

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 Engineering. To alter function of a protein
- Rational Drug Design

Amino acid: Building Block of Protein



Peptide Bond



Dihedral Angles



Techniques of Structure Prediction

- Computer simulation based on energy calculation
 - Based on physico-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 prior assumptions and attempt to locate global minimization.
- Static Minimization Methods
 - Assume that atoms in protein is in static form

Dynamical Minimization Methods

- Motions of atoms also considered
- Monte Carlo simulation (stochastic in nature, time is not considered)
- Molecular Dynamics

Limitations

- large number of degree of freedom, CPU power not adequate
- Interaction potential is not good enough to model

Homology Modeling

- Need homologues of known protein structure
- Backbone modeling
- Side chain modeling
- 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 modeling
 - 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 Modeling
 - 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 Intermediate Step in TSP
 - Most Successful in absence of homology
 - Helix (3), Strand (2) and Coil (3)
 - DSSP for structure assignment

Protein Secondary Structure



Secondary Structure

Regular Secondary Structure (α-helices, β-sheets) Irregular Secondary Structure (Tight turns, Random coils, bulges)

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

Secondary structure prediction

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



Tight turns

Туре	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 $\alpha\text{-turns}$
- 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 $C^{\alpha}(i)-C^{\alpha}(i+3)$ 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)

BetaTurns: A web server for prediction of β**-turn types** (<u>http://www.imtech.res.in/raghava/betaturns/</u>)



Gamma pred: 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 <u>Protein</u> <u>Science</u> 12, 923-929

Query title (optional)	>PREDICTION RESU	LTS
Input sequence format	Plain Text Sequence Gamma Turn Residues	ADTIVAVELDTYPNTDIGDPSYPHIGIDIKSVRSKKTAKWNMQNGKVGTAHIIYNSVDKR CCCEEEEEEECCCCCCCCCCCCCEEEEECCCCCCCEEEEE
Query sequence : (see above for all valid formats) OR	Sequence Secondary Structure Gamma Turn Residues	LSAVVSYPNADSATVSYDVDLDNVLPEWVRVGLSASTGLYKETNTILSWSFTSKLKSNST EEEEEEECCCCEEEEEEECCHHHHCCCEEEEEEEECCCCHHHHEEEEEHHHHHH
Upload Sequence file Please enter your E–mail address:	Sequence Secondary Structure Gamma Turn Residues	HETNALHFMFNQFSKDQKDLILQGDATTGTDGNLELTRVSSNGSPQGSSVGRALFYAPVH CCCCCCCEEEECCCCCCCCCEEEEECCEECCCCCCCCC
	Sequence Run Predict Secondary Structure Gamma Turn Residues	IWESSAVVASFEATFTFLIKSPDSHPADGIAFFISNIDSSIPSGSTGRLLGLFPDAN CECCCCCCCEEEEEEEEEECCCCCCCCEEEEEEECCCCCC

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. <u>Proteins</u>

