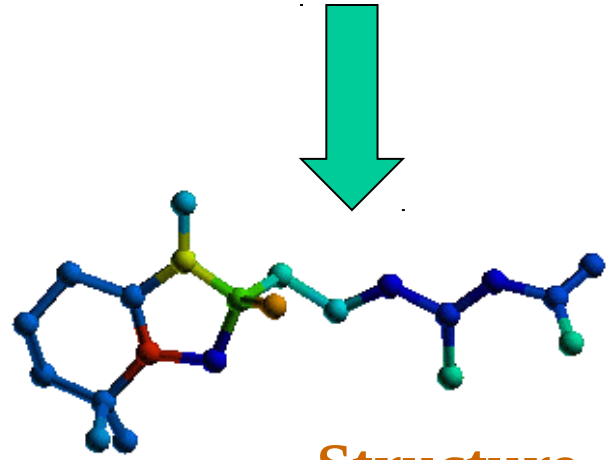
Prediction of Tight Turns In Protein

Protein Sequence +

G.P.S. Raghava, Ph.D., F.N.A.Sc.
Scientist and Head Bioinformatics Centre
Institute of Microbial Technology, Sector-39 A, Chandigarh, India

Email:raghava@imtech.res.in
Web: http://www.imtech.res.in/raghava/



Structure

Protein Structure Prediction

- Experimental Techniques
  - X-ray Crystallography
  - NMR
- Limitations of Current Experimental Techniques
  - Protein DataBank (PDB) -> 27000 protein structures
  - SwissProt -> 100,000 proteins
  - Non-Redudant (NR) -> 1,000,000 proteins
- Importance of Structure Prediction
  - Fill gap between known sequence and structures
  - Protein Engg. To alter function of a protein
  - Rational Drug Design

Techniques of Structure Prediction

- Computer simulation based on energy calculation
  - Based on physio-chemical principles
  - Thermodynamic equilibrium with a minimum free energy
  - Global minimum free energy of protein surface
- Knowledge Based approaches
  - Homology Based Approach
  - Threading Protein Sequence
- Hierarchical Methods
  - Prediction of intermediate state (Secondary Structure)
  - Secondary to tertiary structure

Energy Minimization Techniques

Energy Minimization based methods in their pure form, make no priori assumptions and attempt to locate global minma.

- **Static Minimization Methods**
  - Classical many potential-potential can be construted
  - Assume that atoms in protein is in static form
  - Problems(large number of variables & minima and validity of potentials)
- **Dynamical Minimization Methods**
  - Motions of atoms also considered
  - Monte Carlo simulation (stochastics in nature, time is not cosider)
  - Molecular Dynamics (time, quantum mechanical, classical equ.)
- Limitations
  - large number of degree of freedom,CPU power not adequate
  - Interaction potential is not good enough to model

Knowledge Based Approaches

- Homology Modelling
  - Need homologues of known protein structure
  - Backbone modelling
  - Side chain modelling
  - Fail in absence of homology
- Threading Based Methods
  - New way of fold recognition
  - Sequence is tried to fit in known structures
  - Motif recognition
  - Loop & Side chain modelling
  - Fail in absence of known example

Hierarcial Methods

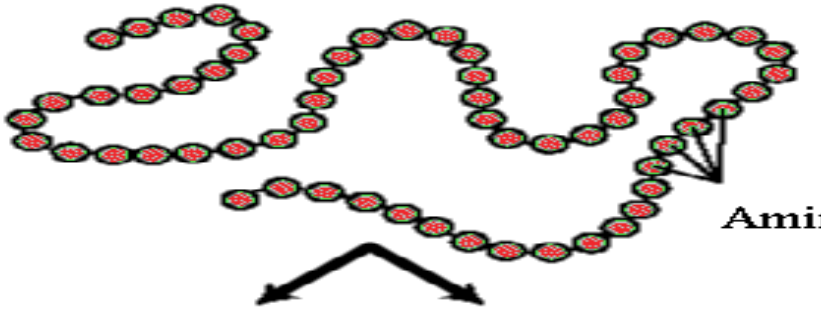
Intermidiate structures are predicted, instead of predicting tertiary structure of protein from amino acids sequence

- Prediction of backbone structure
  - Secondary structure (helix, sheet,coil)
  - Beta Turn Prediction
  - Super-secondary structure
- Tertiary structure prediction
- Limitation

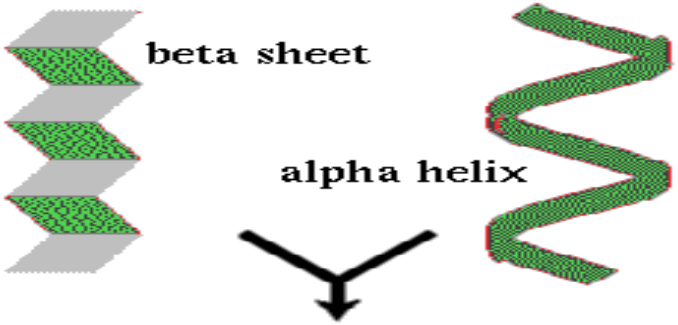
  Accuracy is only 75-80 %

  Only three state prediction

Different Levels of Protein Structure



Primary protein structure
is sequence of amino acids



Secondary protein structure
local conformation
primarily stabilized by
hydrogen bonding

Tertiary protein structure
three dimensional conformation

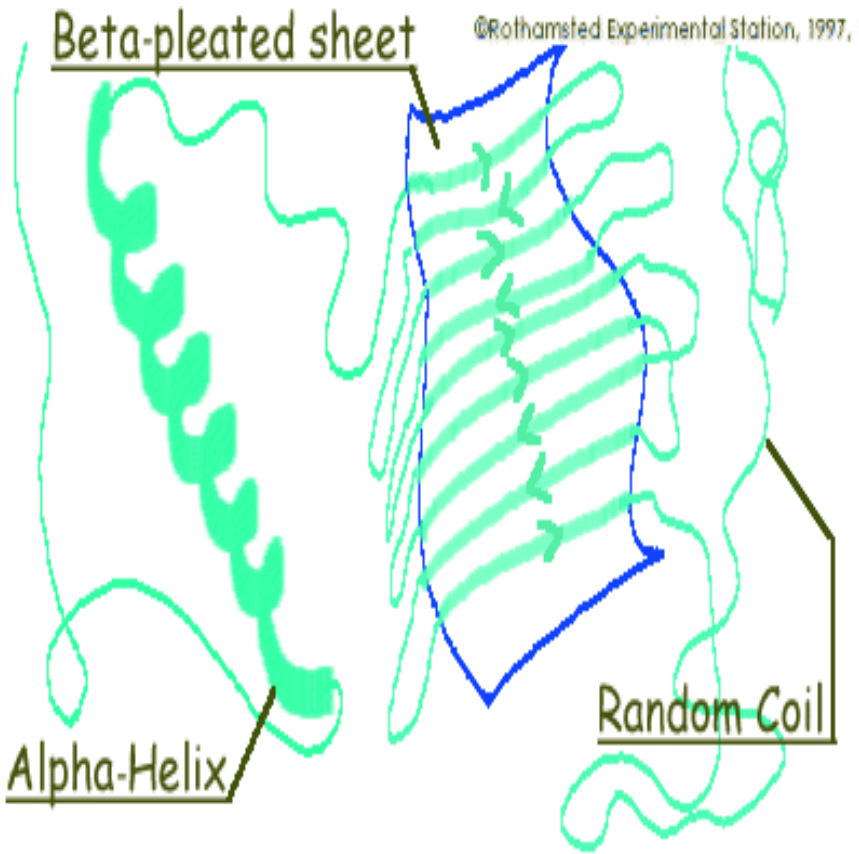
Quaternary protein structure

Combination of multiple
polypeptide chains

Levels of Description of Structural Complexity

- Primary Structure (AA sequence)
- Secondary Structure
  - Spatial arrangement of a polypeptide's backbone atoms without regard to side-chain conformations
    - $\forall$ α, β, coil, turns (Venkatachalam, 1968)
  - Super-Secondary Structure
    - $\forall$ α, β, α/β, $\alpha+\beta$ (Rao and Rassman, 1973)
- Tertiary Structure
  - 3-D structure of an entire polypeptide
- Quarternary Structure
  - Spatial arrangement of subunits (2 or more polypeptide chains)

Protein Secondary Structure



The secondary structure is observed in a localised portion of a protein.

Secondary Structure

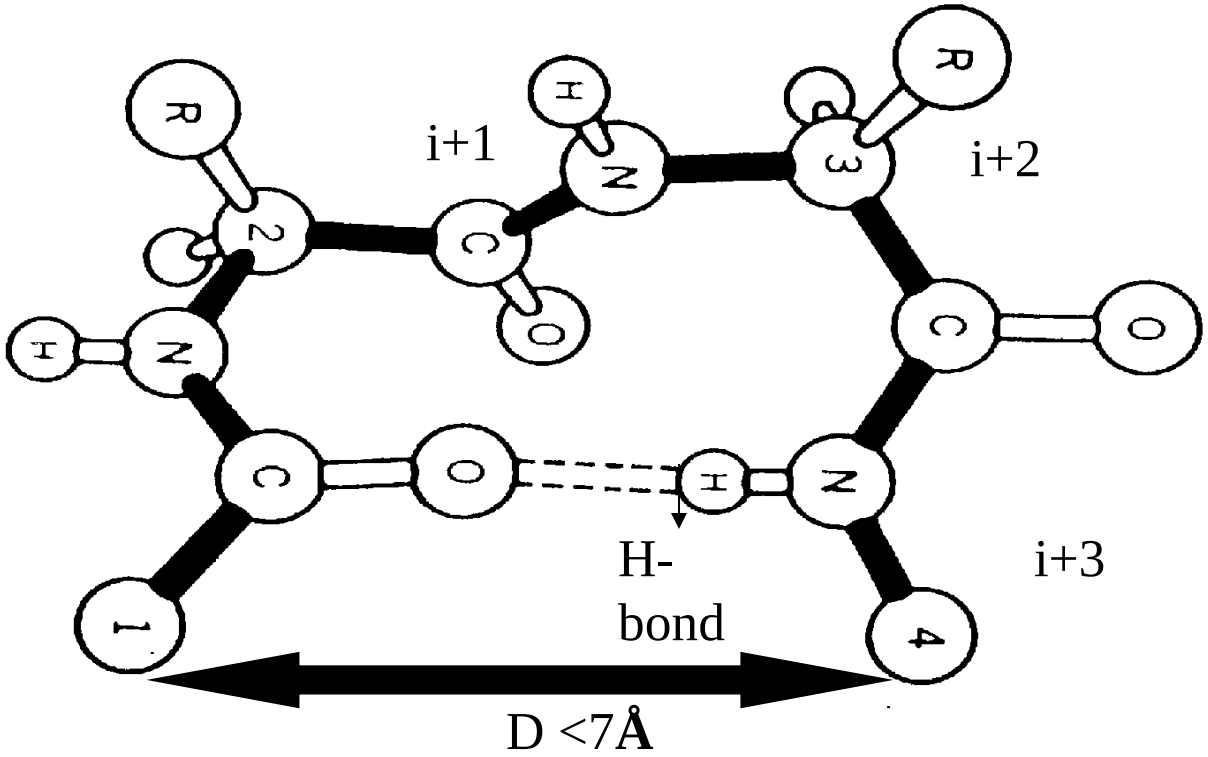
Regular Secondary Structure (α-helices, β-sheets)

Irregular Secondary Structure (Tight turns, Random coils, bulges)

Definition of β-turn

A β-turn is defined by four consecutive residues *i*, *i*+1, *i*+2 and *i*+3 that do not form a helix and have a Cα(*i*)-Cα(*i*+3) distance less than 7Å and the turn lead to reversal in the protein chain. (Richardson, 1981).

The conformation of β-turn is defined in terms of ϕ and ψ of two central residues, *i*+1 and *i*+2 and can be classified into different types on the basis of ϕ and ψ.



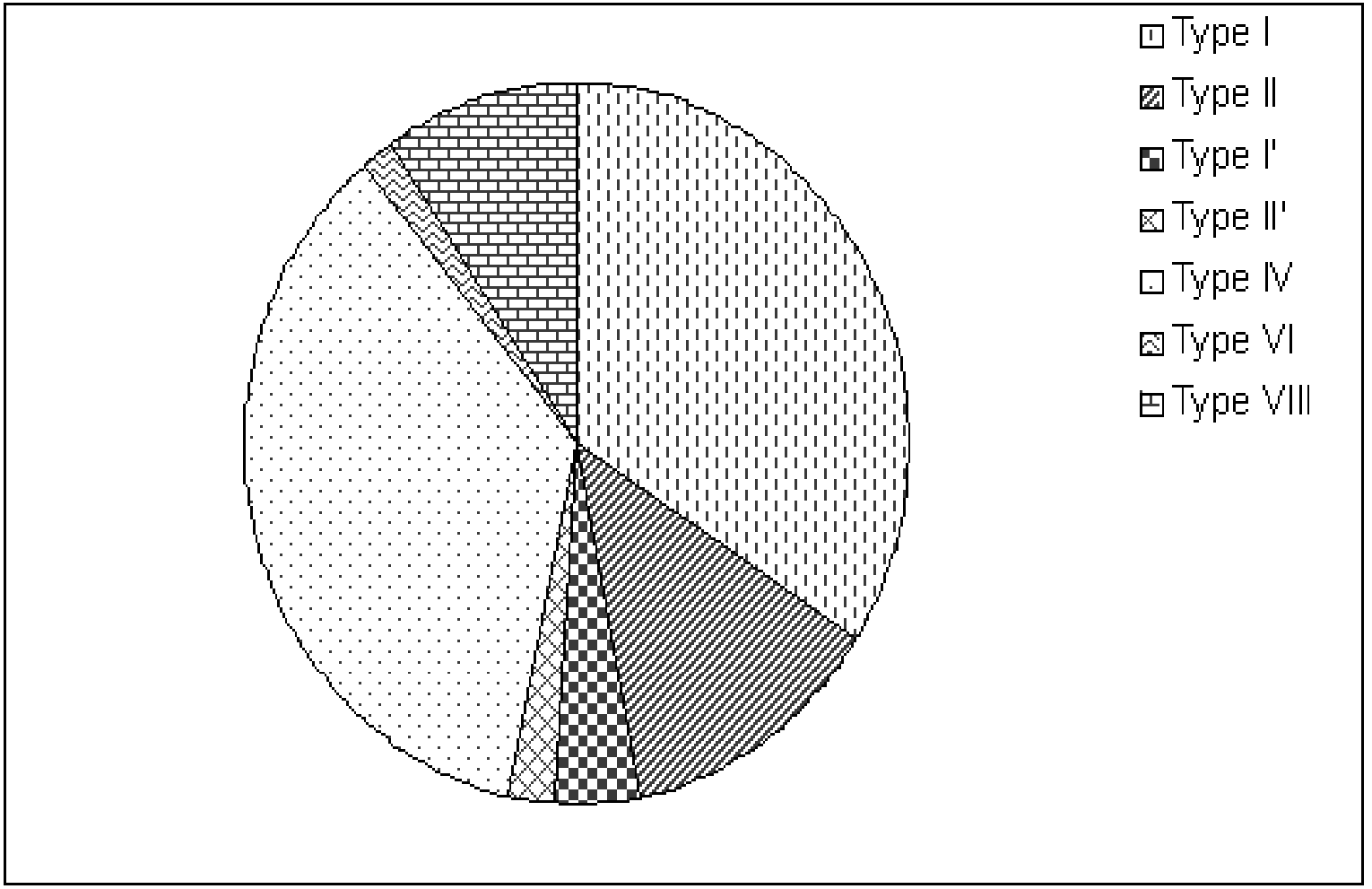
Tight turns

Type	No. of residues	H-bonding
δ-turn	2	NH(i)-CO(i+1)
γ-turn	3	CO(i)-NH(i+2)
β-turn	**4**	**CO(i)-NH(i+3)**
α-turn	5	CO(i)-NH(i+4)
π-turn	6	CO(i)-NH(i+5)

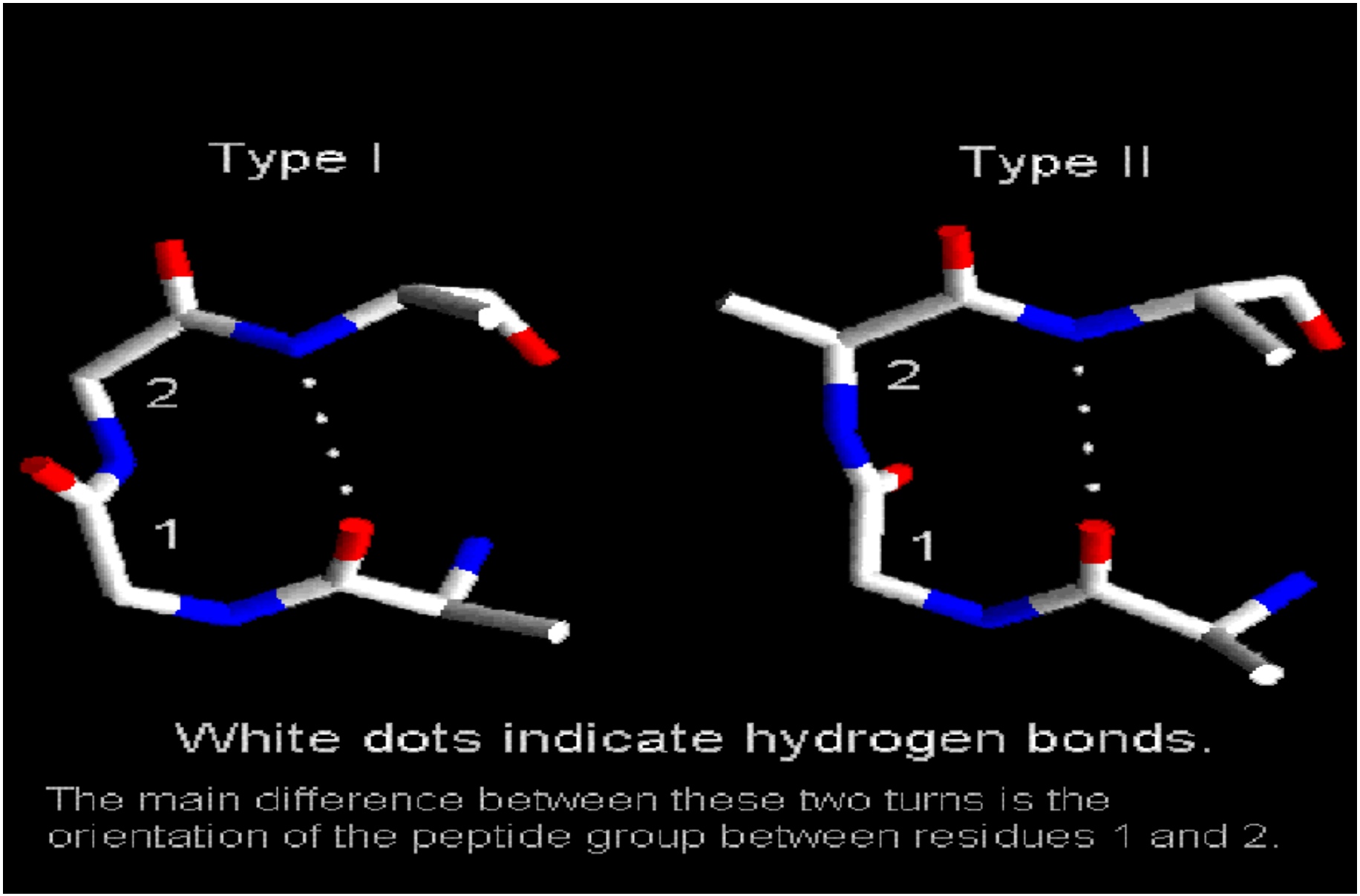
Beta-turn types

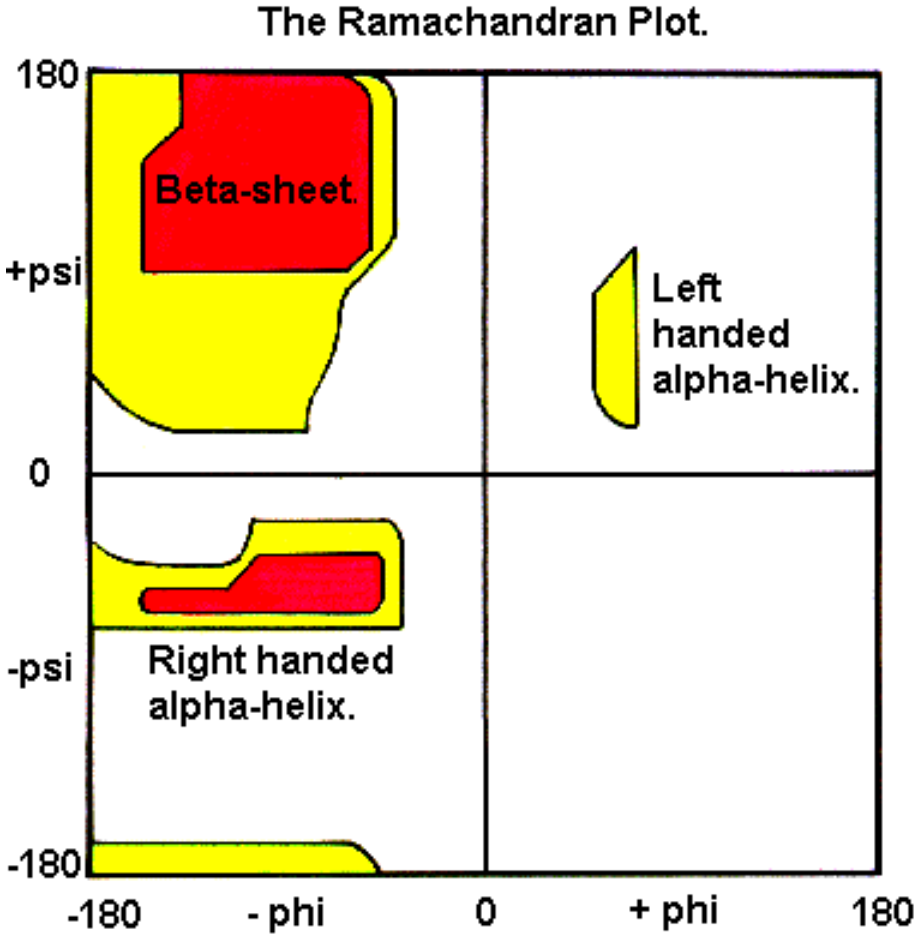
Turn Type	Dihedral angles (°)			
	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
I	-60	-30	-90	0
I'	60	30	90	0
II	-60	120	80	0
II'	60	-120	-80	0
IV	-61	10	-53	17
VIa1	-60	120	-90	0
VIa2	-120	120	-60	0
VIb	-135	135	-75	160
VIII	-60	-30	-120	120

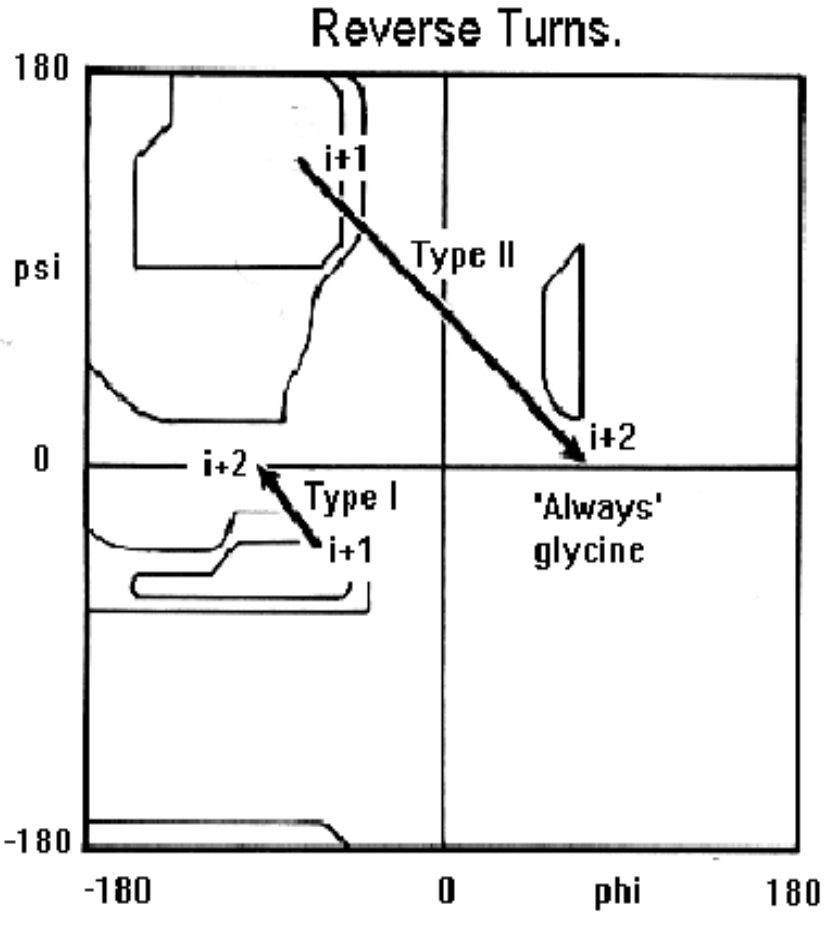
Distribution of β-turn types



Two main types of β-turns







a b

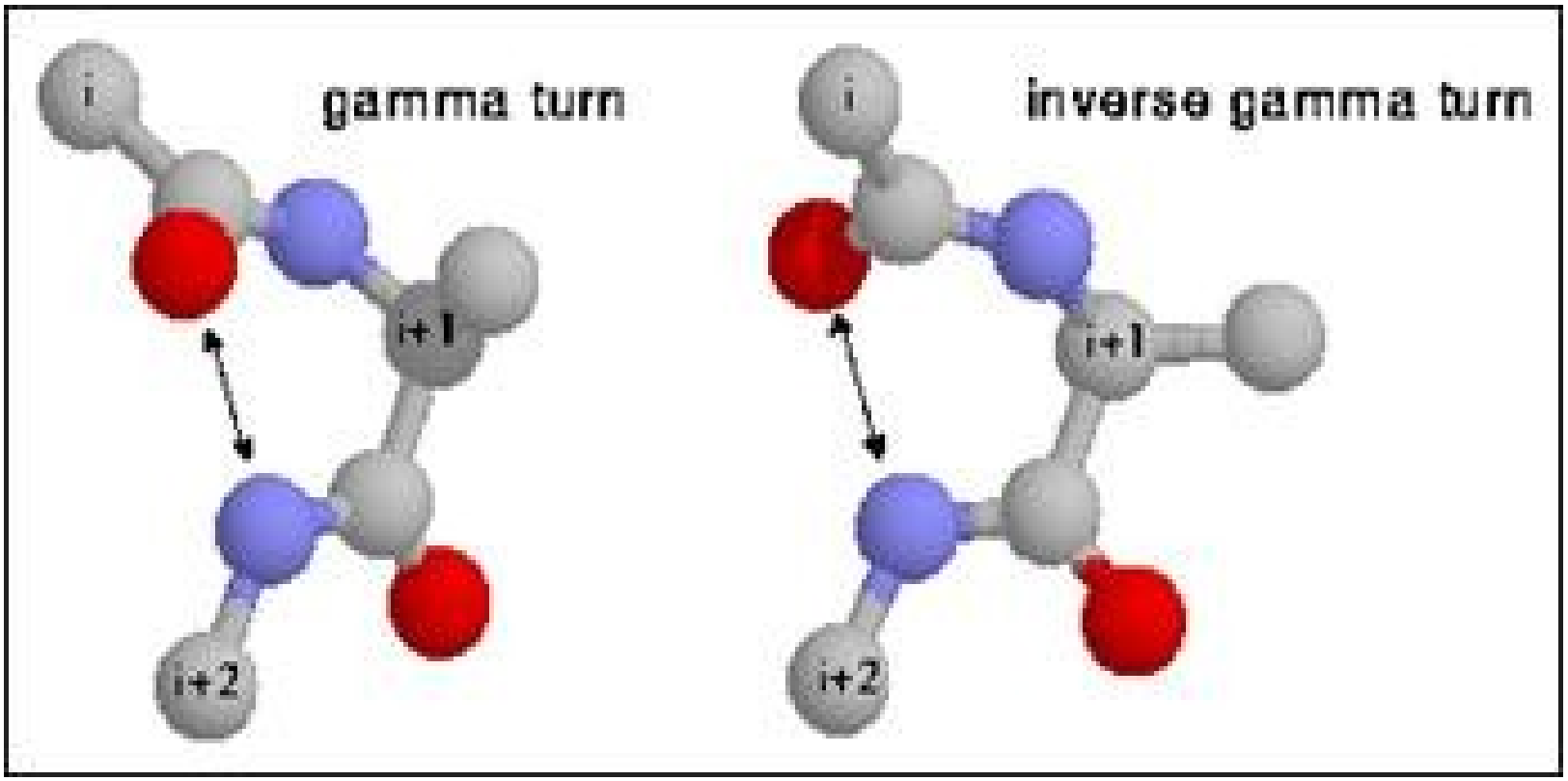
a: Ramachandran plot showing the characteristic region where β-sheet and α-helices are found.

b: Ramachandran plot showing Type I and II turns represented by a vector

Gamma turns

•The γ-turn is the second most characterized and commonly found turn, after the β-turn.

•A γ-turn is defined as 3-residue turn with a hydrogen bond between the Carbonyl oxygen of residue *i* and the hydrogen of the amide group of residue *i*+2. There are 2 types of γ-turns: classic and inverse.



Other rare tight turns

- **δ-turn:** The smallest is a δ-turn. It involves only two amino acid residues. The intra-turn hydrogen bond for a δ-turn is formed between the backbone NH(*i*) and the backbone CO(*i*+1).

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

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

Prediction of tight turns

- Prediction of β-turns
- Prediction of β-turn types
- Prediction of γ-turns
- Prediction of α-turns
- Use the tight turns information, mainly β-turns in tertiary structure prediction of bioactive peptides

Existing β-turn prediction methods

- **Residue Hydrophobicities (*Rose, 1978*)**

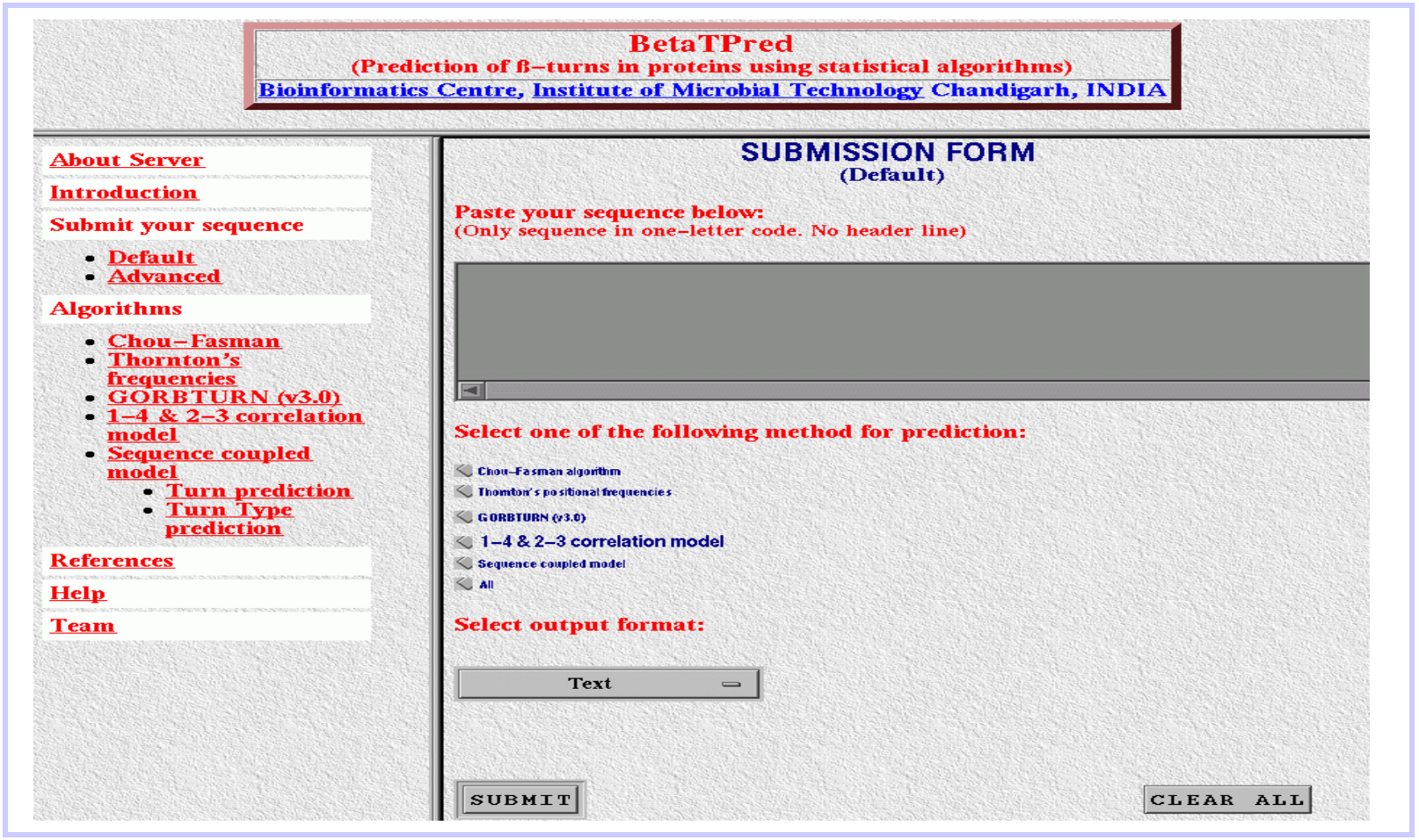
- **Positional Preference Approach**
  - **Chou and Fasman Algorithm (*Chou and Fasman, 1974; 1979*)**
  - **Thornton's Algorithm (*Wilmot and Thornton, 1988*)**
  - **GORBTURN (*Wilmot and Thornton, 1990*)**
  - **1-4 & 2-3 Correlation Model (*Zhang and Chou, 1997*)**
  - **Sequence Coupled Model (*Chou, 1997*)**

- **Artificial Neural Network**
  - **BTPRED (*Shepherd et al., 1999*)**

    (http://www.biochem.ucl.ac.uk/bsm/btpred/)

BetaTPred: Prediction of β-turns using statistical methods

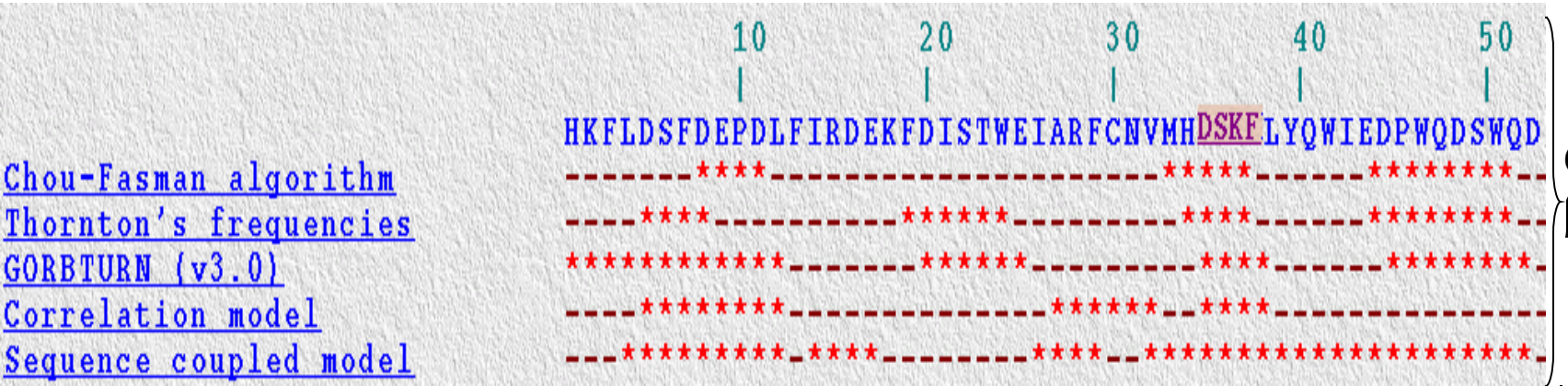
(http://imtech.res.in/raghava/betatpred/)



Harpreet Kaur and G P S Raghava (2002) BetaTPred: Prediction of β-turns in a protein using statistical algorithms. *Bioinformatics* 18(3), 498-499.

Text Output

Graphical (Frames) output



Consensus β-turn

We have evaluated the performance of six methods of β-turn prediction. All the methods have been tested on a set of 426 non-homologous protein chains. In this study, both threshold dependent (Q*total*, Q*pred*., Q*obs*. And MCC) and independent (ROC) measures have been used for evaluation.

Harpreet Kaur and G.P.S Raghava (2002) An evaluation of β-turn prediction methods. ***Bioinformatics*** 18(11), 1508-1514.

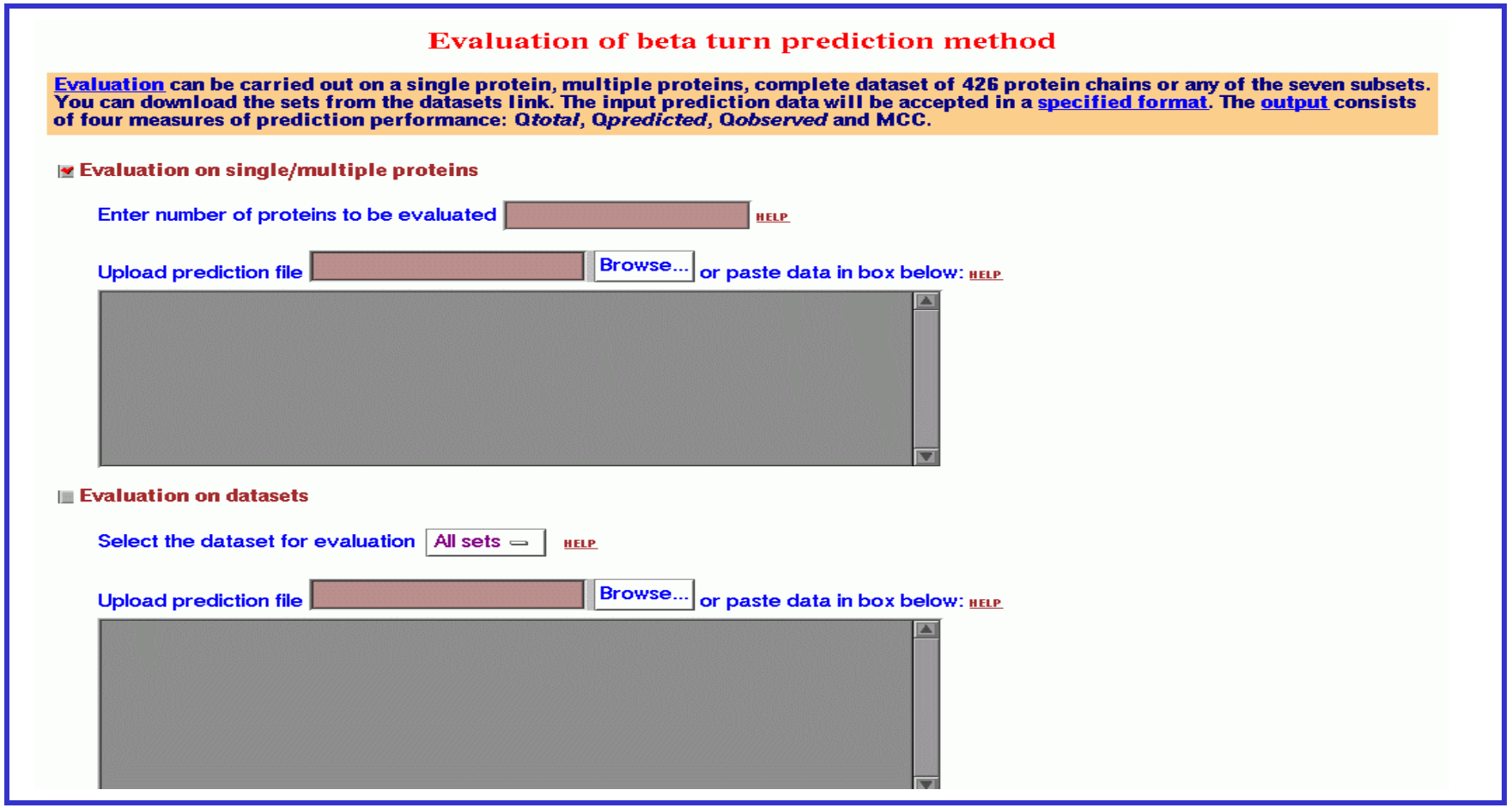
Performance of existing β-turn methods

	Q*total*	Q*pred*.	Q*obs*.	MCC
BTPRED	75.3	49.7	63.4	0.39
Chou-Fasman	75.3	49.6	47.5	0.32
Thornton's algorithm	75.2	49.3	44.9	0.31
1-4 & 2-3 Correlation model	74.8	48.0	39.8	0.28
Sequence Coupled Model	75.4	49.6	40.0	0.28
GORBTURN	75.4	49.6	37.7	0.28

BTEVAL: A web server for evaluation of β-turn prediction methods

(http://imtech.res.in/raghava/bteval/)

Harpreet Kaur and G P S Raghava (2003) BTEVAL: A server for evaluation of β-turn prediction methods. ***Journal of Bioinformatics and Computational Biology*** (in press).



BTEVAL: A web server for evaluation of β-turn prediction methods

INPUT DATA

PDB code: 1ah7

Filename: 1ah7.txt

Amino Acid Sequence:

W S A E D K H K E G V N S H L W I V N R A I D I M S R N T T L V K Q D R V A Q L N E W R T E L E N G
I Y A A D Y E N P Y Y D N S T F A S H F Y D P D N G K T Y I P F A K Q A K E T G A K Y F K L A G E S Y
K N K D M K Q A F F Y L G L S L H Y L G D V N Q P M H A A N F T N L S Y P Q G F H S K Y E N F V D T
I K D N Y K V T D G N G Y W N W K G T N P E E W I H G A A V V A K Q D Y S G I V N D N T K D W F V
K A A V S Q E Y A D K W R A E V T P M T G K R L M D A Q R V T A G Y I Q L W F D T Y G D R

Predicted turns/nonturns:

n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n
n n n n n n n n n n n n t t t t t n n n t t t t n n n n n n n n n n n n n n n n n n n t t t t n n n n n n n n n n n n n n n
n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n t t t t t t n n n n n n n n n n n n n
n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n
n n n n

Data entered by the user

EVALUATION RESULTS

Prediction Method	*Qtotal* (in %)	*Qpredicted* (in %)	*Qobserved* (in %)	MCC
Query method	85.7	78.9	32.6	0.45
Chou-Fasman algorithm (original parameters)	56.5	26.7	76.1	0.22
Thornton's algorithm (original parameters)	62.2	30.9	82.6	0.31
GORBTURN	77.6	40.8	43.5	0.28
1-4 & 2-3 Correlation Model (original parameters)	53.7	24.2	69.6	0.15
Sequence Coupled Model (original parameters)	51.8	21.4	58.7	0.07
BTPRED	85.3	60.9	60.9	0.52

Performance measures of query method plus other existing beta turn prediction methods

BetaTPred2: Prediction of β-turns in proteins from multiple alignment using neural network

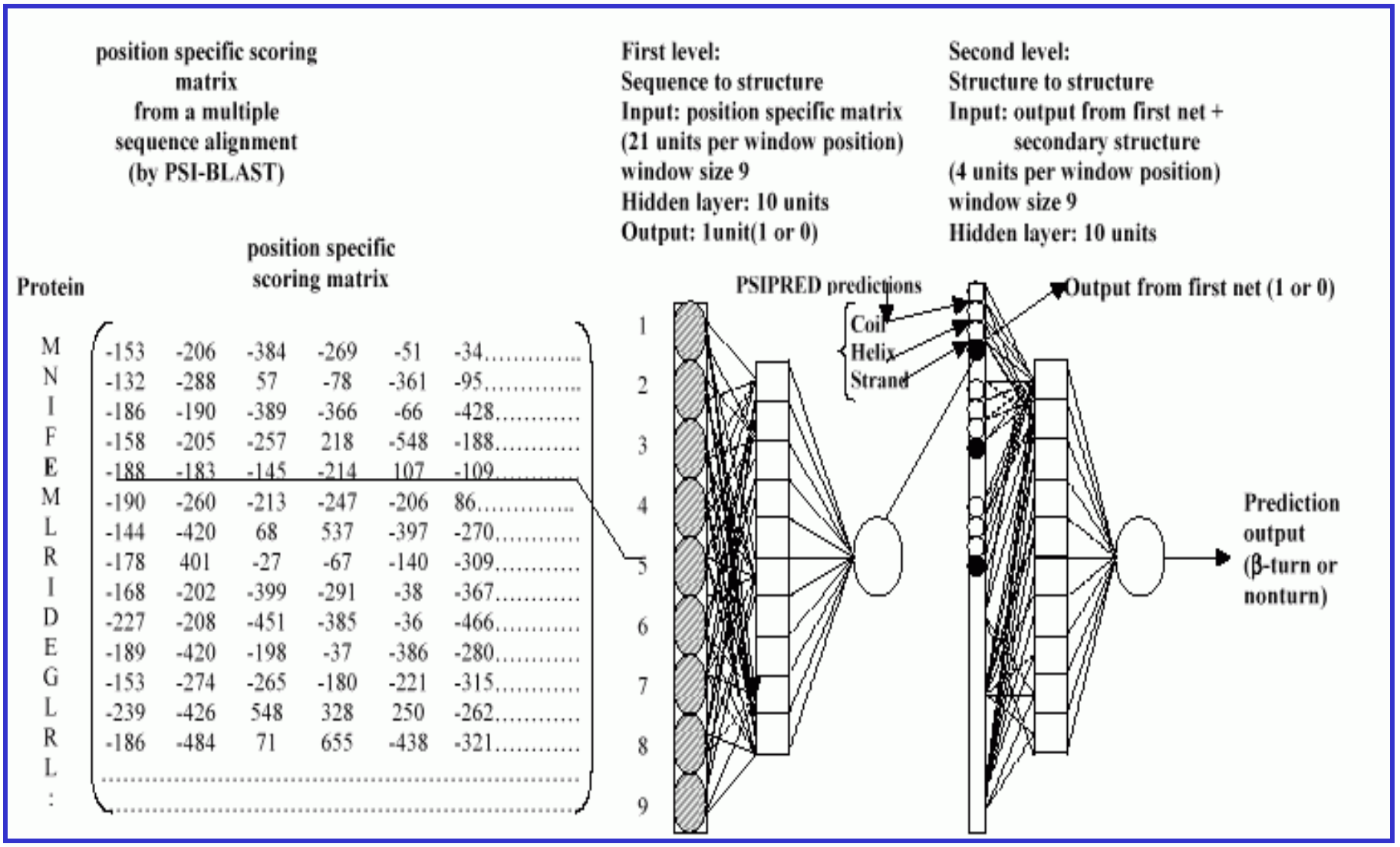
Harpreet Kaur and G P S Raghava (2003) Prediction of β-turns in proteins from multiple alignment using neural network. *Protein Science* 12, 627-634.

- **Two feed-forward back-propagation networks with a single hidden layer are used where the first sequence-structure network is trained with the multiple sequence alignment in the form of PSI-BLAST generated position specific scoring matrices.**

- **The initial predictions from the first network and PSIPRED predicted secondary structure are used as input to the second sequence-structure network to refine the predictions obtained from the first net.**

- **The final network yields an overall prediction accuracy of 75.5% when tested by seven-fold cross-validation on a set of 426 non-homologous protein chains. The corresponding Q*pred*., Q*obs*. and MCC values are 49.8%, 72.3% and 0.43 respectively and are the best among all the previously published β-turn prediction methods. A web server BetaTPred2 (http://www.imtech.res.in/raghava/betatpred2/) has been developed based on this approach.**

Neural Network architecture used in BetaTPred2



BetaTPred2 prediction results using single sequence and multiple alignment.

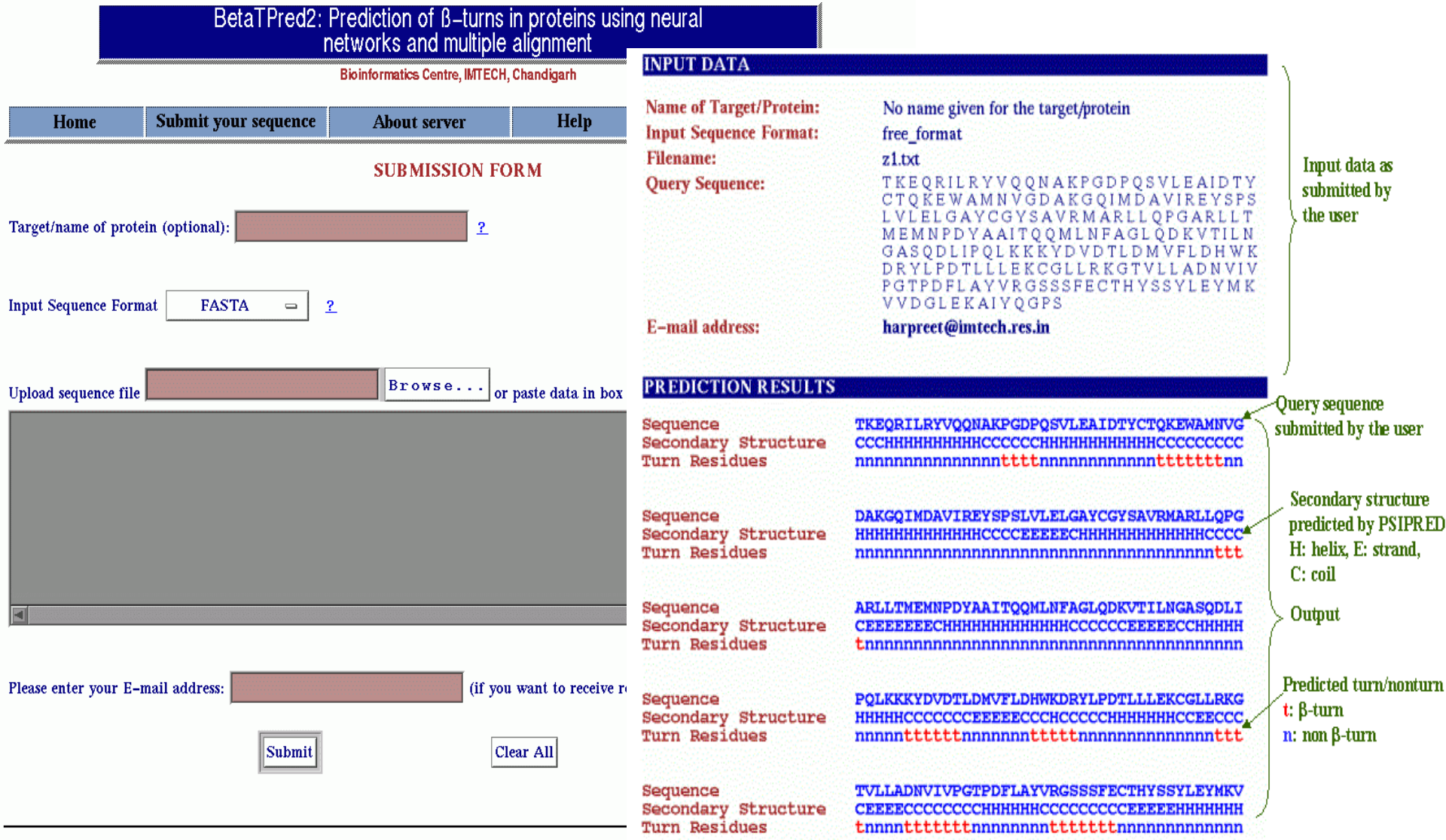
	Single sequence		Multiple alignment	
	First network	Second network	First network	Second network
Q_{total}	71.6 ± 0.7	74.3 ± 1.0 (74.6 ± 1.4)	73.5 ± 1.5	75.5 ± 1.7 (75.8 ± 1.6)
Q_{pred}	44.1 ± 1.3	48.4 ± 1.7 (48.6 ± 1.8)	47.2 ± 1.9	49.8 ± 2.0 (49.9 ± 1.9)
Q_{obs}	57.3 ± 2.6	71.2 ± 2.0 (70.4 ± 2.0)	64.3 ± 2.2	72.3 ± 2.6 (70.8 ± 2.0)
MCC	0.31 ± 0.01	0.41 ± 0.01 (0.41 ± 0.01)	0.37 ± 0.01	0.43 ± 0.01 (0.43 ± 0.01)

Values in parentheses correspond to the prediction results obtained by excluding the proteins that were used to develop PSIPRED method.

Harpreet Kaur and G P S Raghava (2003) Prediction of β-turns in proteins from multiple alignment using neural network. ***Protein Science*** 12, 627-634.

BetaTPred2: A web server for prediction of β-turns in proteins

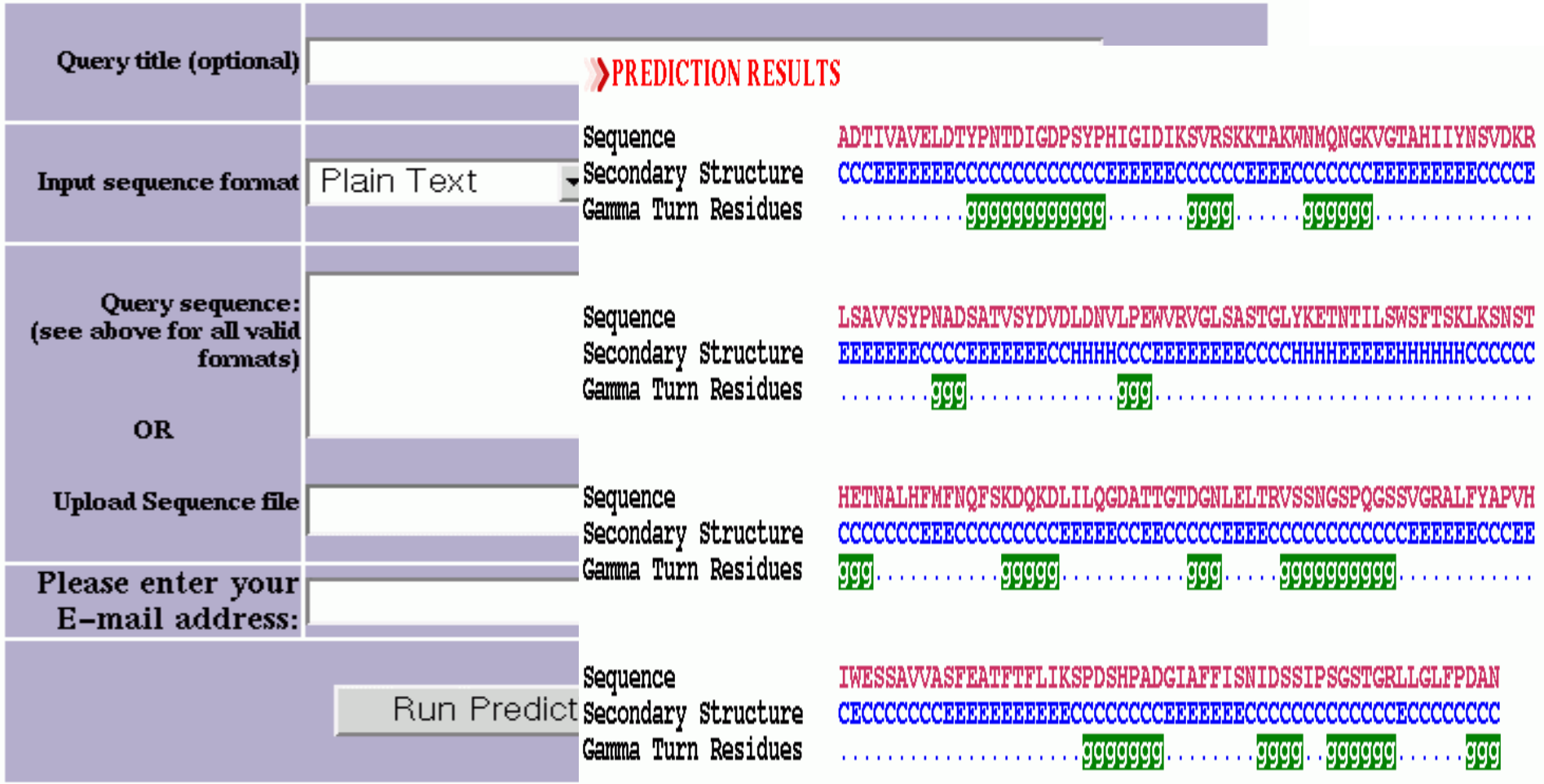
(http://www.imtech.res.in/raghava/betatpred2/)



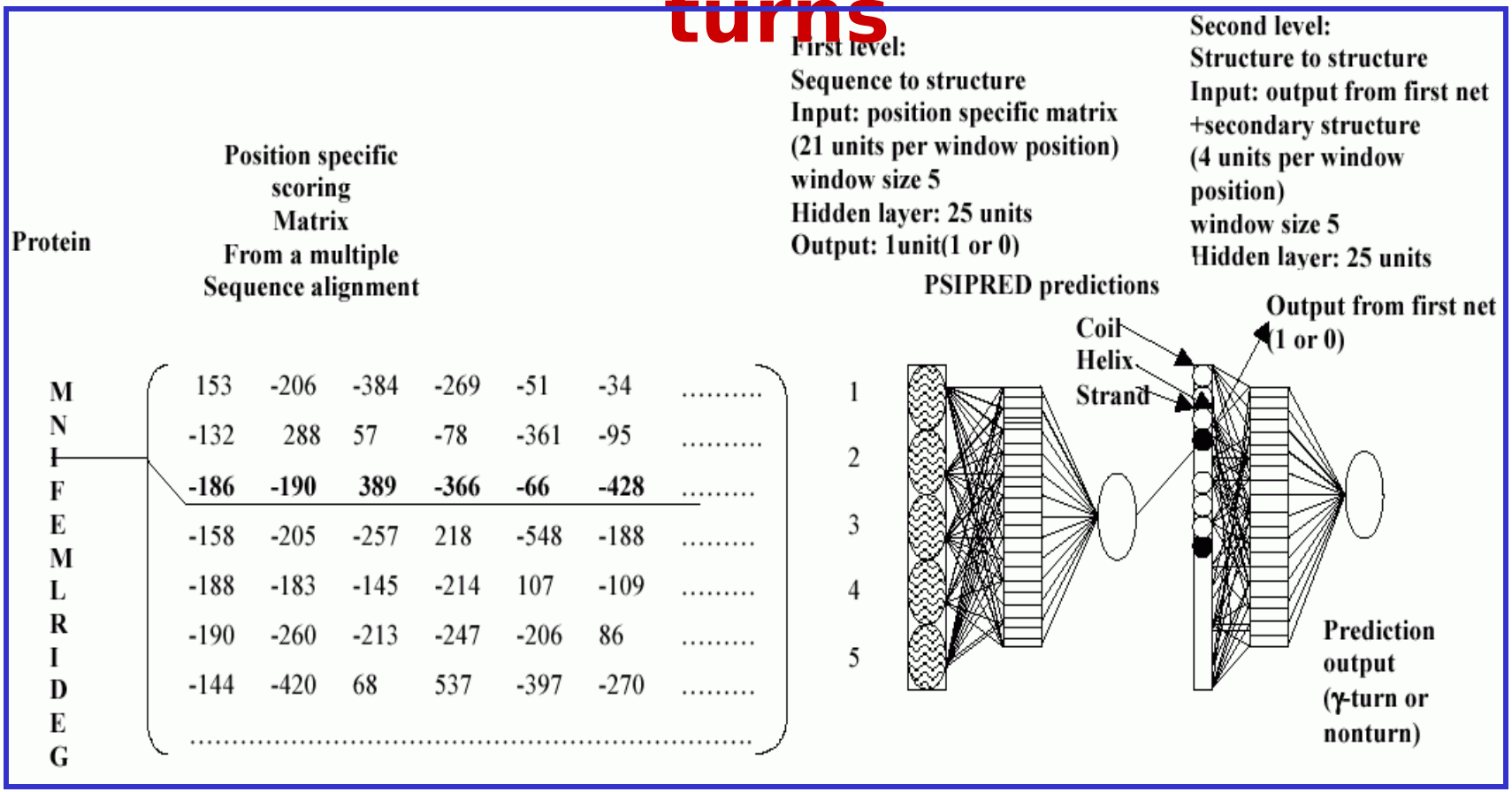
Gammapred: A server for prediction of γ-turns in proteins

(http://www.imtech.res.in/raghava/gammapred/)

Harpreet Kaur and G P S Raghava (2003) A neural network based method for prediction of γ-turns in proteins from multiple sequence alignment. ***Protein Science*** 12, 923-929.



Network architecture for gamma turns

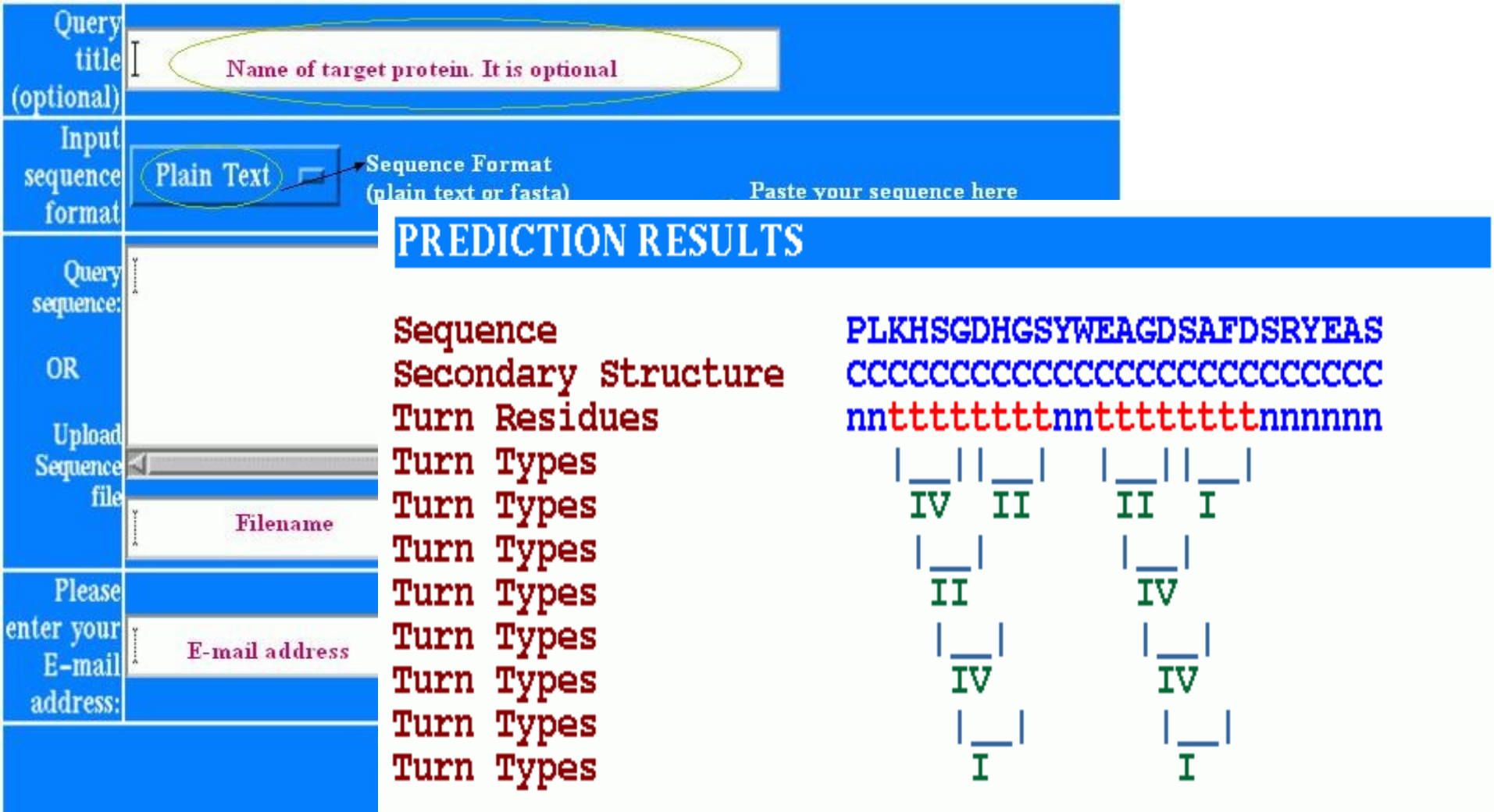


Harpreet Kaur and G P S Raghava (2003) A neural network based method for prediction of γ-turns in proteins from multiple sequence alignment. ***Protein Science*** 12, 923-929.

BetaTurns: A web server for prediction of β-turn types

(http://www.imtech.res.in/raghava/betaturns/)

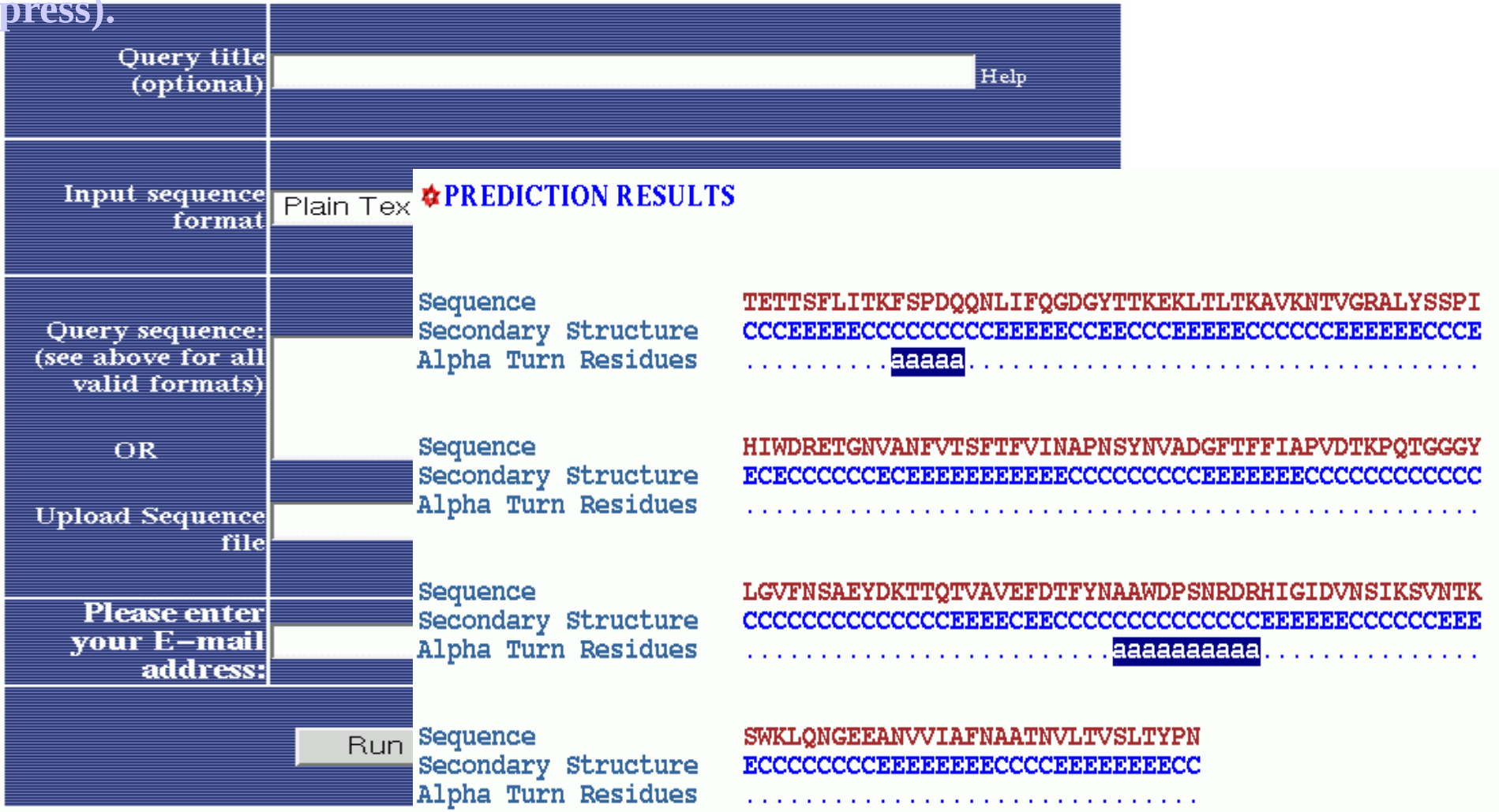
Harpreet Kaur and G P S Raghava (2003) Prediction of β-turn types in proteins from evolutionary information using neural network. Bioinformatics (In Press)



AlphaPred: A web server for prediction of α-turns in proteins

(http://www.imtech.res.in/raghava/alphapred/)

Harpreet Kaur and G P S Raghava (2003) Prediction of α-turns in proteins using PSI-BLAST profiles and secondary structure information. ***Proteins*** (in press).

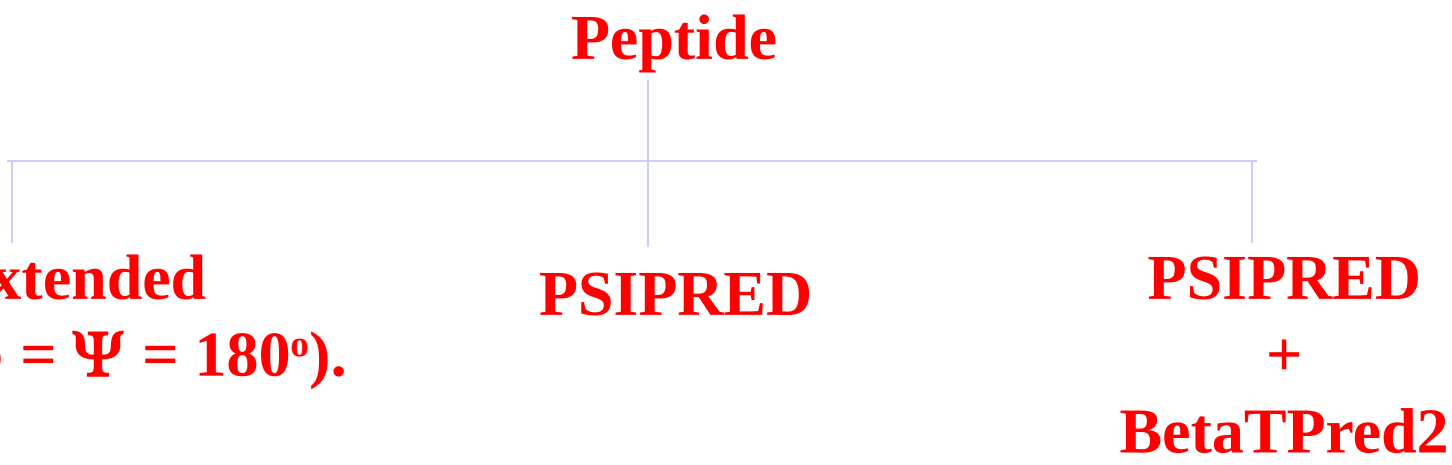


Contribution of β-turns in tertiary structure prediction of bioactive peptides

- 3D structures of 77 biologically active peptides have been selected from PDB and other databases such as PSST (http://pranag.physics.iisc.ernet.in/psst) and PRF (http://www.genome.ad.jp/) have been selected.
- The data set has been restricted to those biologically active peptides that consist of only natural amino acids and are linear with length varying between 9-20 residues.

Secondary structure state	No. of peptides	% of total peptide residues
Helices	46	32.3
β-sheets	10	6.9
β-turns	58	34.9

3 models have been studied for each peptide. The first model has been ($\phi = \Psi = 180^{o}$). The second model is build up by constructed by taking all the peptide residues in the extended conformation assigning the peptide residues the ϕ, Ψ angles of the secondary structure states predicted by PSIPRED. The third model has been constructed with ϕ, Ψ angles corresponding to the secondary states predicted by PSIPRED and β-turns predicted by BetaTPred2.



Root Mean Square Deviation has been calculated…….

Averaged backbone root mean deviation before and after energy minimization and dynamics simulations.

Model	Averaged backbone root mean deviation (Å)	
	before EM & DS[a]	*after EM & DS*
I	10.8	5.9
II	7.6	4.9
III	5.6	4.2

[a] EM and DS denote energy minimization and dynamics simulations respectively.

Thanks