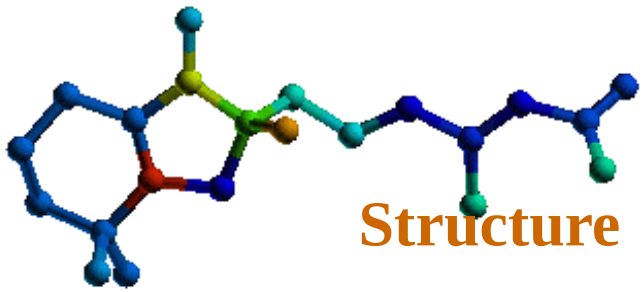
Computer-Aided Protein Structure Prediction

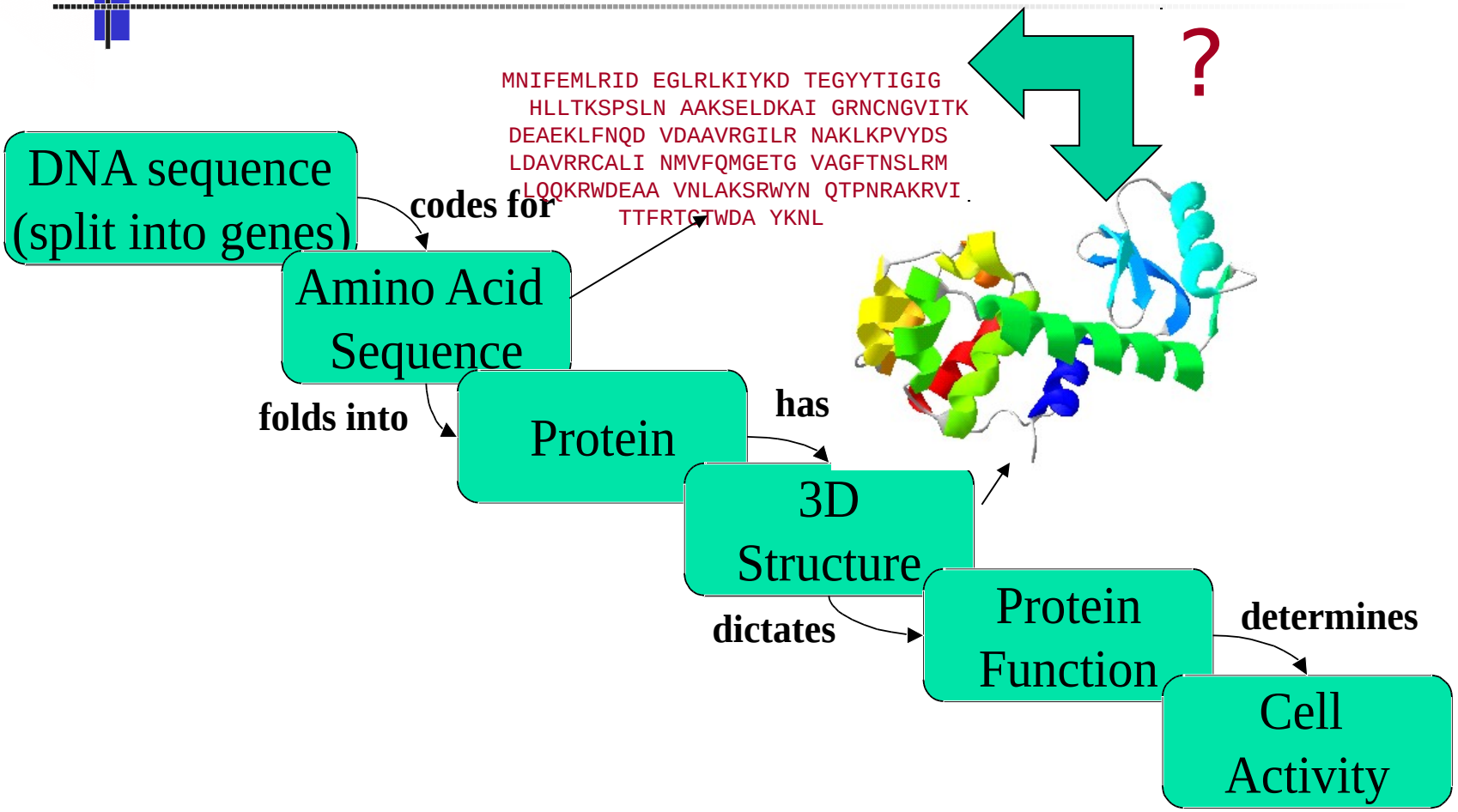
Protein Sequence +

Dr. G.P.S. Raghava, F.N.A. Sc.
Bioinformatics Centre
Institute of Microbial Technology
Chandigarh, INDIA

E-mail:
raghava@imtech.res.in
Web:

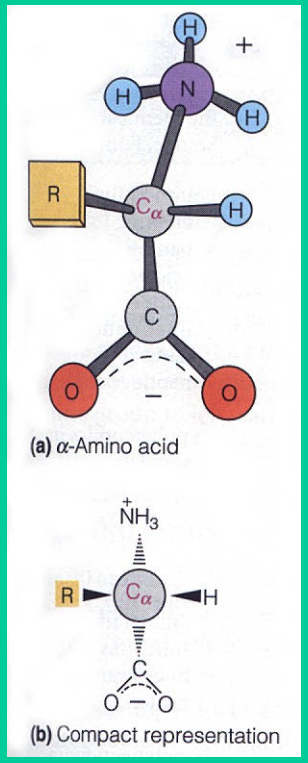


From DNA to Cell Function

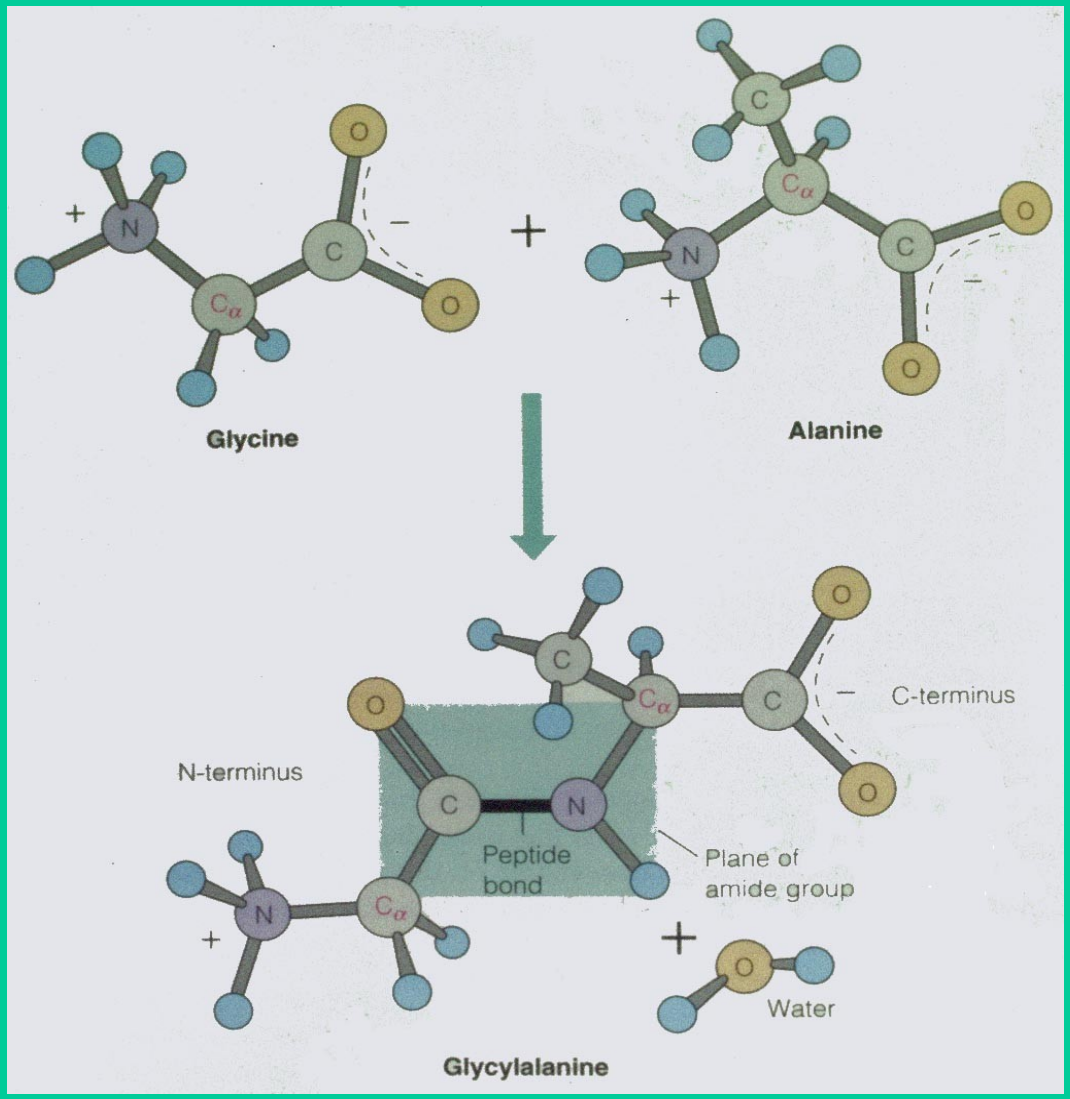


Protein Structure Prediction

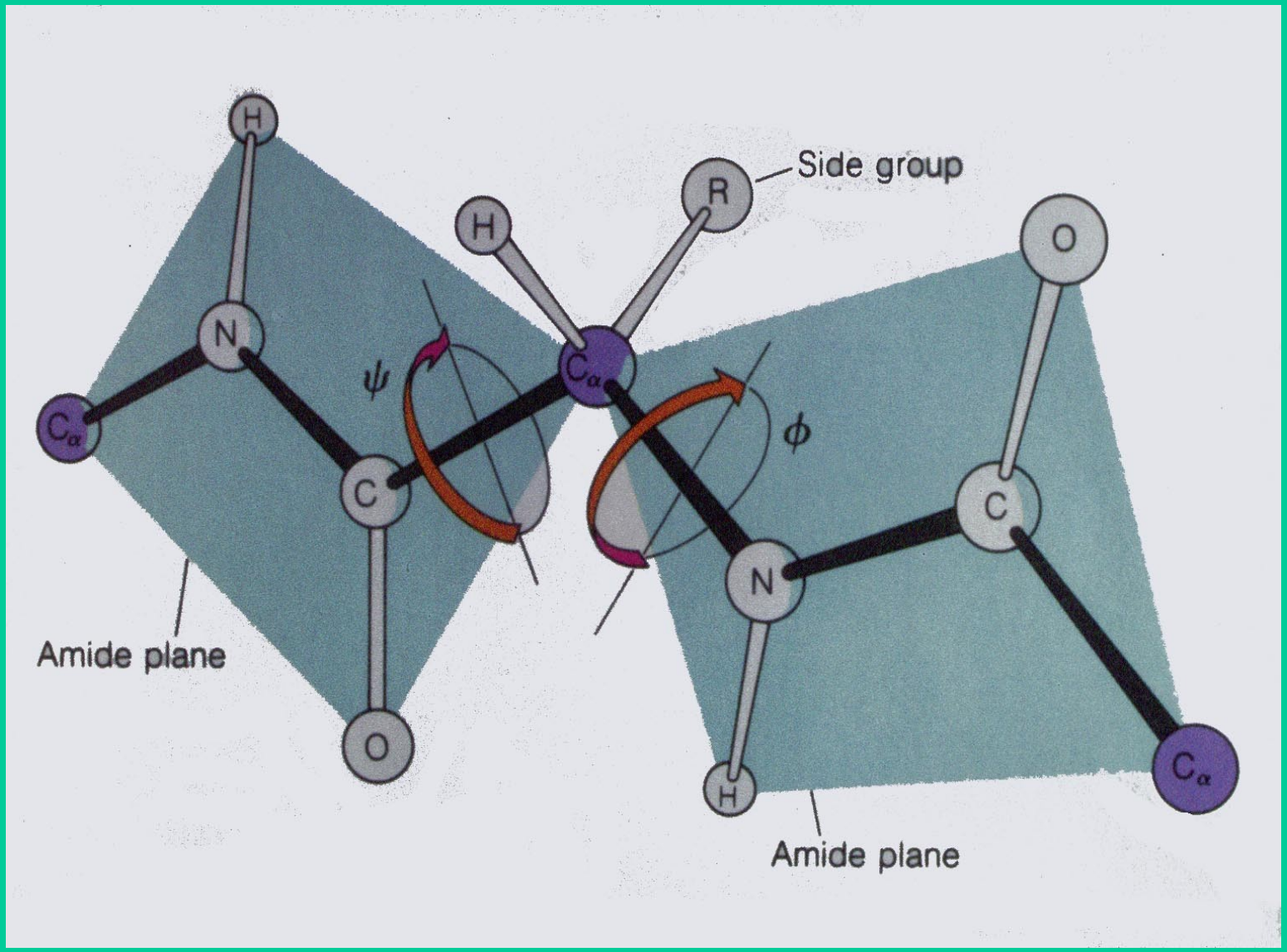
- Experimental Techniques
  - X-ray Crystallography
  - NMR
- Limitations of Current Experimental Techniques
  - Protein DataBank (PDB) -> 30,000 protein structures
  - Unique structure 4000 to 5000 only
  - Non-Redudant (NR) -> 10,00,000 proteins
- Importance of Structure Prediction
  - Fill gap between known sequence and structures
  - Protein Engg. To alter function of a protein
  - Rational Drug Design
- **World Wide Recognition of Problem**
  - **CASP/CAFASP Competition (Olympic 2000)**
  - **Most Wanted (TOP 10)**
  - **Metaserver for Structure Prediction**


H
+
H
N
H
R
Cα
H
C
O
–
O
(a) α-Amino acid
+
NH3
R
Cα
H
C
O
–
O
(b) Compact representation

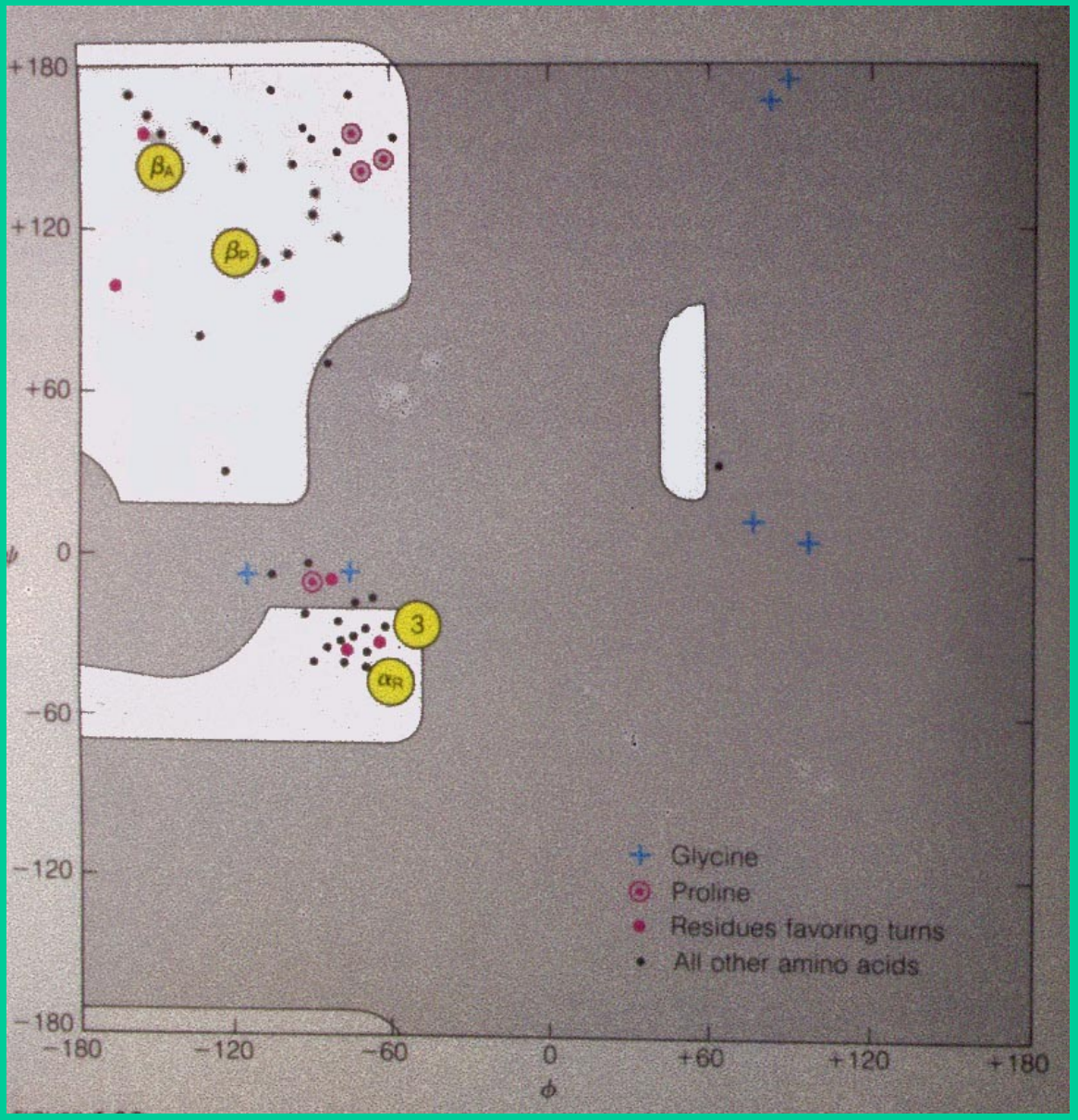
Peptide Bond



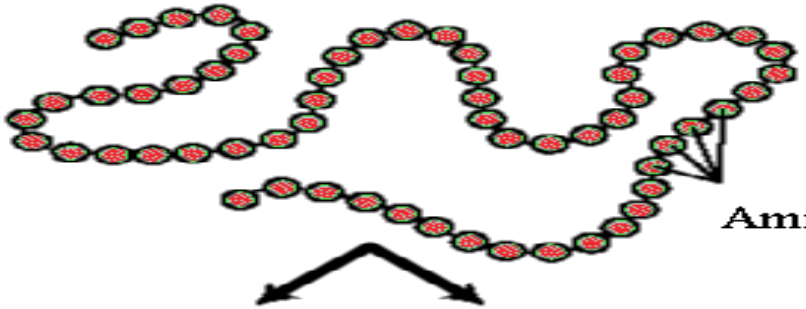
Dihedral Angles



Ramachandran Plot

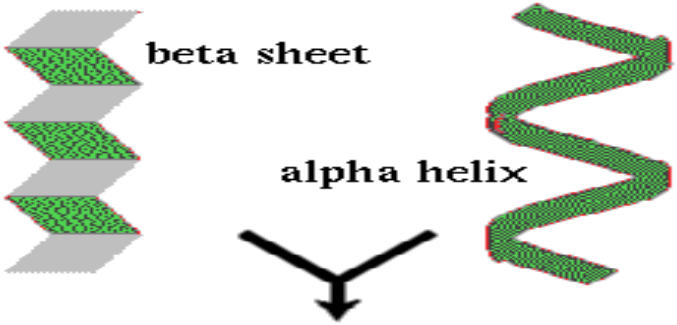


Different Levels of Protein Structure



Primary protein structure
is sequence of amino acids

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

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

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

- 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

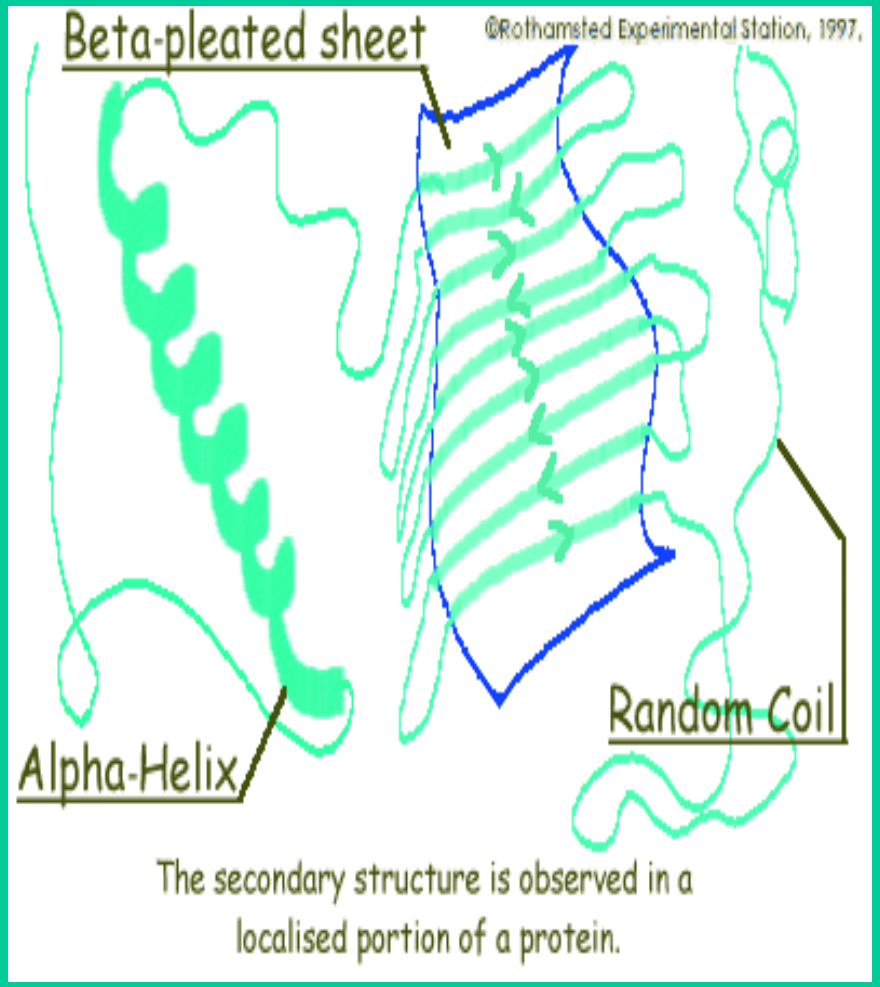
Protein Structure Prediction

- Tertiary Structure Prediction (TSP)
  - Comparative Modelling
  - Energy Minimization Techniques
  - Ab-Initio Prediction (Segment Based)
  - Threading Based Approach
- Limitations of TSP
  - Difficult to predict in absence of homology
  - Computation requirement too high
  - Fail in absence of known examples
- Secondary Structure prediction (SSP)
  - An Intermidiate Step in TSP
  - Most Successful in absence of homology
  - Helix (3), Strand (2) and Coil (3)
  - DSSP for structure assignment

Protein Secondary Structure Prediction

- Existing SSP Methods
  - Statistical Methods (Chou,GOR)
  - Physio-chemical Methods
  - A.I. (Neural Network Approach)
  - Consensus and Multiple Alignment
- Our Method APSSP of SSP
  - Neural Network
  - Example Based Learnning
  - Multiple Alignment
- Steps involved in APSSP
  - Blast search against protein sequence (NR)
  - Multiple Alignment (ClustalW)
  - Profile by HMMER, Result by Email
- Recogntion: CASP,CAFASP,LiveBench, MetaServer

Protein Secondary Structure



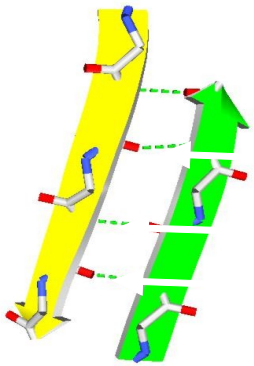
Secondary Structure

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

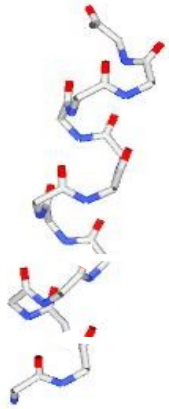
Irregular Secondary Structure (Tight turns, Random coils, bulges)

Secondary structure prediction

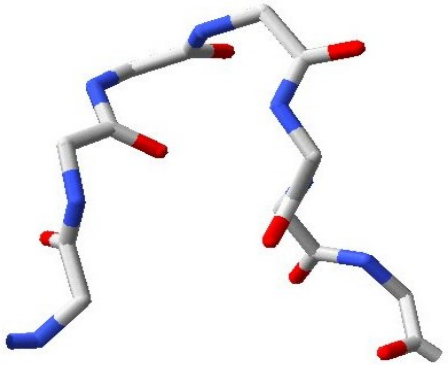
3-state model: Helix (H), Strand (E), Loop (L)



(E) β–Στρανδ, νον–λοχαλ ιντεραχτιονσ



(H) α–Ηελιξ, λοχαλ ιντεραχτιονσ



(Λ) Λοοπ νον–ρεγυλαρ ιντεραχτιονσ

```
SEQ     KELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLD
SS        EEEE    E       E  E    EEEEEE     EEEEEE  EEEEEEHHHEEEE
```

No information about tight turns ?

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)

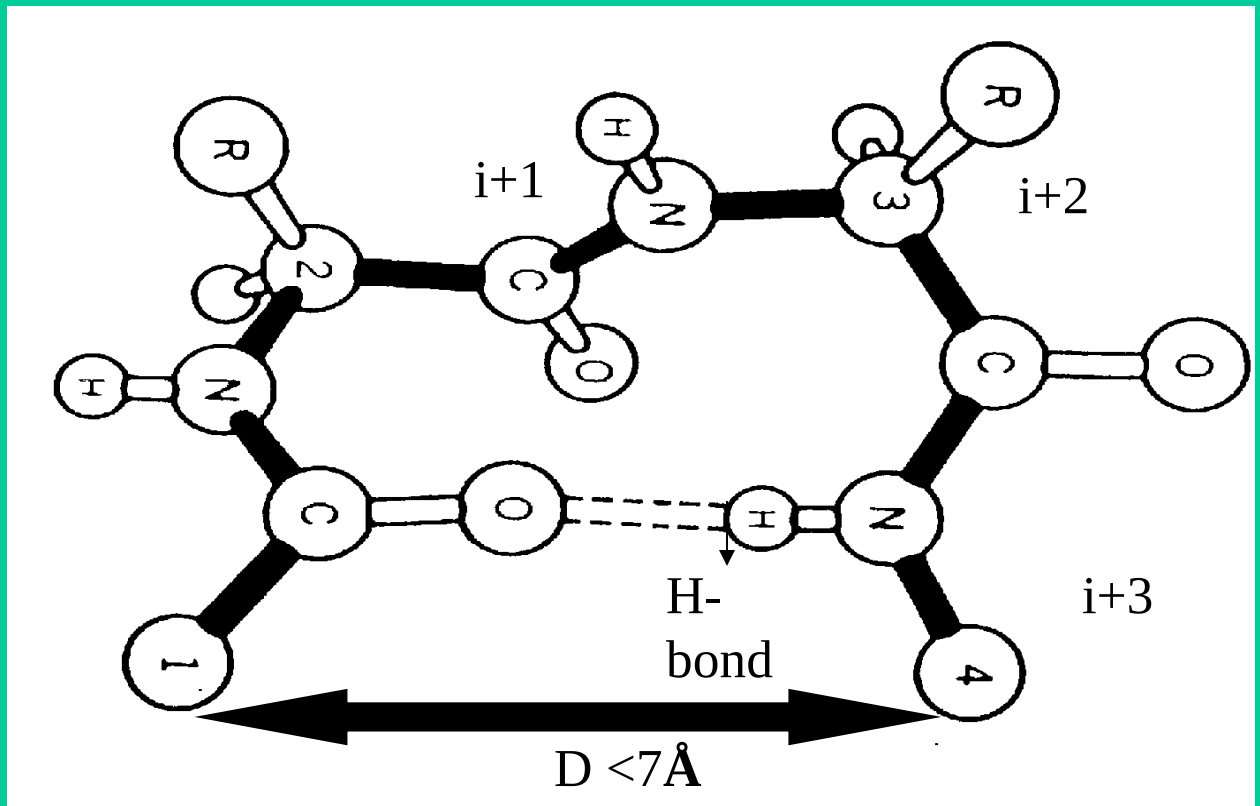
Prediction of tight turns

- 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

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).

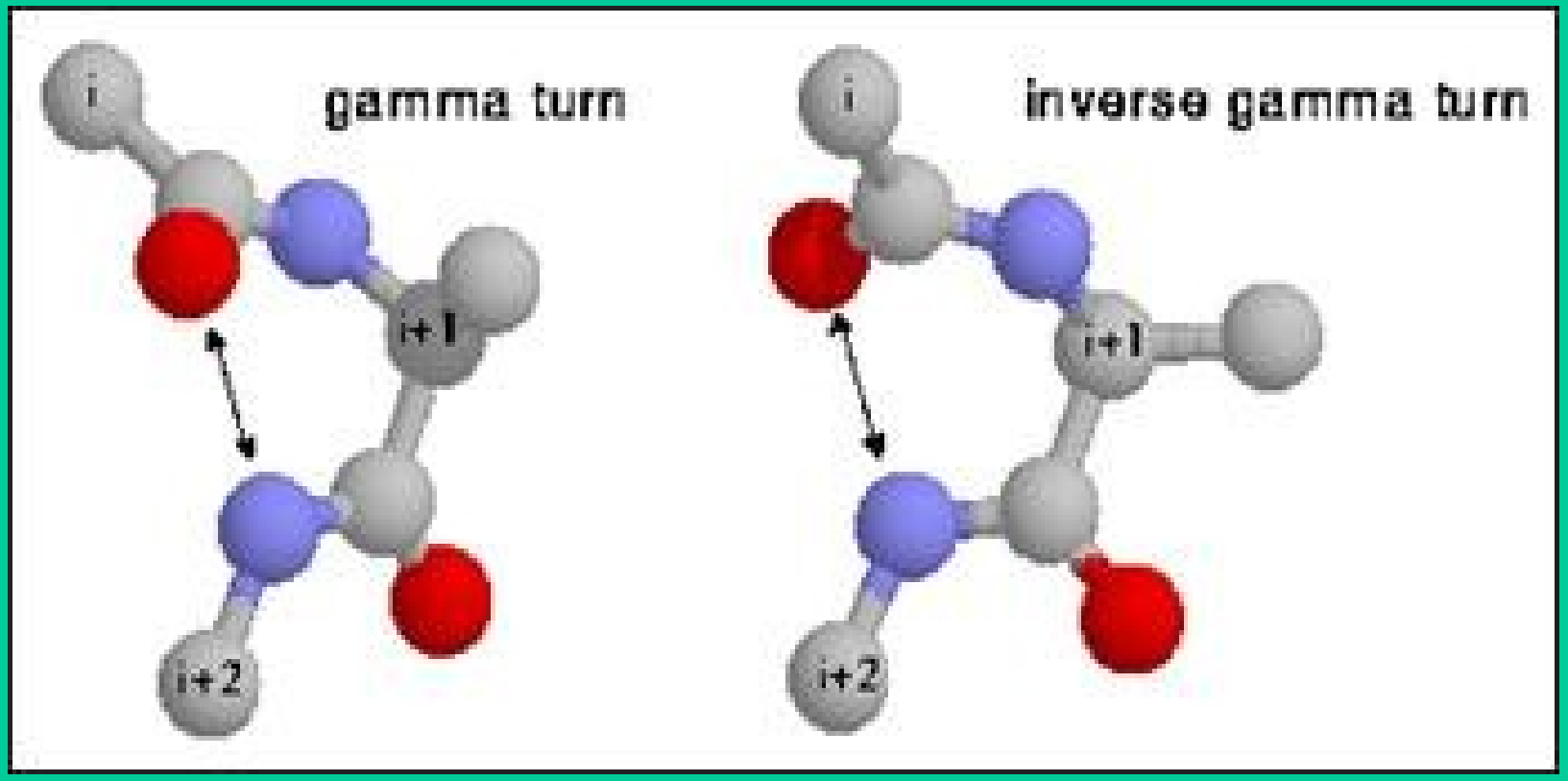
The conformation of β-turn is defined in terms of ϕ and ψ of two central residues, *i*+1 and *i*+2 and can be classified into different types on the basis of ϕ and ψ.



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.



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: Consensus method for Beta Turn prediction (Kaur and Raghava 2002, Bioinformatics)

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.

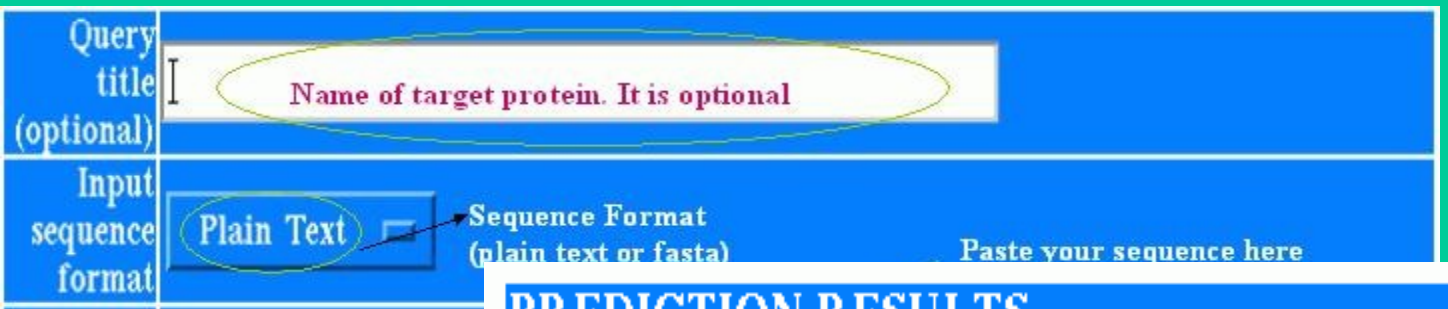
- **Two feed-forward back-propagation networks with a single hidden layer are used where the first sequence-structure network is trained with the multiple sequence alignment in the form of PSI-BLAST generated position specific scoring matrices.**

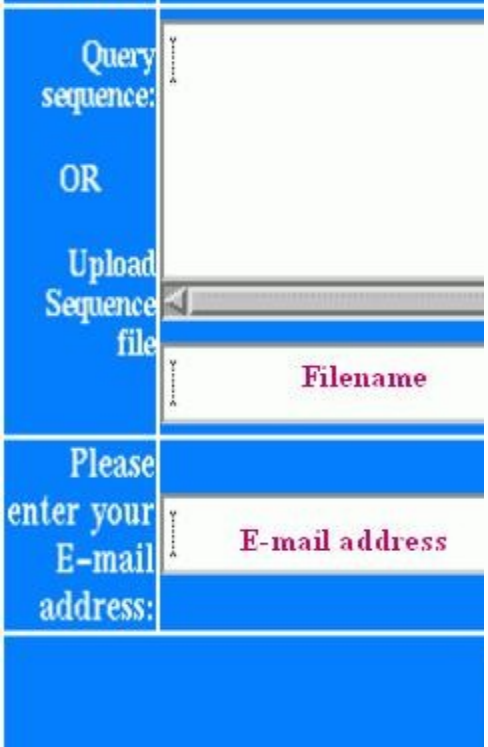
- **The initial predictions from the first network and PSIPRED predicted secondary structure are used as input to the second sequence-structure network to refine the predictions obtained from the first net.**

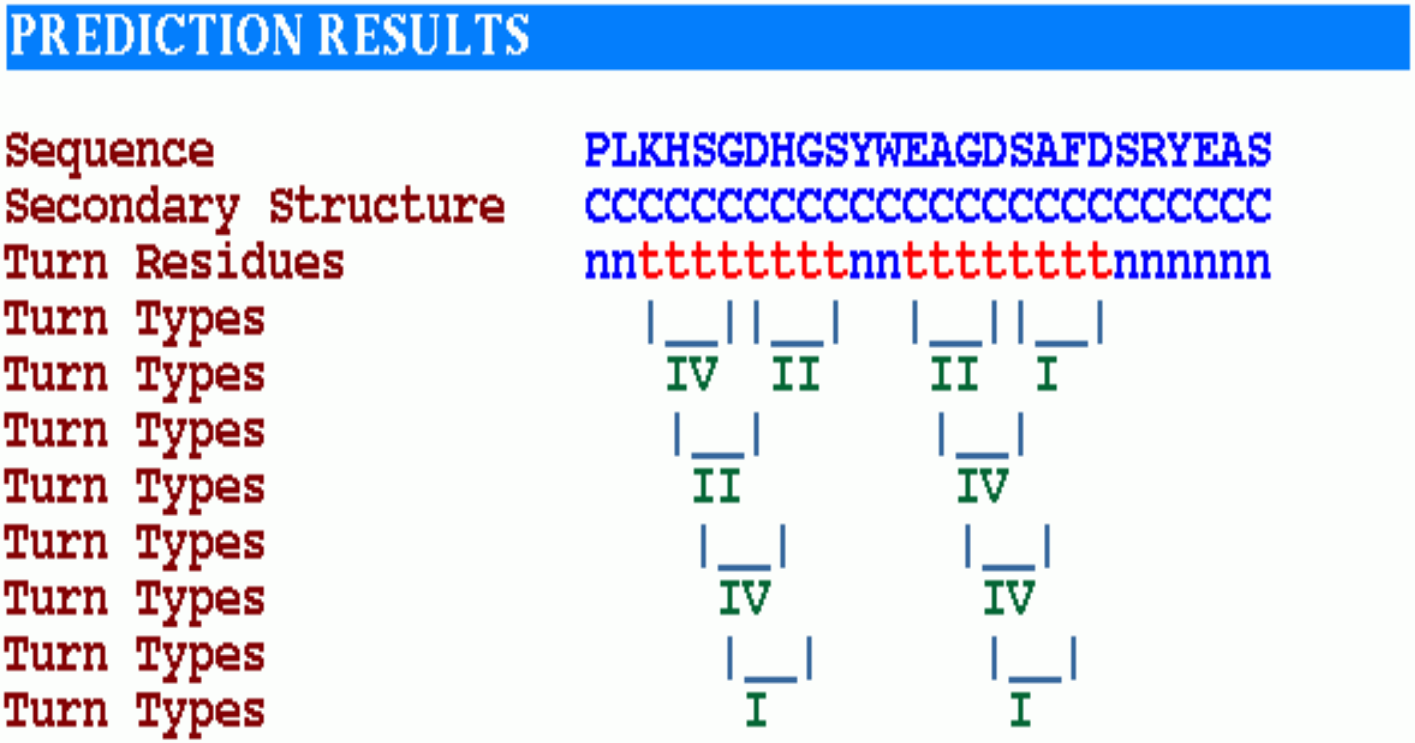
- **The final network yields an overall prediction accuracy of 75.5% when tested by seven-fold cross-validation on a set of 426 non-homologous protein chains. The corresponding Q*pred*., Q*obs*. and MCC values are 49.8%, 72.3% and 0.43 respectively and are the best among all the previously published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

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

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



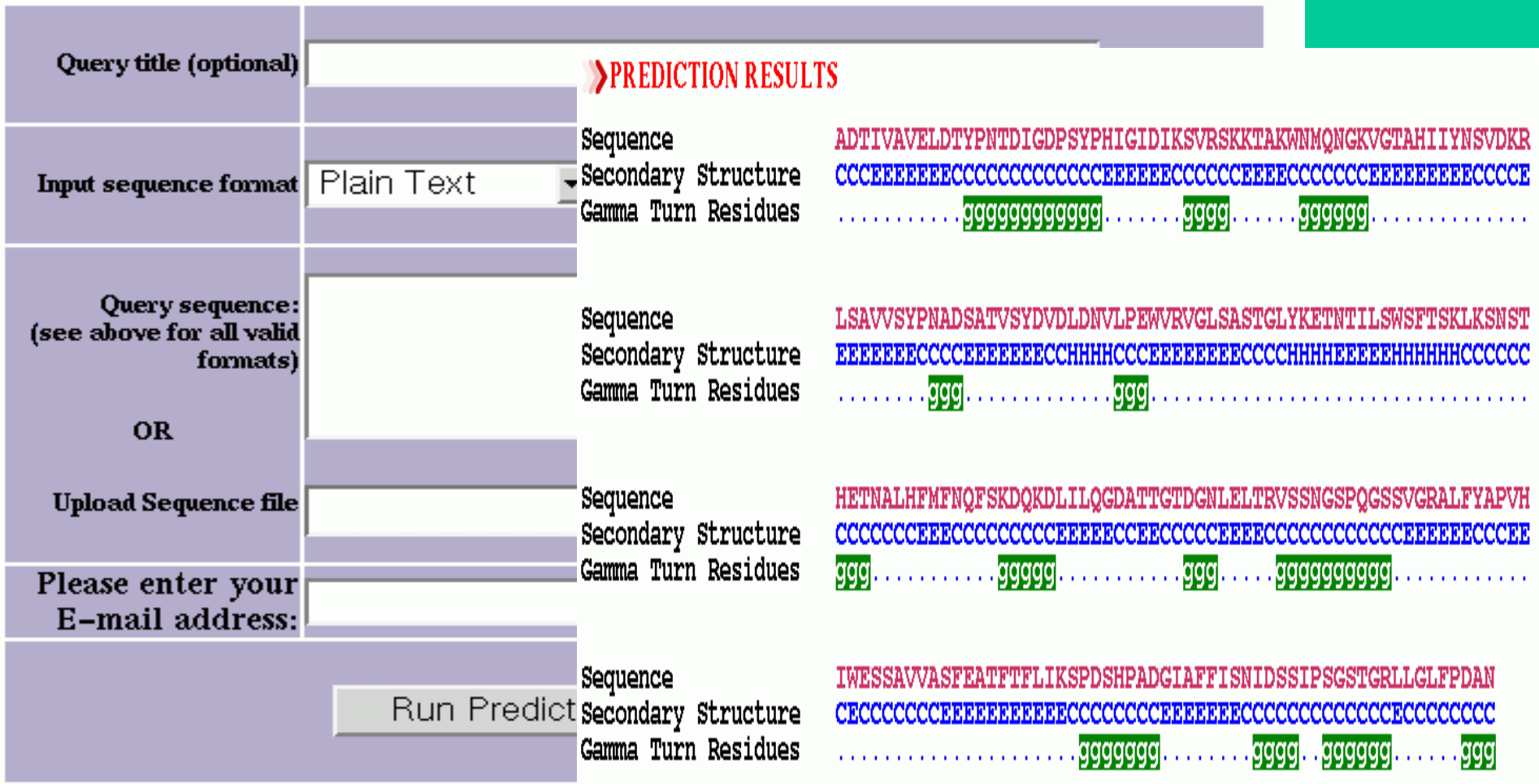




Gammapred: A server for prediction of γ-turns in proteins

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

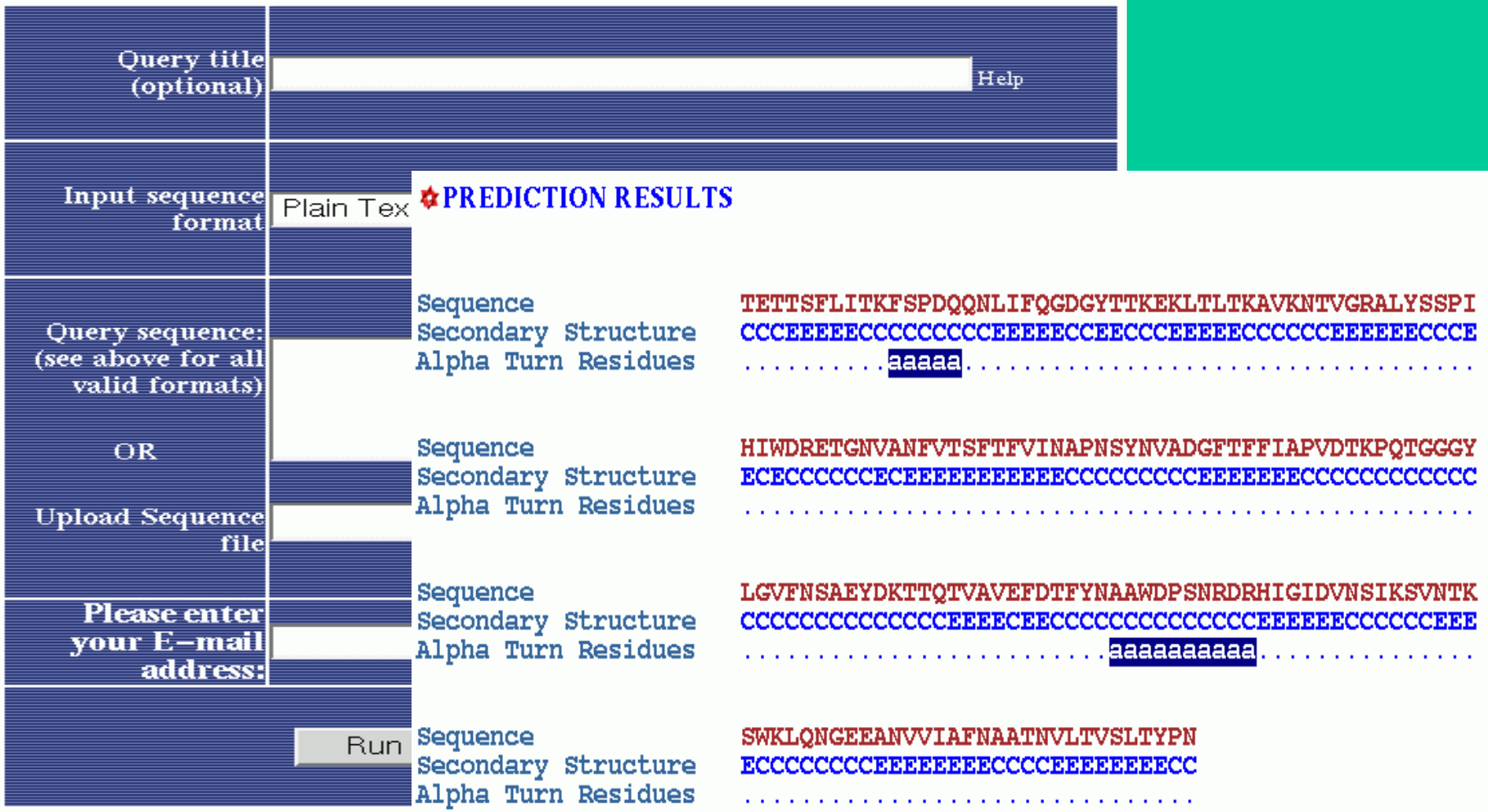
Harpreet Kaur and G P S Raghava (2003) A neural network based method for prediction of γ-turns in proteins from multiple sequence alignment. ***Protein Science*** 12, 923-929.



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* .

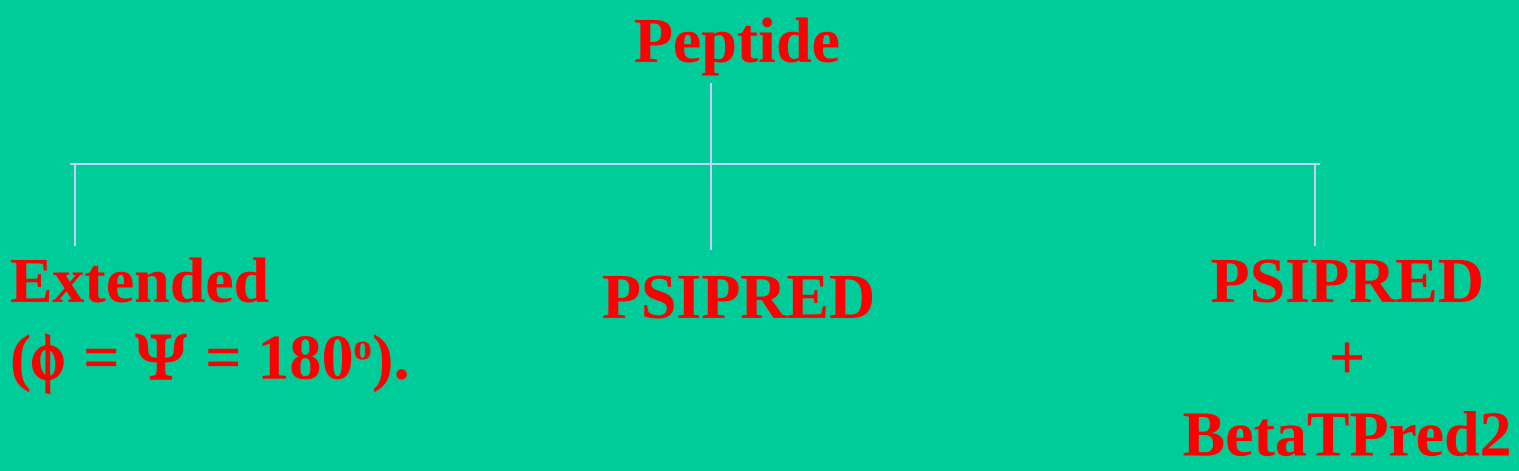


Contribution of β-turns in tertiary structure prediction of bioactive peptides

- 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…….

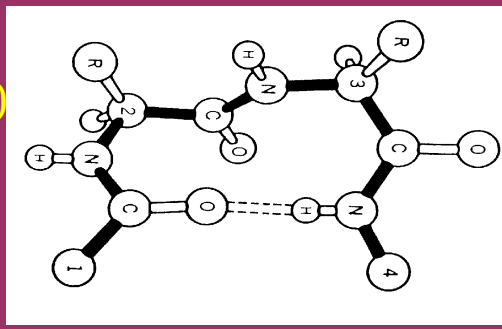
Averaged backbone root mean deviation before and after energy minimization and dynamics simulations.

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.

Protein Structure Prediction

- Regular Secondary Structure Prediction (α-helix β-sheet)
  - APSSP2: Highly accurate method for secondary structure prediction
  - Participate in all competitions like EVA, CAFASP and CASP (In top 5 methods)
  - Combines memory based reasoning (MBR) and ANN methods
- Irregular secondary structure prediction methods (Tight turns)
  - Betatpred: Consensus method for β-turns prediction
    - Statistical methods combined
    - Kaur and Raghava (2001) Bioinformatics
  - Bteval : Benchmarking of β-turns prediction
    - Kaur and Raghava (2002) J. Bioinformatics and Computational Biology, 1:495:504
  - BetaTpred2: Highly accurate method for predicting β-turns (ANN, SS, MA)
    - Multiple alignment and secondary structure information
    - Kaur and Raghava (2003) Protein Sci 12:627-34
  - BetaTurns: Prediction of β-turn types in proteins
    - Evolutionary information
    - Kaur and Raghava (2004) Bioinformatics 20:2751-8.
  - AlphaPred: Prediction of α-turns in proteins
    - Kaur and Raghava (2004) Proteins: Structure, Function, and Genetics 55:83-90
  - GammaPred: Prediction of γ-turns in proteins
    - Kaur and Raghava (2004) Protein Science; 12:923-929.



Protein Structure Prediction

- BhairPred: Prediction of Supersecondary structure prediction
  - Prediction of Beta Hairpins
  - Utilize ANN and SVM pattern recognition techniques
  - Secondary structure and surface accessibility used as input
  - Manish et al. (2005) Nucleic Acids Research (In press)
- TBBpred: Prediction of outer membrane proteins
  - Prediction of trans membrane beta barrel proteins
  - Prediction of beta barrel regions
  - Application of ANN and SVM + Evolutionary information
  - **Natt et al. (2004) Proteins: 56:11-8**
- ARNHpred: Analysis and prediction side chain, backbone interactions
  - Prediction of aromatic NH interactions
  - Kaur and Raghava (2004) FEBS Letters 564:47-57 .
- SARpred: Prediction of surface accessibility (real accessibility)
  - Multiple alignment (PSIBLAST) and Secondary structure information
  - ANN: Two layered network (sequence-structure-structure)
  - **Garg et al., (2005) Proteins (In Press)**
- PepStr: Prediction of tertiary structure of Bioactive peptides

Performance of SARpred, Pepstr and BhairPred were checked on CASP6 proteins

Thankyou