in proteins. Promotif assigns different types of β-turns for example, type I, type I', type II, type II', type IV, type VIa1, type VIa2, and type VIII. The β-turn types were assigned based upon the dihedral angles (φ/ψ) of the two central residues of four consecutive residues forming a β-turn. The ideal values of these dihedral angles are given by Hutchinson and Thornton[35] and are given in Supporting Information Table S1.

Calculation of residue propensities

We calculated the propensity score of occurrence of each amino acid in β-turn using the following equation given by Hutchinson and Thornton:[5]

$$P_t(j) = \frac{f_t(j)}{f_t} \quad (1)$$

where

$$f_t(j) = \frac{\text{Number of residue } j \text{ in turns}}{\text{Number of residue } j \text{ in proteins}}$$

and

$$f_t = \frac{\text{Total number of residues in turns}}{\text{Total number of residues in proteins}}$$

Apart from the amino acid propensity, we also calculated position based propensity of each amino acid in β-turns. Since, four consecutive residues define a β-turn, four position-based propensity scores were calculated: P1, P2, P3, and P4 defining first, second, third, and fourth positions in β-turns. Following equation was used for calculating positional propensities of pairs of amino acids:

$$P_{ti}(j) = \frac{f_i(j)}{f_i} \quad (2)$$

where

$$f_i(j) = \frac{\text{Number of residue } j \text{ at position } i \text{ in turns}}{\text{Number of residue } j \text{ in proteins}}$$

and

$$f_i = \frac{\text{Total number of residues at position } i \text{ in turns}}{\text{Total number of residues in proteins}}$$

Positional preferences of amino acids

We compute positional preferences of each type of residue and pairs of residues. Based on the position of residues in β-turn, we compute propensity of six pairs of residues: (i) P1,2 (residues at positions 1 and 2); (ii) P1,3; (iii) P1,4; (iv) P2,3; (v) P2,4; and (vi) P3,4. Similarly, we also calculate propensity of three consecutive residues in β-turns (P1,2,3 and P2,3,4) and propensity of all four residues in β-turns. We calculated the propensity of all residue pairs in the dataset using the Eq. 4 with i defined as a pair of amino acids.

Input features for prediction of β-turns

Different input features and their combinations were used for development of various models. We used binary profiles, PSSM profiles, predicted secondary structure and β-turn propensity score as input features.

Binary profiles

We converted fixed window length patterns into binary numbers by a vector of dimension $N \times 20$ where N is number of residues in a pattern. Every amino acid is represented by a vector of dimension 20 (for example, Ala is represented 1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0).[36]

Protein profiles

It has been shown in past that evolutionary information in the form of profile provides more information about a protein than its amino acid sequence.[37,38] In this study, we used HHblits (HMM-HMM-based lightning-fast iterative sequence search)[39] for generating PSSM profiles. In the case of protein profile, each residue is represented by a vector of dimension 20.[36]

Secondary structure

We used PSIPRED predicted secondary structure as input feature in β-turn prediction methods.[40] In this study, we used HHblits instead of PSI-BLAST for generating multiple sequence alignments.[39] In order to develop β-turn prediction method, we represented predicted secondary structure by a vector of dimension three where each dimension contains predicted propensity of helix, strand, and coil, respectively. The usage of predicted secondary structure information in the development of models might induce a bias as PSIPRED is also trained on existing structures in the PDB. However, PSIPRED is a standard method for predicting secondary structure of proteins and is used widely in the existing β-turn prediction methods. In order to training our classifier, we used actual secondary structure assigned using DSSP.[40,41]

Propensity scores

We calculate propensity scores of each residue at each position using Eqs. (1) and (2); in addition, we compute

the propensity score of a pair of residues. These propensity scores were used for developing propensity-based models for predicting β-turns.

Classifiers

In this study, we developed prediction models using various classifiers (for example, Random Forest,[42] IBK,[43] Logistic,[44] J48,[45] Multilayer Perceptron,[46] Naïve Bayes[47]) implemented in Weka package.[46] It was observed that Random forest classifier is most suitable for predicting β-turns in a protein. In this study, we have used the FastRandomForest, which is a re-implementation of Random Forest in Weka with better speed and memory utilization.

Turn level prediction

Till date, β-turn prediction methods follow residue level prediction, where each residue was predicted as either β-turn or non-β-turn. As four consecutive residues form a β-turn, a single or double or triple residues predicted as β-turn and its neighboring residues being predicted as non-β-turn, is unreasonable. We therefore focused on turn level prediction, where four consecutive residues were predicted as β-turn. A sliding window of four residues was used for the prediction. If all the four residues of the sliding window make a β-turn then it was defined as a positive pattern and if any of the residue(s) in the window is a non-β-turn then the pattern was defined as a negative pattern. The turn level classification enables realistic prediction as compared with residue level prediction approaches. Various models were developed by increasing the length of sliding window from 6 to 20 to observe the effect of neighboring residues on the prediction. With a sliding window of four residues, a vector of size 93 was constructed, with PSSM ($4 \times 20 = 80$), secondary structure ($4 \times 3 = 12$) and propensity score of tetrapeptide ($1 \times 1 = 1$) as input features. The turn level approach combined with propensity score was used for prediction of nine β-turn types. Separate models were developed for type I, type I', type II, type II', type IV, type VIa1, type VIa2, type VIb, and type VIII using "one versus rest" approach. Therefore, for each turn type, the model was developed considering one β-turn type as positive set and rest of the data as negative set.

Residue level comparison

In order to compare our turn level prediction method with previous β-turn prediction methods, which were developed at residue level, we converted our turn level prediction into residue level prediction for an overall comparison. With a window length of four, each residue occurs in four windows/patterns and each pattern has a predicted turn score. Out of these four turn scores, we assigned the maximum score to the concerned residue.

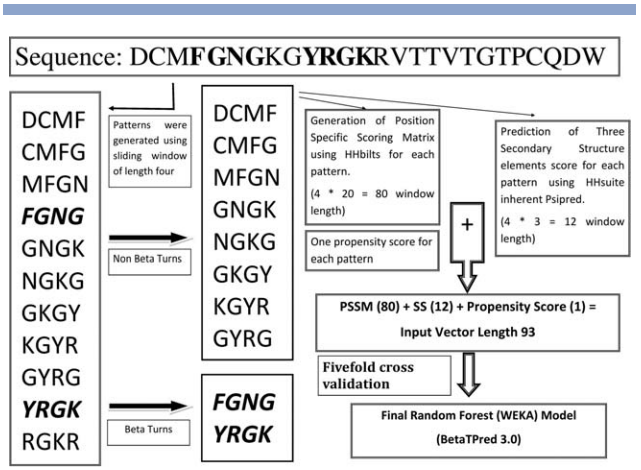

Figure 1

Flowchart of BetaTPred3 Algorithm displaying the development of BetaTPred3.

Propensity based prediction

Statistical models were developed using propensity scores for predicting β-turns at turn level. The statistical method is simple and computationally fast, as it does not require complex models for prediction. We used different propensity scores for prediction of β-turns as given below: (i) propensity of individual residues; (ii) position-wise propensity of residues; (iii) pair-wise residue propensities; (iv) propensities of tripeptides; (v) tetrapeptide propensities; and (vi) Hybrid (average of all propensity scores).

Designing of prediction method

The overall architecture of BetaTPred3 is displayed in Figure 1. The protein sequence is converted into sliding window of four consecutive residues. If these four residues are making a β-turn we label it as positive, else we label it as negative data. Each window is transformed into a feature vector incorporating PSSM profile generated by HHblits and predicted secondary structure by PSIPRED and propensity score of the tetrapeptide. The proposed method predicts whether the four residues make a β-turn or non β-turn instead of predicting a single residue to be in β-turn or non β-turn.

Cross validation and performance measures

In this study, we performed a fivefold cross validation technique on all datasets. We also performed sevenfold cross validation on BT426 dataset to compare with existing methods. For performance measurement, we used various measures: (1) Q_{total} (or prediction accuracy) is

Table I
Propensity Score of Residues in β-Turns and Score Based upon their Positional Preferences in β-Turns

Residue	β-turns	P1	P2	P3	P4
P	1.58	1.62	2.33	1.13	1.25
G	1.58	1.19	1.20	2.10	1.83
N	1.52	1.47	1.27	2.05	1.27
D	1.46	1.57	1.40	1.80	1.09
S	1.20	1.32	1.24	1.13	1.12
H	1.13	1.20	1.00	1.20	1.12
T	1.01	1.06	0.92	0.98	1.09
C	1.01	1.27	0.73	0.86	1.18
K	0.96	0.83	1.16	0.89	0.97
Y	0.93	1.03	0.86	0.91	0.94
W	0.89	0.89	0.84	0.83	0.98
E	0.88	0.75	1.10	0.89	0.79
Q	0.87	0.79	0.89	0.85	0.96
F	0.85	0.99	0.72	0.80	0.89
R	0.84	0.77	0.89	0.82	0.88
A	0.77	0.79	0.88	0.62	0.79
M	0.67	0.75	0.60	0.54	0.79
V	0.65	0.69	0.61	0.48	0.81
L	0.64	0.77	0.59	0.51	0.69
I	0.59	0.68	0.57	0.45	0.68

The table is sorted in descending order based on residue propensity in β-turns.

the percentage of correctly classified residues. (2) Q_{obs} (observed/sensitivity) is the percentage of observed β-turns that are correctly predicted. (3) Q_{pred} (predicted)/PPV (predicted positive value) is the percentage of correct prediction of β-turns. (4) Specificity is the percentage of correct prediction of non β-turns. (5) MCC accounts for both over and underpredictions. (6) Area under curve (AUC) was also calculated by plotting sensitivity against the false positive rate.[48]

RESULTS

Analysis of β-turns in proteins

The BT20142 dataset was used for calculating the propensity score of residues and pair of residues at different positions in β-turn. An excel sheet with all the analysis and propensity values can be accessed at http://crdd.osdd.net/raghava/betatpred3/download.html#analysis. Following is a brief discussion of the analysis we performed.

β-turn forming residues

In this study, we observed that proline and glycine have the highest propensity scores and seem to play a major role in β-turn formation in proteins followed by asparagine and aspartic acid (Table I). Position wise analysis on the preference of amino acids in β-turns emphasizes the importance of residues proline, aspartic acid and asparagine at positions 1 and 2; glycine, asparagine and aspartic acid in position 3; and glycine, asparagine, proline and cysteine in position 4. We further observed that (i) proline is mostly favored at second position in β-turn, (ii) glycine favors third and fourth positions, (iii) asparagine at third position, and (iv) cysteine at first and fourth positions, which might be due to the stabilization of β-turn by forming a disulfide bridge between thiol group of first and fourth cysteine residues in proteins. Cysteine is not a strong β-turn former and is less favored at second and third positions.

β-turn breaking residues

The residues that are least preferred in β-turn formation are also important, as they would help to break β-turns. Analysis of the propensity scores reveals that isoleucine is the least preferred residue in β-turn formation followed by leucine, valine and methionine. Positional preferences suggest that isoleucine is the least preferred in all four positions and can be referred as the strongest β-turn breaker. Following isoleucine, valine (at position first and third) and leucine (at position second and fourth) are least preferred residues (Table I).

Pair of residues as β-turn formers and breakers

Positional preferences of the pair of amino acids correlate well with the analysis of amino acid positional preferences. It was observed that proline is dominated at second position for example, P1-2 (DP, NP, HP, CP, and so forth), P2-3 (PG, PN, PD, PH, and so forth), and P2-4 (PC, PG, PT, PH, and so forth). Similarly, asparagine and glycine are preferred at third and fourth positions, e.g. P1-3 (DN, CN, NN, PN, VG, PG, CG, and so forth), P2-3 (PG, PN, NG, DN, EN, and so forth), and P3-4 (DG, NG, GQ, GK, NN, NP, and so forth). As expected, pair of cysteine residues are the most favored when cysteine occupies first (P1) and fourth positions (P4). Further analysis suggests that when second position is occupied by proline, possible β-turn formers at first position are aspartic acid, asparagine and histidine; at the third position are glycine, asparagine and aspartic acid; while at the fourth position, cysteine and glycine are the most favored ones (Fig. 2). The pairing of valine-isoleucine, valine-leucine, isoleucine-isoleucine, methionine-leucine, and valine-valine are the strongest β-turn breakers, followed by pairs of isoleucine-leucine, methionine-isoleucine and methionine-methionine, and so forth.

Role of tripeptide combinations as β-turn formers and breakers

In the case of tripeptide, aspartic acid-proline-asparagine, cysteine-glutamine-asparagine, histidine-proline-asparagine, methionine-tyrosine-lysine are the most preferred combinations at P1-2-3, while cysteine-tyrosine-cysteine, proline-asparagine-cysteine, proline-aspartic acid-glycine, proline-glycine-glutamine and

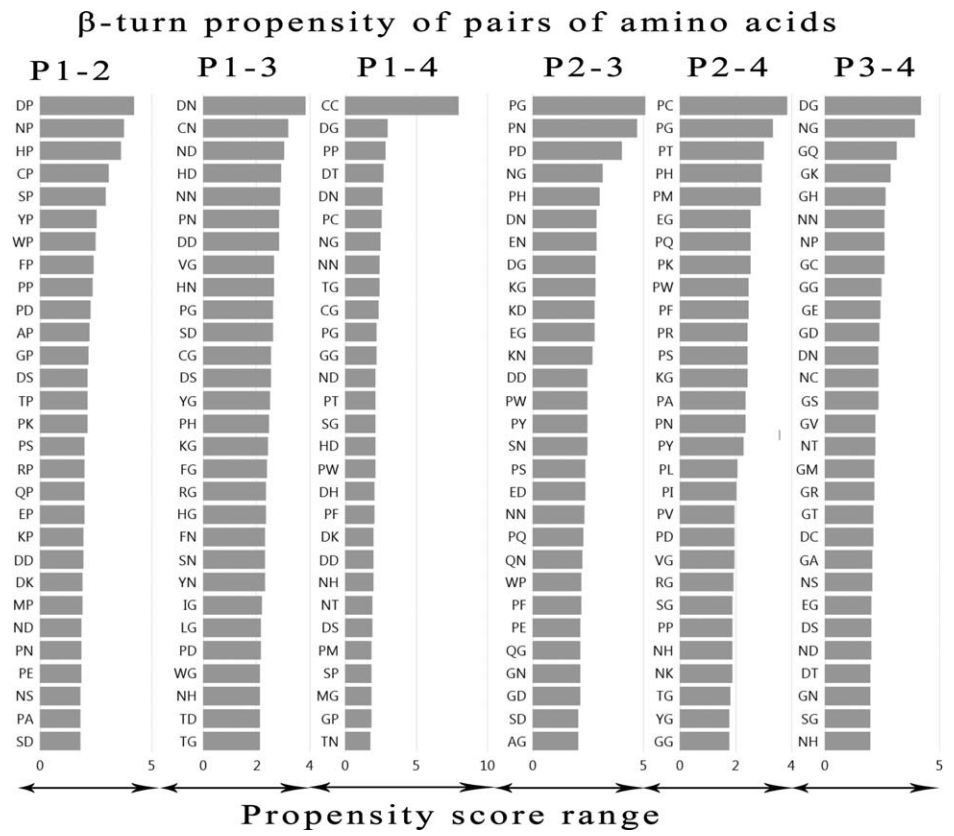

Figure 2

Pairs of amino acids at different positions, which are most favored in the formation of β-turns.

proline-histidine-tryptophan are preferable at P2-3-4 positions. This is consistent with the observations of single as well as pair wise analysis, except glutamine, which is also favored with proline and glycine (Fig. 3). There are few tripeptides having high propensity score for both P1-2-3 and P2-3-4 positions e.g. WPS, WPW, PHW, and so forth. We observed a unique case of cysteine-tyrosine-cysteine (CYC), which is favored in β-turns, if followed by glycine. Out of 81 CYC that form β-turn, 77 have glycine at first or fourth position. Majority of these protein chains having CYC in β-turn are pancreatic or venom phospholipase, suggesting a possible role of CYC tripeptide in stabilization of protein structure.

Tetrapeptides as β-turn formers and breakers

Due to the high number of combinations of tetrapeptides (160,000), the occurrence of some of these combinations in the dataset becomes less, accounting for one or two instances. Therefore, many such low occurrence combinations occurred in β-turns that made their propensity value higher. To identify strong β-turn formers, tetrapeptides having higher propensity score and higher occurrences (at least 50 turns) were selected for analysis. The observations were consistent with the residue wise, pair wise and tripeptide analysis. In general, the propensity of a tetrapeptide was high if the content of β-turn formers are high and are present at their favored positions (Supporting Information Table S2). Some of the most favored tetrapeptides are DGDT, GPDG, NAGD, AGDR, IGIG, DKYG, GDSG, DKGT, EKYG, GIGG, GAGG, LPDG, LSSG, VNGH, YKGQ, DENG, TPDG, and DSDG, and so forth A list of top 20 most frequent and rare tetrapeptides forming β-turns are shown in Supporting Information Table II. It is also interesting to know the tetrapeptides, which always form β-turn (ATWC, WPRR, HKGQ, WPNQ, CTSH, NPHW, and so forth) and that never form β-turn (LRID, LRLK, EMLR, TGTW, AALI, RLKI, etc). A brief list of these interesting peptides is shown in Supporting Information Table S3 and comprehensive list is provided in excel file.

Prediction of β-turns using statistically based method on BT6376 dataset

We developed a propensity-based method for β-turn prediction and achieved reasonable accuracy (Table II). Different propensity scores were pre-calculated from

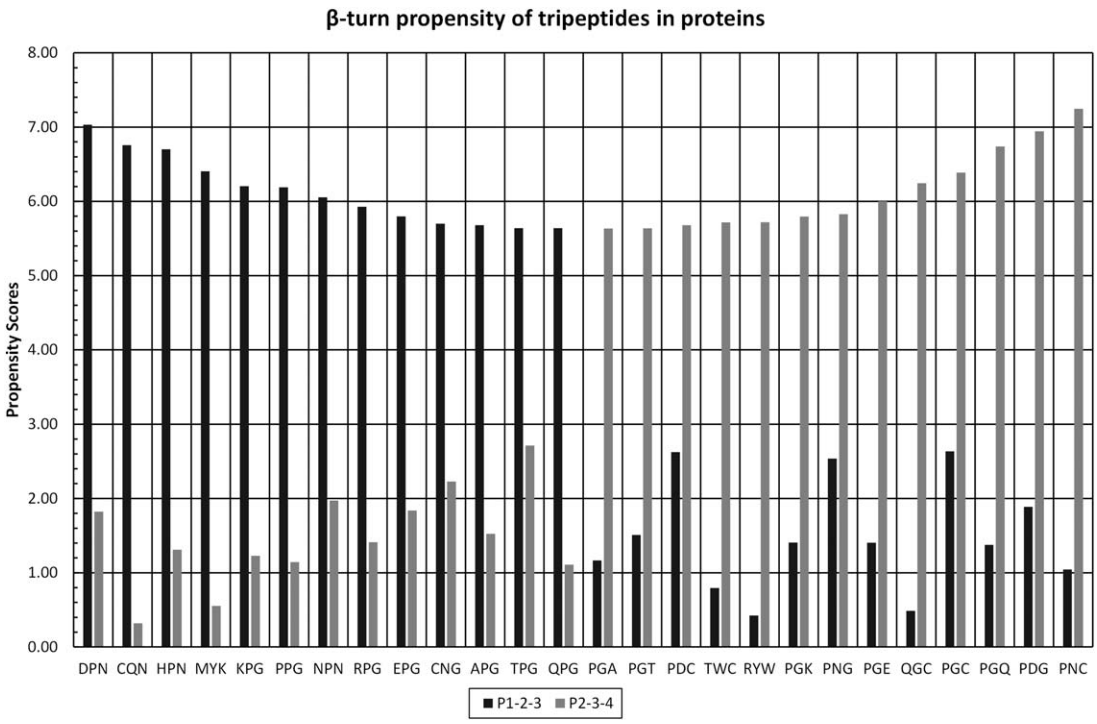

Figure 3

Tripeptides, which are favored by the formation of β-turns and having high propensity score for P1,2,3 and P2,3,4 and low propensity score at other positions.

BT20142 dataset and applied on BT6376 dataset. For performance calculation, we averaged the propensity scores of all residues/residue pairs in a β-turn and used it as a threshold for the prediction of β-turns. The performance of propensity-based method improved from position wise propensity based prediction (0.18 MCC) to pair wise propensity based prediction and finally reaches maximum (0.27 MCC) using tetrapeptide propensity. The results show the effect of pairing; the performance of prediction method increase with number of increase in pair (for example, two to three, three to four).

Prediction of β-turns on BT426 dataset

The BT426 developed in 2000 is the golden dataset for benchmarking of a new β-turn prediction method with existing methods. The BT426 dataset has 9481 β-turns out of 93,702 total patterns of tetrapeptides. We performed fivefold cross validation for evaluating our method on this dataset. Among various classifiers (Random forest, IBK, Logistic, J48, Multilayer Perceptron, and Naïve Bayes) used for the prediction of β-turns, we found Random forest classifier performed the best (Supporting Information Table S4–S7). We developed three binary-based turn level prediction models (Table III) using different input features. The simple binary-based model showed a poor MCC of 0.15 and the performance increased with the inclusion of secondary structure (MCC: 0.31) and β-turn propensity score (MCC: 0.37). These results indicate an important role of secondary structure in the performance improvement. Similarly, we

Table II

The Performance of Propensity Based Method Developed for Predicting β-Turns, Evaluated on BT6376 Dataset

Type of propensity	Q_{pred}	Q_{obs}	Specificity	Q_{total}	MCC	AUC
Residue propensities	19.97	41.86	82.95	79.16	0.18	0.70
Position wise residue propensities	23.35	40.06	86.63	82.34	0.21	0.71
Pair wise propensities	17.88	65.63	69.37	69.02	0.21	0.72
Tripeptide propensities	19.63	57.76	75.96	74.28	0.22	0.72
Tetrapeptide propensities	25.76	52.07	84.75	81.74	0.27	0.77
Hybrid	16.78	73.71	62.85	63.86	0.22	0.73

Table III
The Performance of Turn Level Prediction Method Developed Using Different Features on BT426 Dataset

Input feature	Input features	Qpred	Qobs	Specificity	Qtotal	MCC	AUC
Binary	80	17.93	43.57	77.56	74.12	0.15	0.65
Binary + SS	92	23.07	79.34	70.22	71.14	0.31	0.80
Binary + SS + score	93	30.26	69.35	82.01	80.73	0.37	0.84
PSSM	80	31.99	48.1	88.49	84.4	0.31	0.79
PSSM + SS	92	33.25	55.58	87.44	84.21	0.35	0.84
PSSM + SS + score	93	39.44	61.45	89.38	86.55	0.42	0.87

PSSM: position specific substitution matrix; SS: predicted secondary structure; score: propensity score of β-turn.

Table IV
The Performance of our Models at Residue Level Developed on BT426 Dataset

Input feature	Input features	Q_{pred}	Q_{obs}	Specificity	Q_{total}	MCC	AUC
Binary	80	49.32	19.82	93.36	75.28	0.19	0.67
Binary + SS	92	41.49	89.59	58.81	66.38	0.42	0.80
Binary + SS + score	93	50.37	76.29	75.49	75.69	0.46	0.83
PSSM	80	47.46	72.60	73.80	73.50	0.41	0.81
PSSM + SS	92	50.24	78.87	74.53	75.59	0.47	0.84
PSSM + SS + score	93	55.5	76.0	80.0	79.1	0.51	0.86

developed three PSSM based models; simple PSSM based model yielded an MCC of 0.31 and the inclusion of secondary structure and β-turn propensity enhanced the MCC up to 0.35 and 0.42, respectively as shown in Table III. The combination of PSSM, secondary structure and β-turn propensity score showed the best performance among various features (Q_{pred}: 39.4%; 61.5% Q_{obs}: 61.5%; specificity: 89.4%; Q_{total}: 86.6%; MCC: 0.42 and AUC: 0.87).

It was observed that increasing the window length from four to twenty, has no effect on the performance of turn level prediction (Supporting Information Table S8). β-turns have fixed length of four residues and are present in loop regions of the protein. The PSSM profiles of the loop region are not conserved due to high variation in sequence profile. Thus, the information from the four residues defining a β-turn is sufficient to predict the pattern as β-turn or non-turn.

For a fair comparison with previously developed methods of β-turn prediction, which are based on residue level prediction we transformed the turn level prediction score into residue level prediction score as described in Residue Level Comparison section. The best binary based method achieved the MCC of 0.46 and the best PSSM based method achieved 55.5% Q_{pred}, 76% Q_{obs}, 80% specificity, 79% Q_{total}, 0.51 MCC and 0.86 AUC (Table IV).

Using the best performing model of PSSM, secondary structure and β-turn propensity score, we developed the new method, BetaTPred3. As shown in Table V, the

Table V
Comparison of our Method with Existing Methods Developed for Predicting β-Turn, the Performance was Evaluated at Residue Level on BT426 Dataset

Method	Q_{total}	Q_{pred}/PPV	Sensitivity/Q_{obs}	Specificity	MCC	AUC
BetaTPred3	79.1	55.5	76	80	0.51	0.86
BetaTPred3-Tweak	81.8	58.4	64.4	89.5	0.5	0.86
BetaTPred3-7fold	79	55.3	75.8	80.1	0.51	0.86
NetTurnP	78.2	54.4	75.6	79.1	0.50	0.86
DEBT	79.2	54.8	70.1	N/A	0.48	0.84
E-SSpred	80.9	63.6	49.2	N/A	0.44	0.84
BTNpred	80.9	62.7	55.6	N/A	0.47	N/A
SVM	79.8	55.6	68.9	N/A	0.47	0.87
MOLEBRNN	77.9	53.9	66	N/A	0.45	0.83
BTSVM	78.7	56	62	N/A	0.45	N/A
BetaTPred2	75.5	49.8	72.3	N/A	0.43	0.77
COUDES	75.5	49.8	66.6	N/A	0.41	N/A
KNN	75	46.5	66.7	N/A	0.4	N/A
BTPRED	74.9	55.3	48	N/A	0.35	N/A
1-4 and 2-3 Correlation model	59.1	32.4	61.9	N/A	0.17	N/A

Table VI
The Performance of our Model and Existing Methods Developed for Predicting Type of β-Turns, all Methods Evaluate at Residue Level on Dataset BT426

		β-Turn prediction methods					
β-Turn types	β-Turn count	Mole-brnn	Cou-des	Beta-turns	DEBT	Net-TurnP	BetaT-pred3
Type I	2752	0.31	0.30	0.29	0.36	0.36	0.39
Type I'	301	0.35	0.22	N/A	N/A	0.23	0.47
Type II	982	0.33	0.30	0.29	0.29	0.31	0.42
Type II'	167	0.13	0.10	N/A	N/A	0.16	0.31
Type IV	2871	0.23	0.10	0.23	0.27	0.27	0.26
Type VIa1	43	N/A	N/A	N/A	N/A	N/A	0.27
Type VIa2	18	N/A	N/A	N/A	N/A	N/A	0.13
Type VIb	70	N/A	N/A	N/A	N/A	N/A	0.38
Type VIII	724	0.10	0.07	0.02	0.14	0.16	0.14

performance of BetaTPred3 was comparable to existing residue level β-turn prediction methods. The BetaTPred3 with sevenfold cross validation achieved 55.3% Q_{pred}, 75.8% Q_{obs}, 80.1% specificity, 79% Q_{total}, 0.51 MCC and 0.86 AUC. After tweaking the results, we achieved the highest accuracy of 81.8% with highest MCC of 0.51. The term tweaking refers to reporting the highest Qpred, Specificity, and Qobs at 65% from the same prediction model. The turn level prediction method is nonambiguous in terms of β-turn prediction that is, either the four consecutive residues will be classified as β-turn or non-turn that represents the realistic approach.

The turn level prediction approach was further used to predict the nine β-turn types in BT426 dataset. Due to the small size of BT426 dataset, type II', VIa1, VIa2 and VIb comprises less than 0.5% of total data. For a fair comparison between BetaTPred3 with different β-turn type prediction methods, the turn level prediction score was transformed to residue level prediction score. We achieved an MCC of more than 0.30 for β-turn type I, I', II, II' and VIb (Table VI). In the case of β-turn type VIa2 and VIII the MCC was below 0.20. It was observed that BetaTPred3 performs better than existing methods in the prediction of all β-turn types, except type IV and VIII. In the case of type IV there are sufficient data for training the model, yet the performance remains poor because type IV turns have, no defined φ and ψ angle for $i+1$ and $i+2$ residues. For the first time, BetaTPred3 was able to predict β-turn type VIa1 and VIb with acceptable MCC.

Prediction of β-turns on BT6376 dataset

In real life scenario, the performance of a method on unknown data improves with an increase in the data size. Thus, we created a new and large nonredundant dataset having 6376 protein sequences using the ccPDB server. We performed fivefold cross validation of our method on this dataset. The same optimization parameters were used to develop Random forest based model. The model achieved 87.08% Q_{total}, 38.02% Q_{pred}, 63.61% Q_{obs}, 89.46% specificity, 0.43 MCC and 0.88 AUC (Table VII, Fig. 4). We observed an increase of 0.1 in MCC and 4% increase in AUC as compared with the performance on BT426 dataset. As Qpred is inversly proportional to Qobs, if we increase Qpred then Qobs will decrease accordingly. For example at 61.24% Qpred, the Qobs decreases to 29.89%, but specificity increases to 98.08% (Supporting Information Table S9). Thus, at higher Qpred the predicted β-turn will be highly accurate, but fewer β-turns will be predicted. Similarly, at lower Qpred, more β-turn will be predicted with low accuracy of correct prediction. The β-turn type prediction performance improved due to sufficient training data as compared with smaller BT426 dataset. BetaTPred3 achieved

Table VII
The Performance of our Models Developed for Predicting β-Turn Types, Models were Evaluated at Turn Level on BT6376 Dataset

β-Turn types	No. turns	Q_{pred}	Q_{obs}	Specificity	Q_{total}	MCC	Max. MCC
β-turn	131,862	38.02	63.61	89.46	87.08	0.43	0.43
Type I	42,393	16.35	68.95	88.91	88.3	0.3	0.35
Type I'	4353	45.03	44.84	99.68	99.36	0.45	0.47
Type II	13,559	21.09	62.01	97.31	96.9	0.35	0.44
Type II'	2643	32.98	34.54	99.66	99.34	0.33	0.37
Type IV	38,201	11	54.46	87.46	86.55	0.2	0.22
Type VIa1	654	28.18	38.23	99.65	99.42	0.33	0.38
Type VIa2	188	19.12	34.57	99.51	99.29	0.25	0.28
Type VIb	1082	26.85	46.3	99.48	99.26	0.35	0.38
Type VIII	10,111	12.97	25.25	98.45	97.78	0.17	0.21

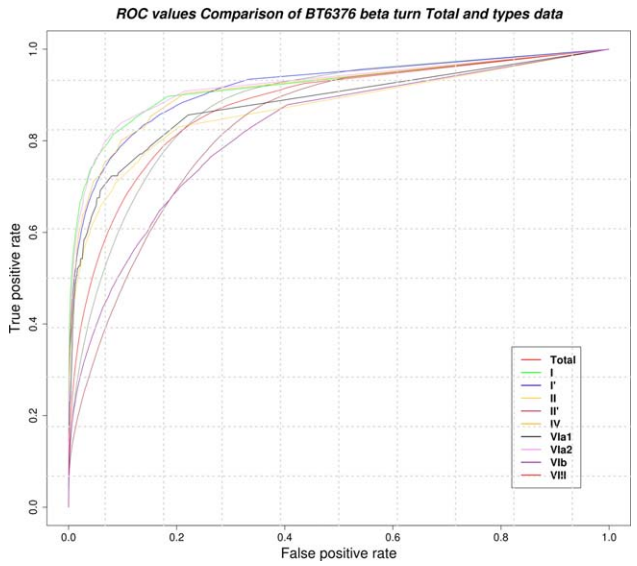

Figure 4

ROC plot using turn level prediction of β-turn types on BT6376 dataset by BetaTPred3.

MCC above 0.31 for all β-turn types, except type IV and type VIII prediction (Table VII, Fig. 4). The variation between best MCC and maximum MCC is due to less number of β-turn types as compared with non-β-turn in the respective dataset.

Figure 4 shows the ROC/AUC value of BT6376 dataset and nine β-turn types. It was observed that β-turn and types have ROC value higher than 0.85, except for type IV (0.84) and type VIII (0.82). The ROCR package[49] in R statistical language[50] is used to illustrate the Figure 4. Using the same optimization parameter of Random Forest, there is a marginal difference in prediction quality across smaller BT426 and larger BT6376 dataset. Evaluating BetaTPred3 on small and large dataset showed that both models are very stable with only 0.53% difference in Q_{total}.

IMPLEMENTATION OF WEB SERVER

To serve the scientific community, we developed a web server BetaTPred3, available at http://crdd.osdd.net/raghava/betatpred3. The web server can be used to predict β-turns and types in proteins using turn level prediction approach. The BetaTPred3 web server is divided into four modules, each having different function.

1. Prediction: This module implements Random forest based model build on BT6376 dataset for the prediction of β-turns.
2. Propensity: This module implements various propensity scores based method for the prediction of β-turns. The output displays the normalized propensity score for each turn.
3. Turn type: This module helps in the prediction of nine β-turn types in a given protein sequence.
4. Design: This module is a unique feature of BetaTPred3, which helps the user in designing (initiating or breaking) a β-turn at desired positions in the protein. In this module, first all possible 80 mutants are generated for a given pattern of four residues. Secondly, β-turns propensity of each mutant peptide is computed using our statistical models. The mutant amino acid of the mutant pattern having the highest and the lowest probability score to form β-turn are represented as "β-turn initiating mutation" and "β-turn breaking mutation" respectively. The design module displays the probable mutation to increase or decrease the β-turn probability score based on the residue and pair wise residue β-turn propensity score obtained from PDB.

DISCUSSION

β-turns play an important role in defining the tertiary structure of proteins. Initially, statistical based methods were developed to predict β-turns, which achieved the maximum MCC of 0.26. These methods were based on the frequency and positional preference of amino acids in β-turn using very small dataset of protein structures. Later, machine learning methods were developed, which enhanced the MCC to 0.50. The latest method NetTurnP utilizes the positional preference of amino acids at first, second, third, and fourth positions for improving the performance. Thus, it can be concluded that positional preference of amino acids in β-turns enhances the prediction performance. In lieu of these methods, we have performed an exhaustive study of positional preferences of amino acids, pair of residues, tripeptides and tetrapeptides on a large dataset of 20142 PDB chains.

We observed that glycine, proline, asparagine, and aspartic acid are favored in β-turns and are called β-turn formers. Glycine having flexible side-chain movements can turn the polypeptide chain to 180° and initiate turns. Proline is known to form kinks and de-regularize the ordered local structure of proteins and therefore, is preferred to form irregular secondary structures like β-turns. Asparagine side-chain can form hydrogen bonds with the peptide backbone; it is found near the beginning and ending of helices and turns. Aspartic acid and asparagine differ only in their functional groups at C-gamma position in which the former has amide group and the latter has a hydroxyl group. Despite different functional groups, both asparagine and aspartic acid have similar capability of forming hydrogen bonds. β-turn breakers include isoleucine, leucine, valine and methionine, which are mainly hydrophobic in nature. It was also observed that different amino acids have different positional preferences e.g. proline is favored at first and second positions, asparagine and aspartic acid at third position, while glycine at third and fourth positions.

We compare residue propensity scores obtained in this study and propensity score described in the previous study obtained from 12 proteins;[15] the trend of prominent β-turn formers is generally the same. However, we noticed a change in the order of amino acids (Supporting Information Table S10). Glutamate and glutamine, which were prominent β-turn breakers in the previous analysis, are neutral while methionine and isoleucine, which were not prominent β-turn breakers, are observed as potential breakers in this study. Position wise analysis has a similar trend except that the first and fourth positions prefer proline and glycine, respectively as the prominent β-turn forming residues in this study, instead of tryptophan, which occurred in the previous analysis. Another analysis performed by Chou and Fasman[14] based on 29 proteins is more similar to our study with respect to the residue wise propensities of prominent β-turn formers and breakers (Supporting Information Table S10). The residue wise propensities calculated by Hutchinson and Thornton[5] (1994) also follow the same trend with our analysis, with a change in the order of amino acids as prominent β-turn formers and breakers (Supporting Information Table S10).

Next, we calculated the pair wise propensity at positions P1-2, P1-3, P1-4, P2-3, P2-4, and P3-4. In the case of P1-2, proline is favored at second position; asparagine, glycine, and aspartate are favored at third position in P1-3. The pair of CC is the most favored pair at P1-4, due to the formation of disulfide bonds, followed by glycine dominated at fourth position. Similarly, proline is more favored at P2-3, P2-4, and glycine and asparagine more favored at P3-4 than other amino acid residues. Therefore, the positional preferences of pair of amino acids have a similar tendency to the positional preference of single amino acids. Although Zhang and Chou have performed the pair wise analysis, they focused only on P1-4 and P2-3 pairs (1-4 and 2-3 correlation model) for improving β-turn prediction. Compared with the previous analysis by Zhang and Chou (based on only 29 proteins), there is a vast difference in the residue pair preferences at positions P1-4 and P2-3. Residue pairs such as GQ, GR, FF, GM, KW, QW, DI, GE, and so forth (generally having glycine at P1) are prominent at P1-4 in the previous analysis. However, our results showed the dominance of residue pairs CC, DG, PP, DT, NG, TG, CG, PG, GG, and so forth (glycine at P4) at P1-4. In the case of P2-3, residue pairs such as GF, CH, GH, YH, GN, NC, GA, GK, and so forth are favored in the previous analysis whereas PG, PN, PD, NG, PH, DN, EN, and so forth are preferred in the present analysis. This noticeable difference is due to the under-representation of residue pairs with a small dataset of 29 proteins. In the present study, β-turn propensity score for all the residue pairs have been computed with an updated large dataset and the results are comprehensive and more reliable than other existing studies.

Further, tripeptides at P1-2-3 favors proline and glycine/asparagine at second and third positions while glycine is more favored at third and fourth positions in P2-3-4 tripeptides. In the case of tetrapeptides, β-turn forming residues are dominated in β-turns, but pairs of β-turn breakers strongly restrain the formation of β-turns, especially pairs having isoleucine, leucine, or valine. We also observed few tetrapeptide, which always form β-turns. These tetrapeptides have β-turn formers at their preferred positions. These tetrapeptides are devoid of hydrophobic residues, except tryptophan, which is the most frequent residue among all hydrophobic residues. We also observed that few tetrapeptides never occurred in β-turns; these tetrapeptides have two (or more) residues being hydrophobic in nature. Together, these tetrapeptides (always or never observed in β-turns) can be used to initiate or break β-turns in proteins of interest.

We introduced a turn level prediction approach for the prediction of complete β-turns and non-β-turns in proteins. Although the ability of turn level prediction is almost equal to the existing methods, it has two major advantages over residue level prediction. The prediction results are realistic with no ambiguity that is, four consecutive residues are predicted as either β-turn or nonturn. The proposed algorithm consists of a single model for prediction as compared with NetTurnP that has two steps using six different models. For the first time, BetaTPred3 was able to predict the β-turn type VIa1 and VIb, which are rare in BT426 dataset and has better/comparable performance for the rest of β-turn types. In order to improve the β-turn prediction method, we developed a new model based upon 6376 PDB chains. Finally, we have developed a web server BetaTPred3 for the prediction and designing of β-turn and its types. For the first time, we have developed a systematic module that helps biologists in understanding the positional effect of pairs of amino acids in β-turn formation in a given protein. Further, the users will be able to initiate or break a β-turn in a given protein using statistical based prediction method.

Our study will be helpful to the scientific community in better understanding of β-turn formation. In the past, experiments have been performed to introduce β-turn, especially in peptides, for better stability. The propensity scores of residues and pairs of residues at different positions of β-turns and whole β-turn propensity will be helpful in better understanding and designing of β-turn in proteins/peptides.

ACKNOWLEDGMENTS

The authors thank Prof. Michael Gromiha for critically reading the manuscript. H.S. collected and organized the data. H.S. and S.S. performed the experiments. H.S. and S.S. developed the web interface. H.S. and S.S. analyzed the data. H.S. and S.S. prepared the manuscript. G.P.S.R. conceived the idea and coordinated the project.

REFERENCES

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

2. James Milner-White E, Poet R. Loops, bulges, turns and hairpins in proteins. Trends Biochem Sci 1987;12:189–192.
3. Rose GD, Glerasch LM, Smith JA. Turns in peptides and proteins. In: Anfinsen CB, Frederic MR, editors. Advances in Protein Chemistry, Vol. 37. Academic Press; 1985. pp 1–109.
4. Petersen B, Lundegaard C, Petersen TN. NetTurnP–neural network prediction of beta-turns by use of evolutionary information and predicted protein sequence features. PLoS One 2010;5:e15079
5. Hutchinson EG, Thornton JM. A revised set of potentials for beta-turn formation in proteins. Protein Sci 1994;3:2207–2216.
6. Rubinstein N, Mayrose I, Halperin D, Yekutieli D, Gershoni J, Pupko T. Computational characterization of B-cell epitopes. Mol Immunol 2008;45:3477–3489.
7. Pellequer JL, Westhof E, Van Regenmortel MH. Predicting location of continuous epitopes in proteins from their primary structures. Methods Enzymol 1991;176:201–203.
8. Li SZ, Lee JH, Lee W, Yoon CJ, Baik JH, Lim SK. Type I beta-turn conformation is important for biological activity of the melanocyte-stimulating hormone analogues. Eur J Biochem 1999;265:430–440.
9. Ohage EC, Graml W, Walter MM, Steinbacher S, Steipe B. Beta-turn propensities as paradigms for the analysis of structural motifs to engineer protein stability. Protein Sci 1997;6:233–241.
10. Ramirez-Alvarado M, Blanco FJ, Niemann H, Serrano L. Role of beta-turn residues in beta-hairpin formation and stability in designed peptides. J Mol Biol 1997;273:898–912.
11. Shao Q, Yang L, Gao YQ. Structure change of beta-hairpin induced by turn optimization: an enhanced sampling molecular dynamics simulation study. J Chem Phys 2011;135:235104
12. Ybe JA, Hecht MH. Sequence replacements in the central beta-turn of plastocyanin. Protein Sci 1996;5:814–824.
13. Kaur H, Garg A, Raghava GP. PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides. Protein Pept Lett 2007;14:626–631.
14. Chou PY, Fasman GD. Prediction of beta-turns. Biophys J 1979;26: 367–383.
15. Chou PY, Fasman GD. Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. Biochemistry 1974;13:211–222.
16. Zhang C-T, Chou K-C. Prediction of β-turns in proteins by 1-4 and 2-3 correlation model. Biopolymers 1997;41:673–702.
17. Chou KC. Prediction of beta-turns. J Pept Res 1997;49:120–144.
18. Kaur H, Raghava GP. BetaTPred: prediction of beta-TURNS in a protein using statistical algorithms. Bioinformatics 2002;18:498–499.
19. McGregor MJ, Flores TP, Sternberg MJ. Prediction of beta-turns in proteins using neural networks. Protein Eng 1989;2:521–526.
20. Shepherd AJ, Gorse D, Thornton JM. Prediction of the location and type of beta-turns in proteins using neural networks. Protein Sci 1999;8:1045–1055.
21. Kim S. Protein beta-turn prediction using nearest-neighbor method. Bioinformatics 2004;20:40–44.
22. Fuchs PF, Alix AJ. High accuracy prediction of beta-turns and their types using propensities and multiple alignments. Proteins 2005;59: 828–839.
23. Kaur H, Raghava GP. A neural network method for prediction of beta-turn types in proteins using evolutionary information. Bioinformatics 2004;20:2751–2758.
24. Guruprasad K, Shukla S. Prediction of beta-turns from amino acid sequences using the residue-coupled model. J Pept Res 2003;61:159–162.
25. Kirschner A, Frishman D. Prediction of beta-turns and beta-turn types by a novel bidirectional Elman-type recurrent neural network with multiple output layers (MOLEBRNN). Gene 2008;422:22–29.
26. Hu X, Li Q. Using support vector machine to predict beta- and gamma-turns in proteins. J Comput Chem 2008;29:1867–1875.
27. Zheng C, Kurgan L. Prediction of beta-turns at over 80% accuracy based on an ensemble of predicted secondary structures and multiple alignments. BMC Bioinformatics 2008;9:430
28. Kountouris P, Hirst JD. Predicting beta-turns and their types using predicted backbone dihedral angles and secondary structures. BMC Bioinformatics 2010;11:407
29. Singh H, Chauhan JS, Gromiha MM, Raghava GP. ccPDB: compilation and creation of data sets from Protein Data Bank. Nucleic Acids Res 2012;40(Database issue):D486-D489.
30. Guruprasad K, Rajkumar S. Beta-and gamma-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials. J Biosci 2000;25:143–156.
31. Kaur H, Raghava GP. An evaluation of beta-turn prediction methods. Bioinformatics 2002;18:1508–1514.
32. Kaur H, Raghava GP. Prediction of beta-turns in proteins from multiple alignment using neural network. Protein Sci 2003;12:627–634.
33. T.H. Pham KSaTBH. Prediction and analysis of β-turns in proteins by support vector machine. In: 14th International Conference on Genome Informatics (GIW 2003). Yokohama; 2003. pp 196–205. Available at: http://giw.hgc.jp/giw2003/accepted_paper.htm.
34. Liu L, Fang Y, Li M, Wang C. Prediction of beta-turn in protein using E-SSpred and support vector machine. Protein J 2009;28:175–181.
35. Hutchinson EG, Thornton JM. PROMOTIF–a program to identify and analyze structural motifs in proteins. Protein Sci 1996;5:212–220.
36. Agarwal S, Mishra NK, Singh H, Raghava GP. Identification of mannose interacting residues using local composition. PLoS One 2011;6:e24039
37. Soding J. Protein homology detection by HMM-HMM comparison. Bioinformatics 2005;21:951–960.
38. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
39. Remmert M, Biegert A, Hauser A, Soding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods 2012;9:173–175.
40. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000;16:404–405.
41. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
42. Breiman L. Random forests. Mach Learn 2001;45:5–32.
43. Kibler DAaD. Instance-based learning algorithms. Mach Learn 1991; 6:37–66.
44. le Cessie SavH JC. Ridge estimators in logistic regression. Appl Statist 1992;41:191–201.
45. Quinlan R. C4.5: Programs for machine learning. San Mateo, CA: Morgan Kaufmann Publishers; 1993.
46. Mark Hall EF, Geoffrey H, Bernhard P, Peter R, Ian HW. The WEKA data mining software: an update. SIGKDD Explor 2009;11. Available at: http://www.cs.waikato.ac.nz/ml/weka/citing.html.
47. Langley GHJAP. Estimating continuous distributions in bayesian classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence. San Mateo, Morgan Kaufmann Publishers; 1995. pp 338–345.
48. Kaur H, Raghava GP. A neural-network based method for prediction of gamma-turns in proteins from multiple sequence alignment. Protein Sci 2003;12:923–929.
49. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics 2005;21:3940–3941.
50. Team RDC R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2012. Available at: http://cran.r-project.org/doc/FAQ/R-FAQ.html#Citing-R.