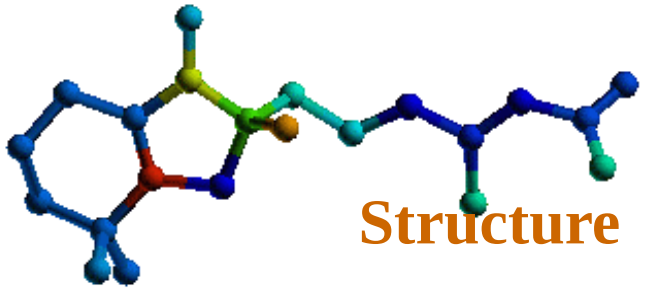
Prediction of protein Secondary, Supper Secondary and tertiary Structure

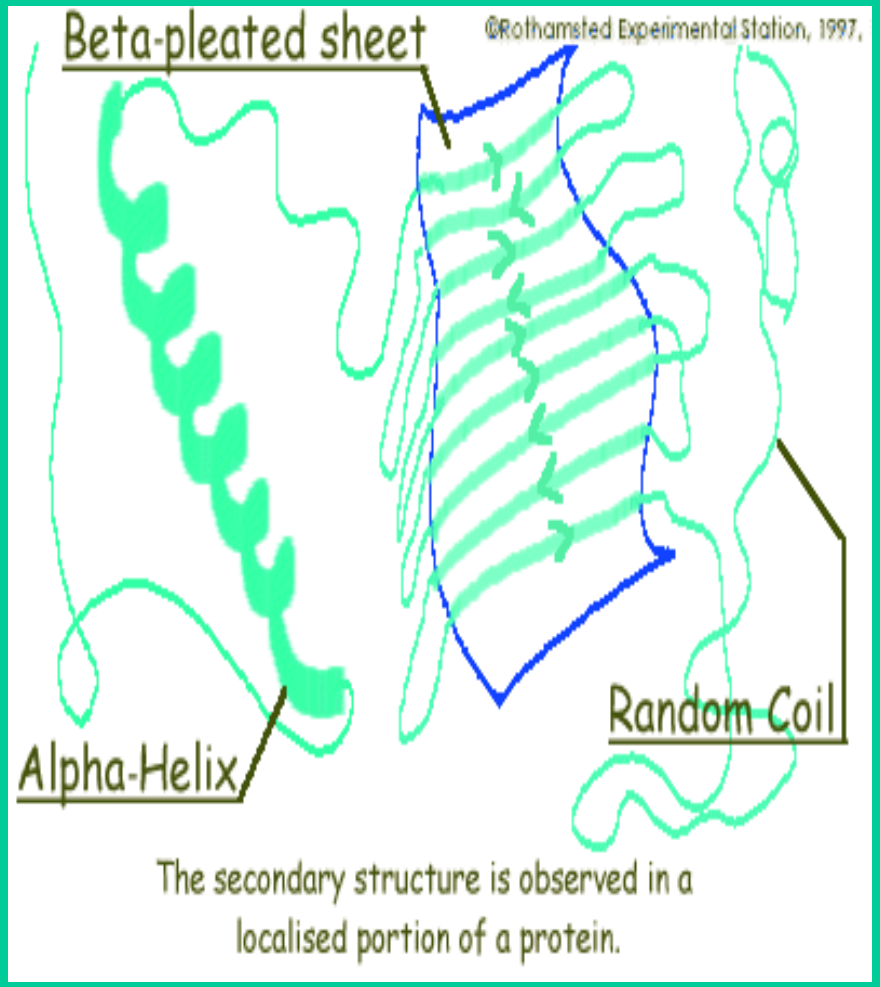
Protein Sequence +

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

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



Protein Secondary Structure



Secondary Structure

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

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

Assignment of Secondary Structure

- Program
  - DSSP (Sander Group)
  - Stride (Argos Group)
  - Pcurve
- DSSP
  - 3 helix states (I=3,4,5)
  - 2 Sheets (isolated and extended)
  - Irregular Regions

dssp

- The DSSP program defines secondary structure, geometrical features and solvent exposure of proteins, given atomic coordinates in Protein Data Bank format
- Usage: dssp [-na] [-v] pdb_file [dssp_file]
- Output :

```
   24   26   E  H  < S+     0   0  132
   25   27   R  H  < S+     0   0  125
   26   28   N     <        0   0   41
   27   29   K              0   0  197
   28        !              0   0    0
   29   34   C              0   0   73
   30   35   I  E     -cd  58  89B   9
   31   36   L  E     -cd  59  90B   2
   32   37   V  E     -cd  60  91B   0
   33   38   G  E     -cd  61  92B   0
```

Automatic assignment programs

- DSSP (http://www.cmbi.kun.nl/gv/dssp/)
- STRIDE (http://www.hgmp.mrc.ac.uk/Registered/Option/stride.html)

#	RESIDUE	AA	STRUCTURE	BP1	BP2	ACC	N-H-->O	O-->H-N	N-H-->O	O-->H-N	TCO	KAPPA	ALPHA	PHI	PSI	X-CA	Y-CA	Z-CA
1	4	A	E	0	0	205	0, 0.0	2,-0.3	0, 0.0	0, 0.0	0.000	360.0	360.0	360.0	113.5	5.7	42.2	25.1
2	5	A	H -	0	0	127	2, 0.0	2,-0.4	21, 0.0	21, 0.0	-0.987	360.0	-152.8	-149.1	154.0	9.4	41.3	24.7
3	6	A	V -	0	0	66	-2,-0.3	21,-2.6	2, 0.0	2,-0.5	-0.995	4.6	-170.2	-134.3	126.3	11.5	38.4	23.5
4	7	A	I E -A	23	0A	106	-2,-0.4	2,-0.4	19,-0.2	19,-0.2	-0.976	13.9	-170.8	-114.8	126.6	15.0	37.6	24.5
5	8	A	I E -A	22	0A	74	17,-2.8	17,-2.8	-2,-0.5	2,-0.9	-0.972	20.8	-158.4	-125.4	129.1	16.6	34.9	22.4
6	9	A	Q E -A	21	0A	86	-2,-0.4	2,-0.4	15,-0.2	15,-0.2	-0.910	29.5	-170.4	-98.9	106.4	19.9	33.0	23.0
7	10	A	A E +A	20	0A	18	13,-2.5	13,-2.5	-2,-0.9	2,-0.3	-0.852	11.5	172.8	-108.1	141.7	20.7	31.8	19.5
8	11	A	E E +A	19	0A	63	-2,-0.4	2,-0.3	11,-0.2	11,-0.2	-0.933	4.4	175.4	-139.1	156.9	23.4	29.4	18.4
9	12	A	F E -A	18	0A	31	9,-1.5	9,-1.8	-2,-0.3	2,-0.4	-0.967	13.3	-160.9	-160.6	151.3	24.4	27.6	15.3
10	13	A	Y E -A	17	0A	36	-2,-0.3	2,-0.4	7,-0.2	7,-0.2	-0.994	16.5	-156.0	-136.8	132.1	27.2	25.3	14.1
11	14	A	L E >> -A	16	0A	24	5,-3.2	4,-1.7	-2,-0.4	5,-1.3	-0.929	11.7	-122.6	-120.0	133.5	28.0	24.8	10.4
12	15	A	N T 45S+	0	0	54	-2,-0.4	-2, 0.0	2,-0.2	0, 0.0	-0.884	84.3	9.0	-113.8	150.9	29.7	22.0	8.6
13	16	A	P T 45S+	0	0	114	0, 0.0	-1,-0.2	0, 0.0	-2, 0.0	-0.963	125.4	60.5	-86.5	8.5	32.0	21.6	6.8
14	17	A	D T 45S-	0	0	66	2,-0.1	-2,-0.2	1,-0.1	3,-0.1	0.752	89.3	-146.2	-64.6	-23.0	33.0	25.2	7.6
15	18	A	Q T <5 +	0	0	132	-4,-1.7	2,-0.3	1,-0.2	-3,-0.2	0.936	51.1	134.1	52.9	50.0	33.3	24.2	11.2
16	19	A	S E < +A	11	0A	44	-5,-1.3	-5,-3.2	2, 0.0	2,-0.3	-0.877	28.9	174.9	-124.8	156.8	32.1	27.7	12.3
17	20	A	G E -A	10	0A	28	-2,-0.3	2,-0.3	-7,-0.2	-7,-0.2	-0.893	15.9	-146.5	-151.0	-178.9	29.6	28.7	14.8
18	21	A	E E -A	9	0A	14	-9,-1.8	-9,-1.5	-2,-0.3	2,-0.4	-0.979	5.0	-169.6	-158.6	146.0	28.0	31.5	16.7
19	22	A	F E +A	8	0A	3	12,-0.4	12,-2.3	-2,-0.3	2,-0.3	-0.982	27.8	149.2	-139.1	120.3	26.5	32.2	20.1
20	23	A	M E -AB	7	30A	0	-13,-2.5	-13,-2.5	-2,-0.4	2,-0.4	-0.983	39.7	-127.8	-152.1	161.6	24.5	35.4	20.6
21	24	A	F E -AB	6	29A	45	8,-2.4	7,-2.9	-2,-0.3	8,-1.0	-0.934	23.9	-164.1	-112.5	137.7	21.7	37.0	22.6
22	25	A	D E -AB	5	27A	6	-17,-2.8	-17,-2.8	-2,-0.4	2,-0.5	-0.948	6.9	-165.0	-123.7	138.3	18.9	38.9	20.8
23	26	A	F E > S-AB	4	26A	76	3,-3.5	3,-2.1	-2,-0.4	-19,-0.2	-0.947	78.4	-27.2	-127.3	111.5	16.4	41.3	22.3
24	27	A	D T 3 S-	0	0	74	-21,-2.6	-20,-0.1	-2,-0.5	-1,-0.1	0.904	128.9	-46.6	50.4	45.0	13.4	42.1	20.2
25	28	A	G T 3 S+	0	0	20	-22,-0.3	2,-0.4	1,-0.2	-1,-0.3	0.291	118.8	109.3	84.7	-11.1	15.4	41.4	17.0
26	29	A	D E < S-B	23	0A	114	-3,-2.1	-3,-3.5	109, 0.0	2,-0.3	-0.822	71.8	-114.7	-103.1	140.3	18.4	43.4	18.1
27	30	A	E E -B	22	0A	8	-2,-0.4	-5,-0.3	-5,-0.2	3,-0.1	-0.525	24.9	-177.7	-74.1	127.5	21.8	41.8	19.1

Type of Secondary Structure Prediction

- Information based classification
  - Property based methods (Manual / Subjective)
  - Residue based methods
  - Segment or peptide based approaches
  - Application of Multiple Sequence Alignment
- Technical classification
  - Statistical Methods
    - Chou & fashman (1974)
    - GOR
  - Artificial Itellegence Based Methods
    - Neural Network Based Methods (1988)
    - Nearest Neighbour Methods (1992)
    - Hidden Markove model (1993)
    - Support Vector Machine based methods

Lim

J. Molecular Biology (1974) 88, 873.

Predicts: alpha helix, beta strand, irregular.

Accuracy: 80-85 %.

Based on: short-range and long-range interactions to stabilize the secondary structures. Takes into account packing issues.

Data (training) set: 25 proteins (structures known to date)
Test set: 25 proteins

Method:

1. predict secondary structure based on RULES developed from known structures
2. plot schemes on primary protein sequence
3. following known rules (1-6), locate helical regions
4. following known rules (1-8), locate beta regions

Chou Fasman

Biochemistry (1974) 13:2, 222.

Predicts: alpha helix, beta strand, beta (reverse) turn, none.

Accuracy: 77 %.

Based on: short-range and medium-range interactions play a predominant role in determining secondary structure.

Data (training) set: 19 proteins (structures known to date)
Test set: 19 proteins

Method:

1. assign each residue helix potential, beta potential, turn potential
2. locate clusters of helix formers, helix breakers, etc.
   H(a), h(a), I(a), i(a), b(a), B(a)
3. search for helical regions (4 out of 6 H or h)
4. search for beta regions
5. search for turns

Chou-Fasman Rules (Mathews, Van Holde, Ahern)

Amino Acid	α-Helix	β-Sheet	Turn
Ala	1.29	0.90	0.78
Cys	1.11	0.74	0.80
Leu	1.30	1.02	0.59
Met	1.47	0.97	0.39
Glu	1.44	0.75	1.00
Gln	1.27	0.80	0.97
His	1.22	1.08	0.69
Lys	1.23	0.77	0.96
Val	0.91	1.49	0.47
Ile	0.97	1.45	0.51
Phe	1.07	1.32	0.58
Tyr	0.72	1.25	1.05
Trp	0.99	1.14	0.75
Thr	0.82	1.21	1.03
Gly	0.56	0.92	1.64
Ser	0.82	0.95	1.33
Asp	1.04	0.72	1.41
Asn	0.90	0.76	1.23
Pro	0.52	0.64	1.91
Arg	0.96	0.99	0.88

Favors α-Helix

Favors β-Sheet

Favors Turns

Chou-Fasman

- First widely used procedure
- If propensity in a window of six residues (for a helix) is above a certain threshold the helix is chosen as secondary structure.
- If propensity in a window of five residues (for a beta strand) is above a certain threshold then beta strand is chosen.
- The segment is extended until the average propensity in a 4 residue window falls below a value.
- Output-helix, strand or turn.

GOR method

- Garnier, Osguthorpe & Robson
- Assumes amino acids up to 8 residues on each side influence the ss of the central residue.
- Frequency of amino acids at the central position in the window, and at -1, ... -8 and +1,....+8 is determined for α, β and turns (later other or coils) to give three 17 x 20 scoring matrices.
- Calculate the score that the central residue is one type of ss and not another.
- Correctly predicts ~64%.

$$I(S_j;R_1,R_2,\ldots..R_{last}) \simeq \sum_{m=-8}^{m=+8} I(S_j;R_{j+m})$$

Directional information measure for the α-helical conformation†

Amino acid residue	Residue position‡ (centinats)															
	$j-8$		$j-6$		$j-4$		$j-2$	j	$j+2$		$j+4$		$j+6$		$j+8$	
Gly	−5	−10	−15	−20	−30	−40	−50	−60	−86	−60	−50	−40	−30	−20	−15	−10	−5
Ala	5	10	15	20	30	40	50	60	65	60	50	40	30	20	15	10	5
Val	0	0	0	0	0	0	5	10	14	10	5	0	0	0	0	0	0
Leu	0	5	10	15	20	25	28	30	32	30	28	25	20	15	10	5	0
Ile	5	10	15	20	25	20	15	10	6	0	−10	−15	−20	−25	−20	−10	−5
Ser	0	−5	−10	−15	−20	−25	−30	−35	−39	−35	−30	−25	−20	−15	−10	−5	0
Thr	0	0	0	−5	−10	−15	−20	−25	−26	−25	−20	−15	−10	−5	0	0	0
Asp	0	−5	−10	−15	−20	−15	−10	0	5	10	15	20	20	20	15	10	5
Glu	0	0	0	0	10	20	60	70	78	78	78	78	78	70	60	40	20
Asn	0	0	0	0	−10	−20	−30	−40	−51	−40	−30	−20	−10	0	0	0	0
Gln	0	0	0	0	5	10	20	20	10	−10	−20	−20	−10	−5	0	0	0
Lys	20	40	50	55	60	60	50	30	23	10	5	0	0	0	0	0	0
His	10	20	30	40	50	50	50	30	12	−20	−10	0	0	0	0	0	0
Arg	0	0	0	0	0	0	0	0	−9	−15	−20	−30	−40	−50	−50	−30	−10
Phe	0	0	0	0	0	5	10	15	16	15	10	5	0	0	0	0	0
Tyr	−5	−10	−15	−20	−25	−30	−35	−40	−45	−40	−35	−30	−25	−20	−15	−10	−5
Trp	−10	−20	−40	−50	−50	−10	0	10	12	10	0	−10	−50	−50	−40	−20	−10
Cys	0	0	0	0	0	0	−5	−10	−13	−10	−5	0	0	0	0	0	0
Met	10	20	25	30	35	40	45	50	53	50	45	40	35	30	25	20	10
Pro	−10	−20	−40	−60	−80	−100	−120	−140	−77	−60	−30	−20	−10	0	0	0	0

Garnier, Osguthorpe, Robson

J. Molecular Biology (1978) 120, 97.

Predicts: alpha helix, beta strand, beta (reverse) turn, coil.

Accuracy: 60 %.

Based on: Based on single residue determination vs. neighboring interactions determination.
Optimized for predicted protein to include expected percentage secondary structure.

Data (training) set: 25 proteins (structures known to date)
Test set: 25 proteins

Method:

1. evaluation of information state for each residue, each conformational state

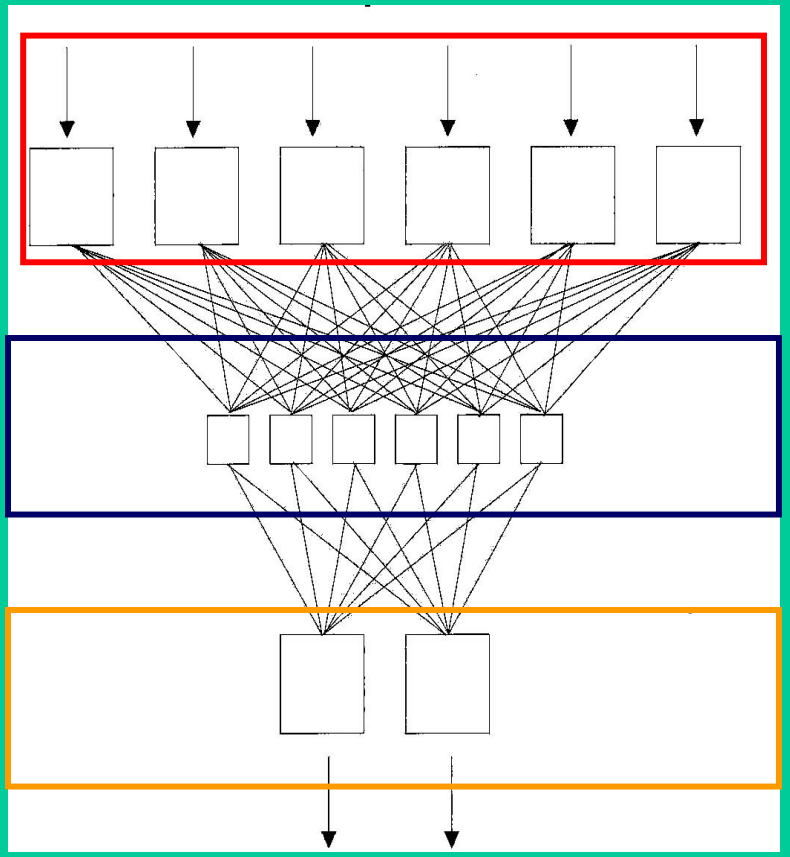
$$I\ (S_j\ ;\ R_1\ R_2\ \ldots\ R_{last})\ \sim\ \sum_{m=-8}^{m=+8} I\ (S_j;\ R_{j+m})$$

2. locate conformation with highest content
3. variables: decision constant (optimized to experimental CD)
   run constant (includes neighboring effects)
4. optional: homologous sequences
   information content from each homolog is added and divided by # homologs

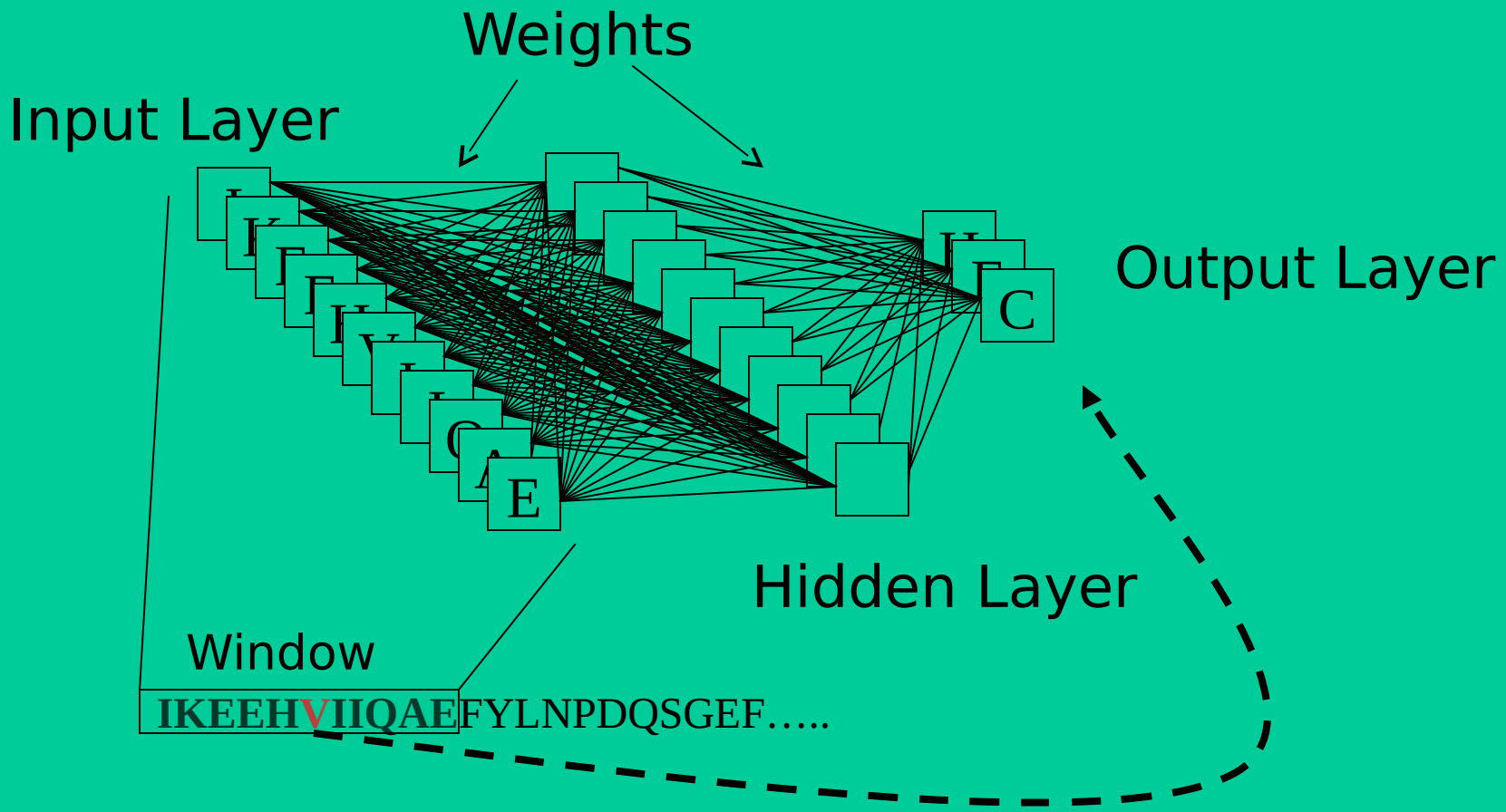
Artificial Neural Network

General structure of ANN :

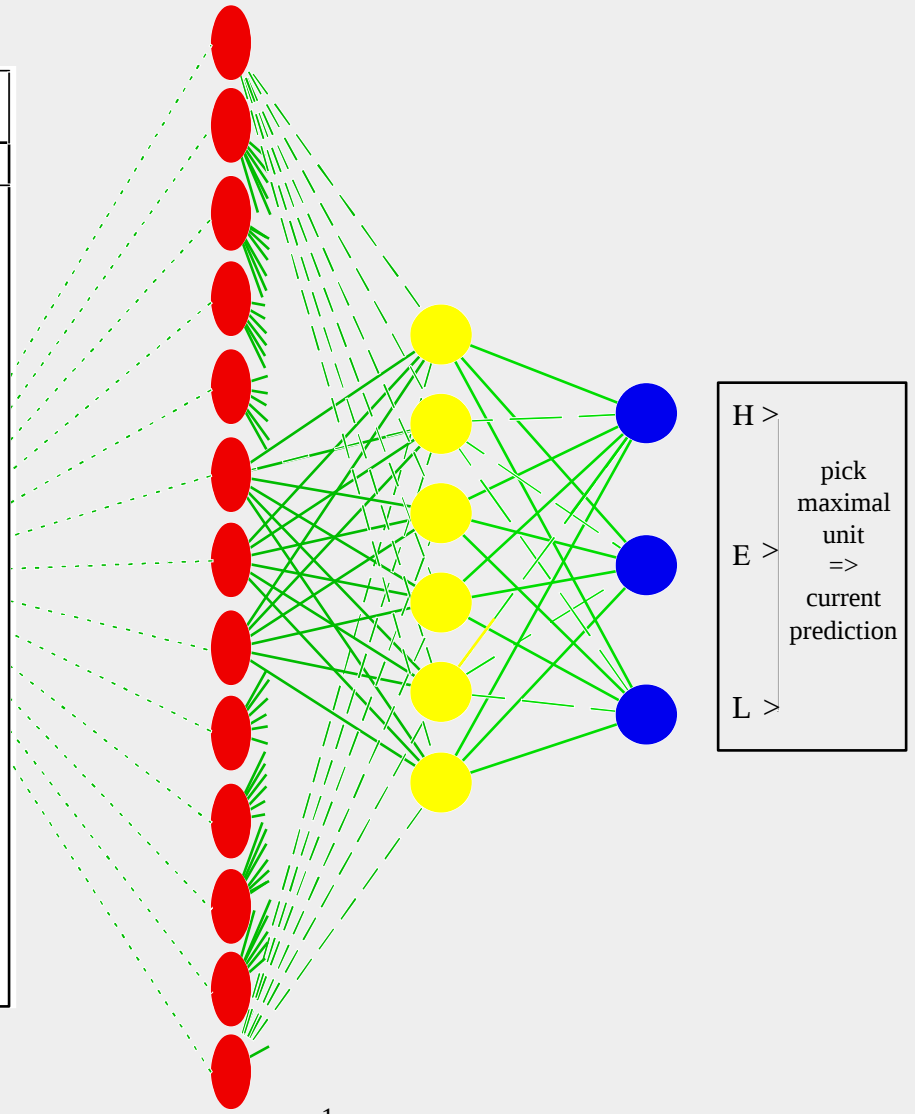
- One input layer.
- Some hidden layers.
- One output layer.



Architecture



Protein	Alignments	profile table
		GSAPD NTEKQ CVHIR LMYFW
:	: : : :	
G	G G G G	5.... ....
Y	Y Y Y Y	....5..
I	I I E E	....2.. ...3. ....
Y	Y Y Y Y	....5..
D	D D D D	....5 ....
P	P P P P	...5. ....
E	A E A A	..3.. ..2.. ....
D	V V E E	....1 ..2.. .2... ....
G	G G G G	5.... ....
D	D D D D	....5 ....
P	P P P P	...5. ....
D	D T D D	....4 .1... ....
D	N Q N N	....1 3...1 ....
G	G N G G	4.... 1.... ....
V	V I V V	.... .4.1. ....
N	E P K K	...1. 1.12. ....
P	P P P P	...5. ....
G	G G G G	5.... ....
T	T T T T	..... .5... ....
D	E K S A	.11.1 ..11. ....
F	F F F F	....5. .....
:	: : : :	



corresponds to the the 21*3 bits coding for the profile of one residue

Secondary Structure Prediction

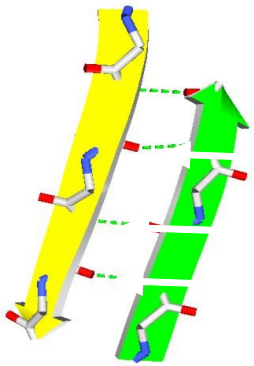
- Application of Multiple sequence alignment
  - Segment based (+8 to -8 residue)
  - Input Multiple alignment instead of single seq uence
  - Application of PSIBLAST
- Current methods (combination of)
  - Segment based
  - Neural network
  - Multiple sequence alignment (PSIBLAST)
  - Combination of Neural Network + Nearest Neighbour Method

PSIPRED

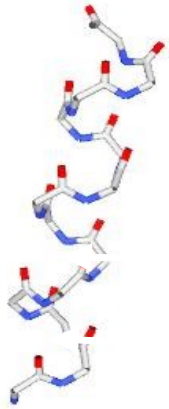
- Uses multiple aligned sequences for prediction.
- Uses training set of folds with known structure.
- Uses a two-stage neural network to predict structure based on position specific scoring matrices generated by PSI-BLAST (Jones, 1999)
  - First network converts a window of 15 aa's into a raw score of h,e (sheet), c (coil) or terminus
  - Second network filters the first output. For example, an output of hhhhehhhh might be converted to hhhhhhhhh.
- Can obtain a Q_3 value of 70-78% (may be the highest achievable)

Secondary structure prediction

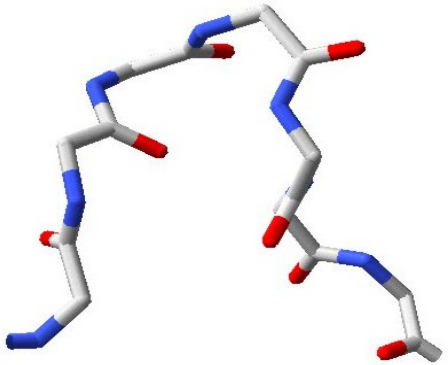
3-state model: Helix (H), Strand (E), Loop (L)



(E) β–Στρανδ, νον–λοχαλ ιντεραχτιονσ



(H) α–Ηελιξ, λοχαλ ιντεραχτιονσ



(Λ) Λοοπ νον–ρεγυλαρ ιντεραχτιονσ

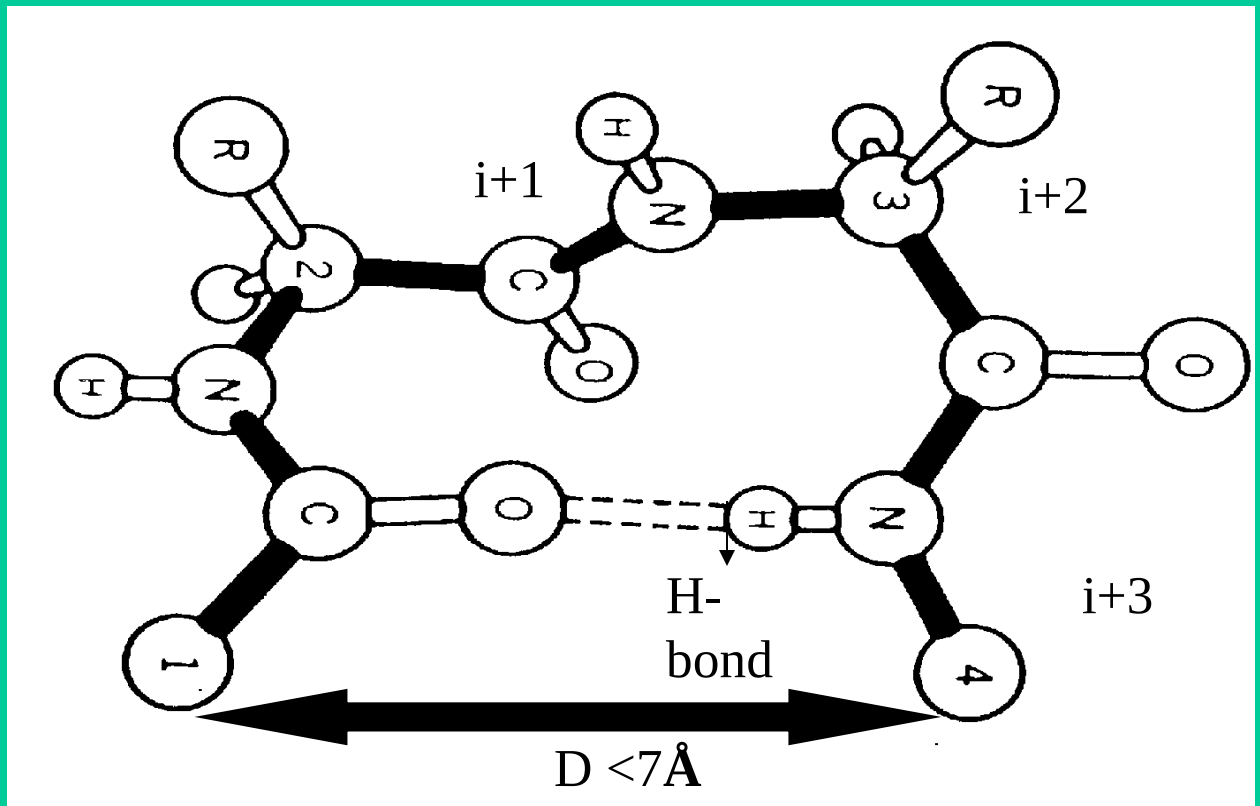
```
SEQ     KELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLD
SS        EEEE    E       E  E    EEEEEE     EEEEEE  EEEEEEHHHEEEE
```

No information about tight turns ?

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)

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

    http://www.imtech.res.in/raghava/betatpred/

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

INPUT DATA

PDB code:	1ah7
Filename:	1ah7.txt
Amino Acid Sequence:	WSAEDKHKEGVNSHLWIVNRAIDIMSRNTTLVKQDRVAQLNEWRTELENG IYAADYENPYYDNSTFASHFYDPDNGKTYIPFAKQAKETGAKYFKLAGESY KNKDMKQAFFYLGLSLHYLGDVNQPMHAANFTNLSYPQGFHSKYENFVDT IKDNYKVTDGNGYWNWKGTNPEEWIHGAAVVAKQDYSGIVNDNTKDWFV KAAVSQEYADKWRAEVTPMTGKRLMDAQRVTAGYIQLWFDTYGDR
Predicted turns/nonturns:	nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnntttttnnnttttnnnnnnnnnnnnnnnnnnnttttnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnttttttnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnn

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

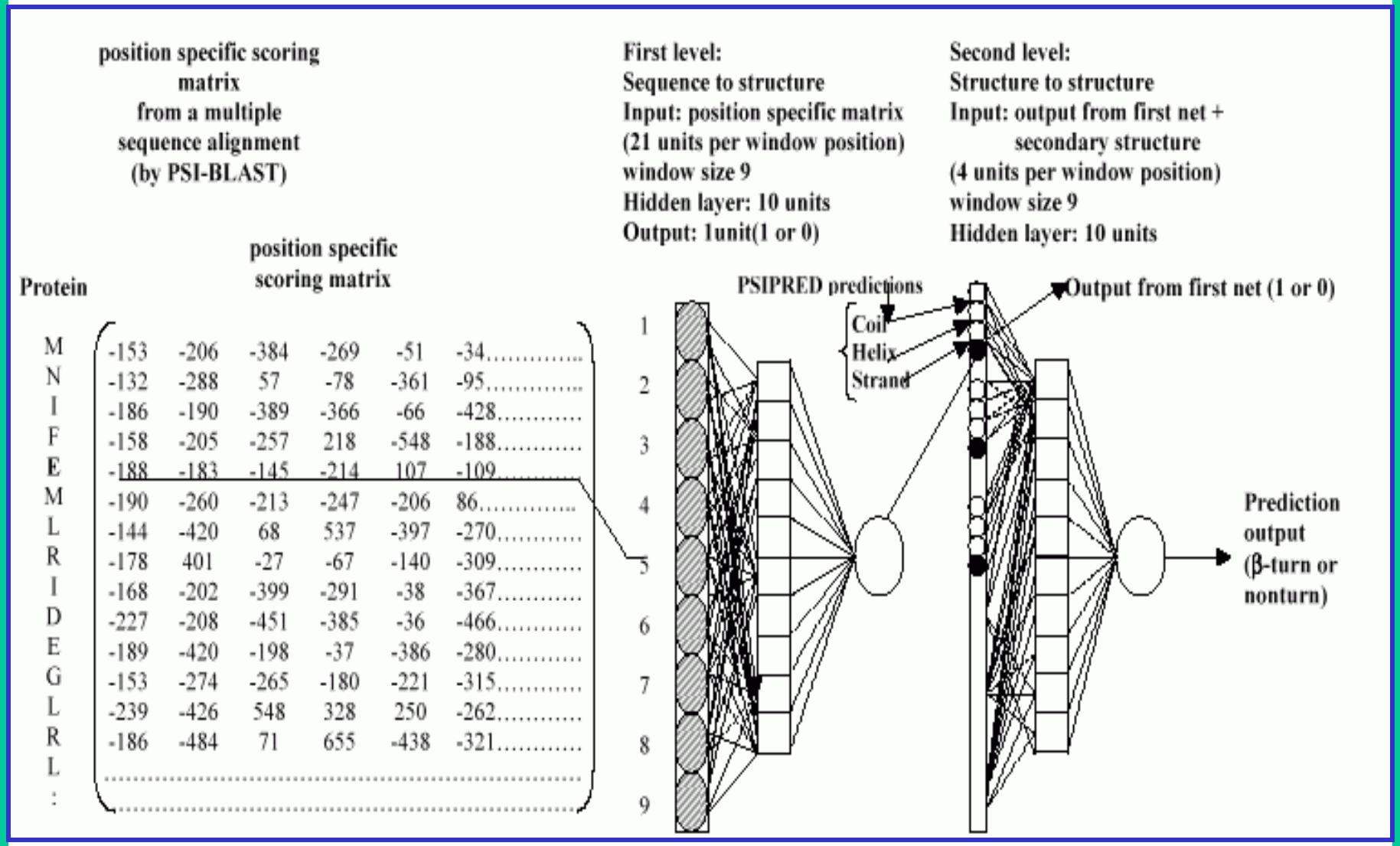
Performance measures of query method plus other existing beta turn prediction methods

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



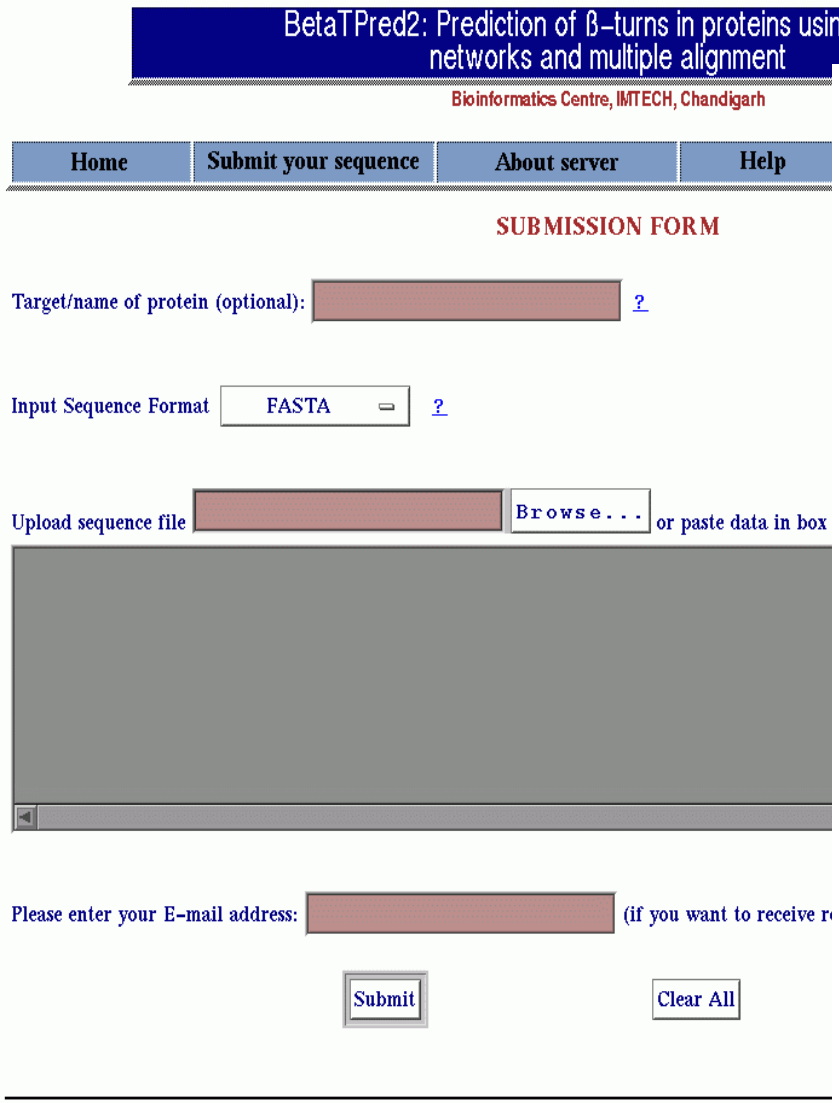
BetaTPred2 prediction results using single sequence and multiple alignment.

	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.

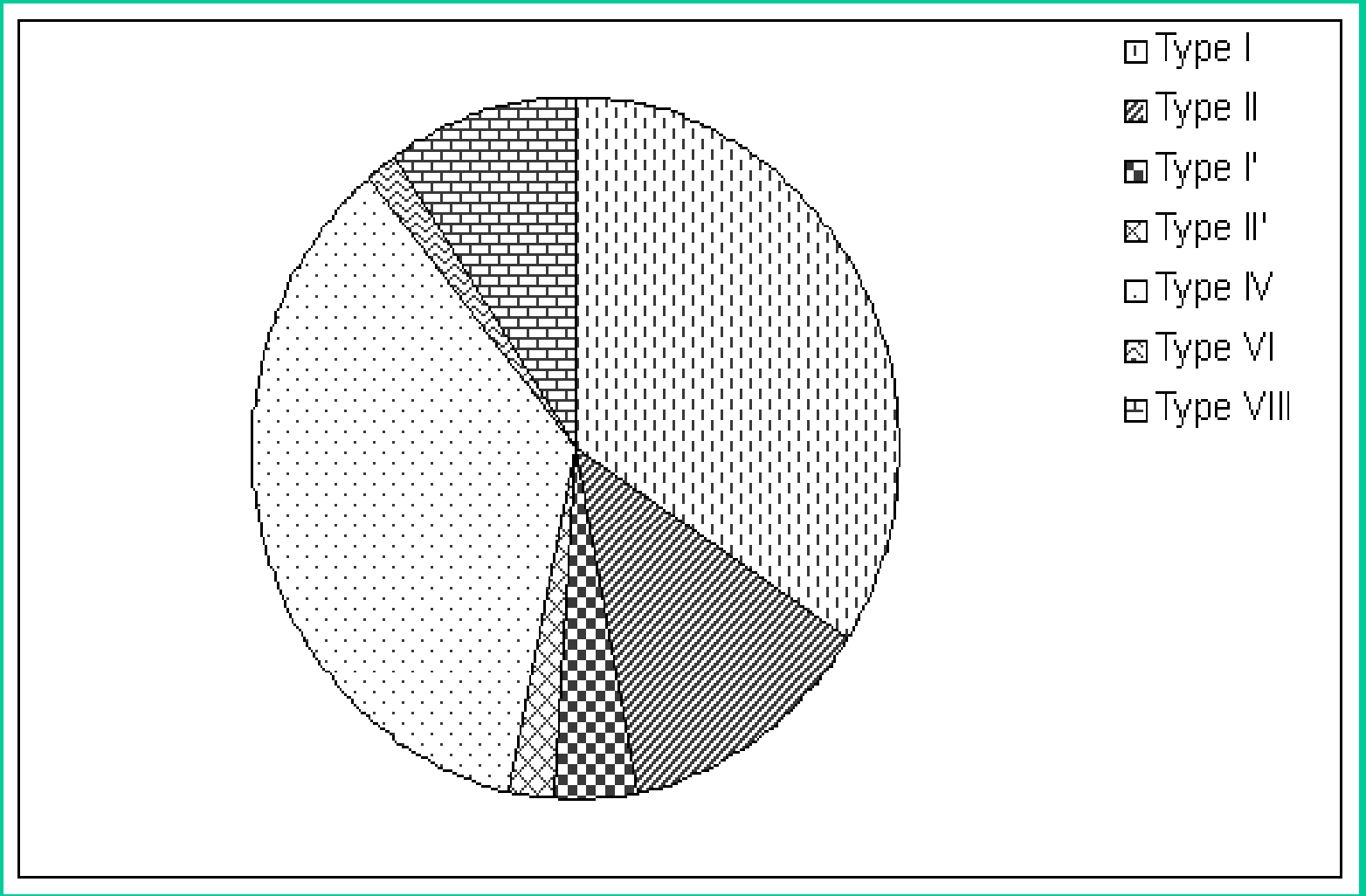
BetaTPred2: A web server for prediction of β-turns in proteins (http://www.imtech.res.in/raghava/betatpred2/)



Beta-turn types

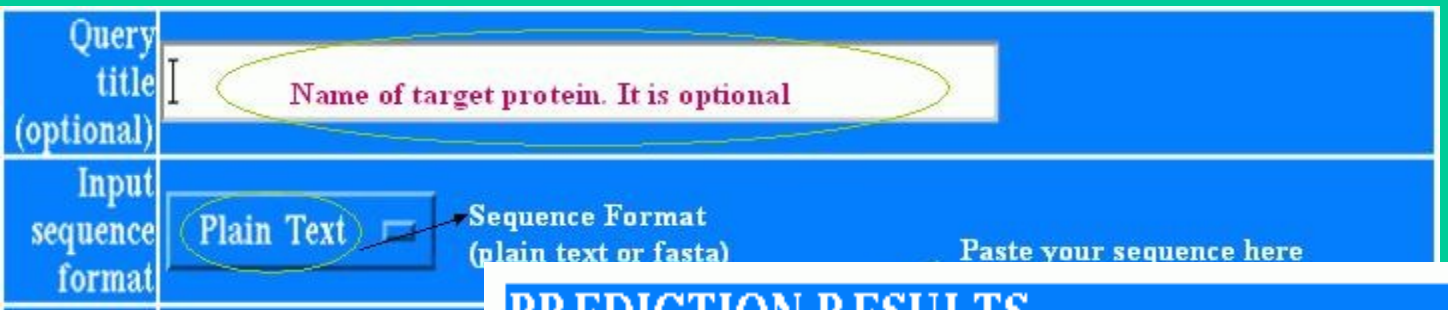
Turn Type	Dihedral angles (°)			
	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
I	-60	-30	-90	0
I'	60	30	90	0
II	-60	120	80	0
II'	60	-120	-80	0
IV	-61	10	-53	17
VIa1	-60	120	-90	0
VIa2	-120	120	-60	0
VIb	-135	135	-75	160
VIII	-60	-30	-120	120

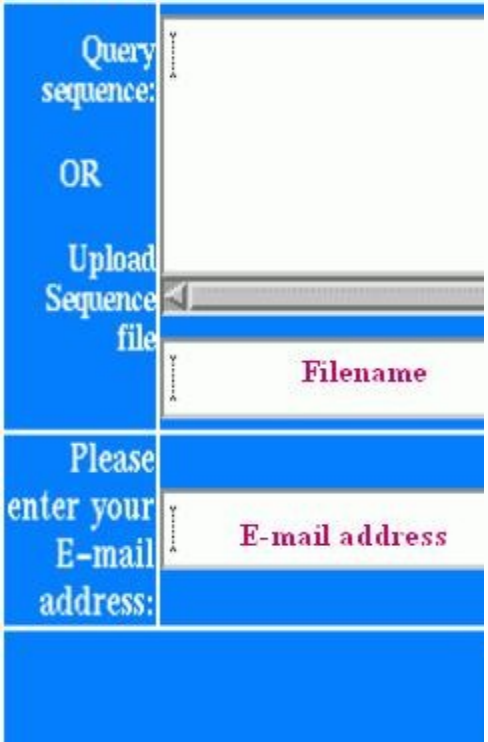
Distribution of β-turn types

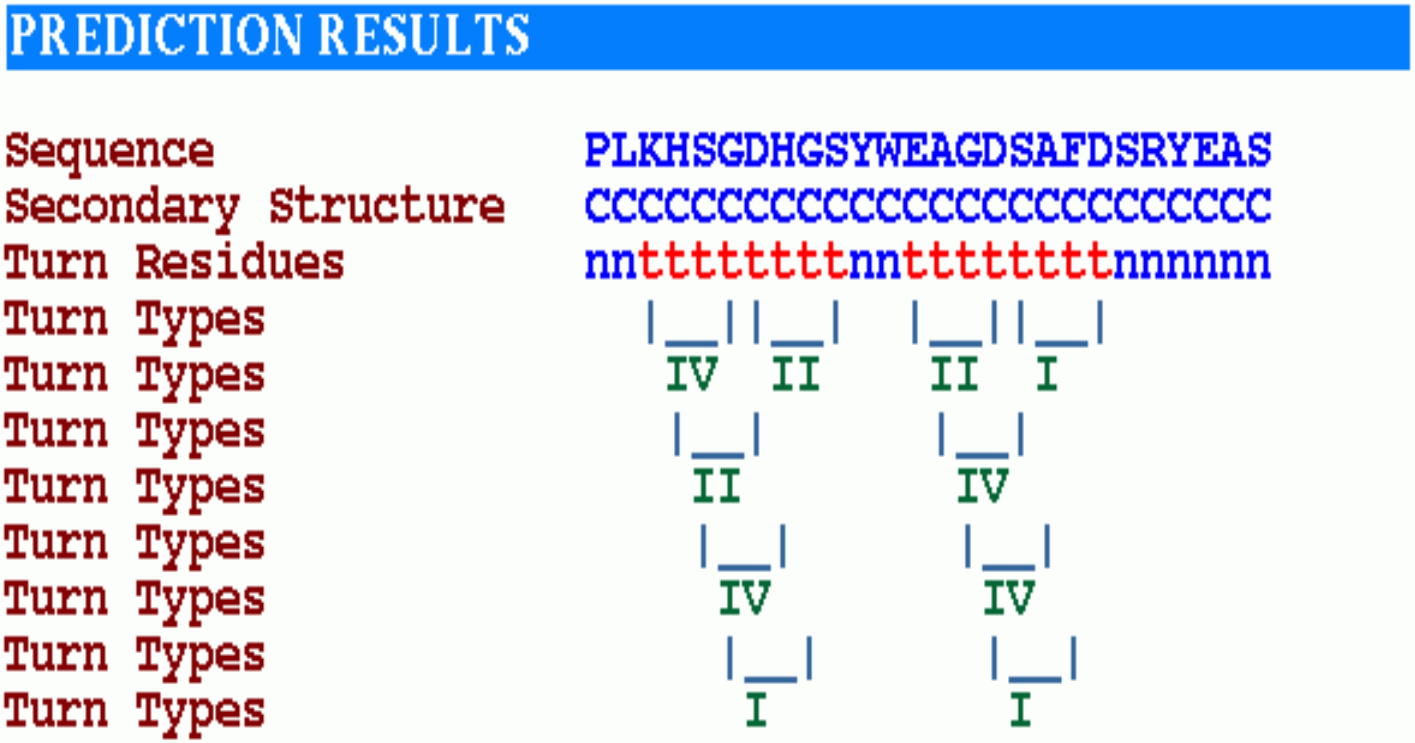


BetaTurns: A web server for prediction of β-turn types

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



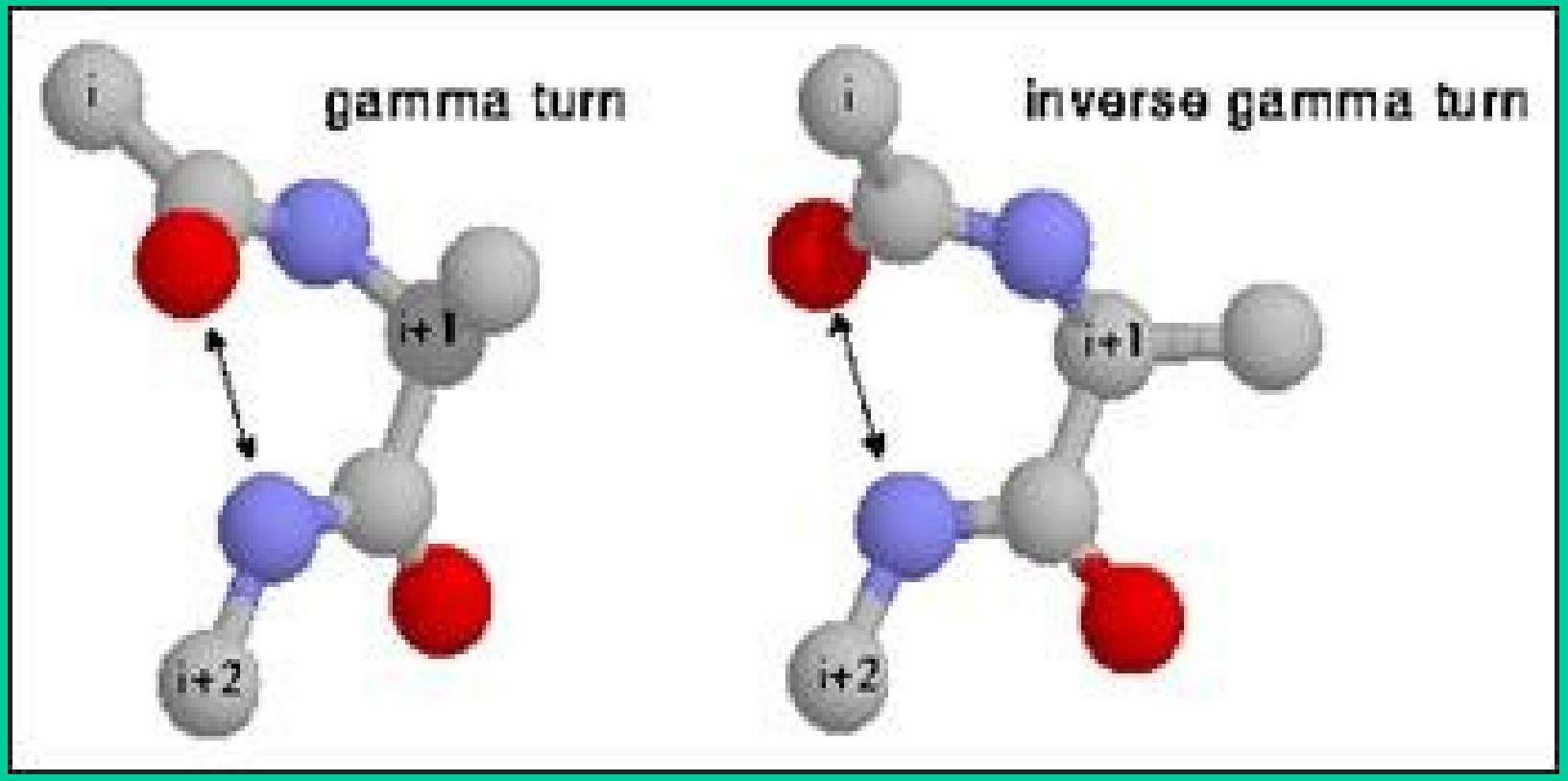




Gamma turns

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

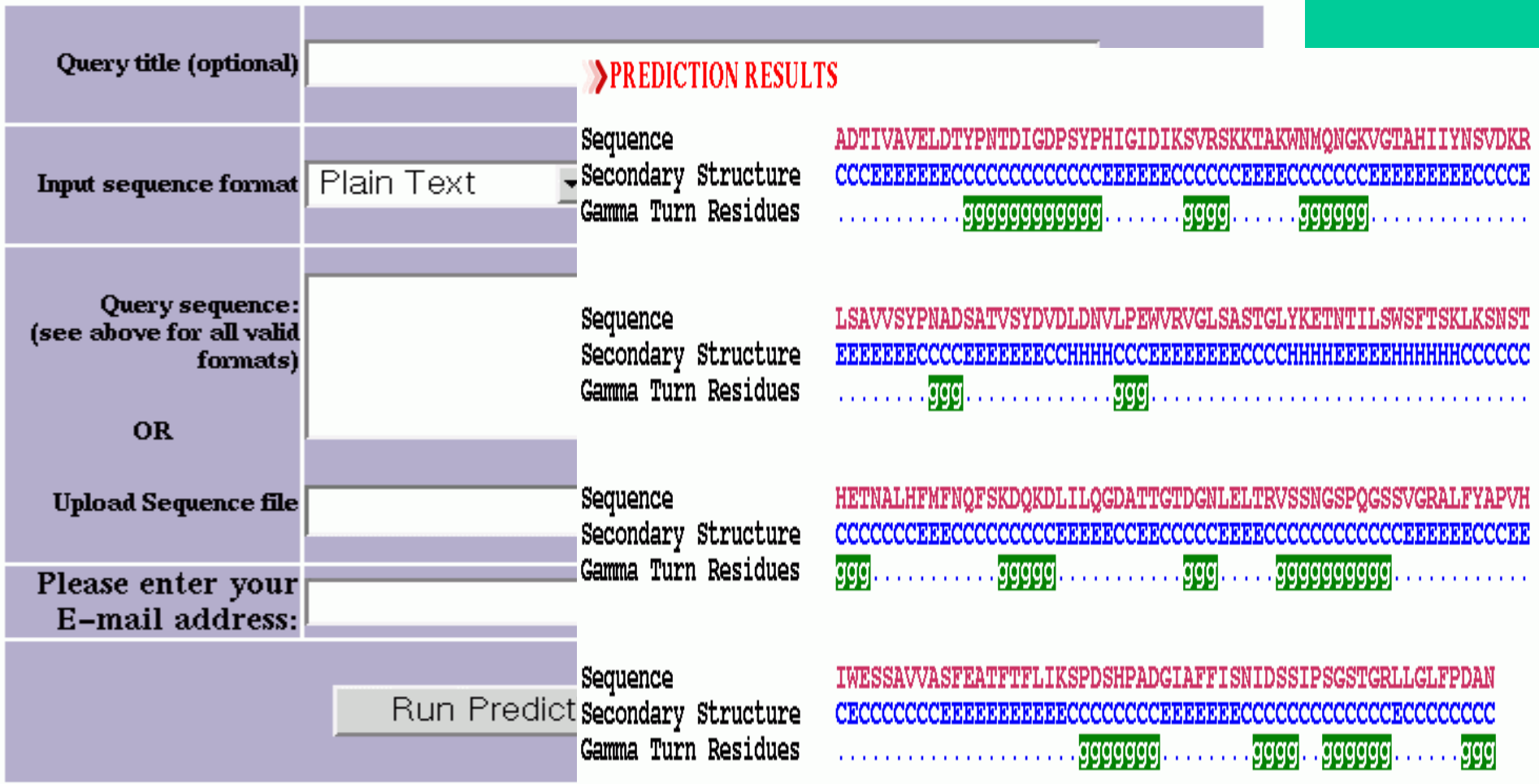
•A γ-turn is defined as 3-residue turn with a hydrogen bond between the Carbonyl oxygen of residue *i* and the hydrogen of the amide group of residue *i*+2. There are 2 types of γ-turns: classic and inverse.



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

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

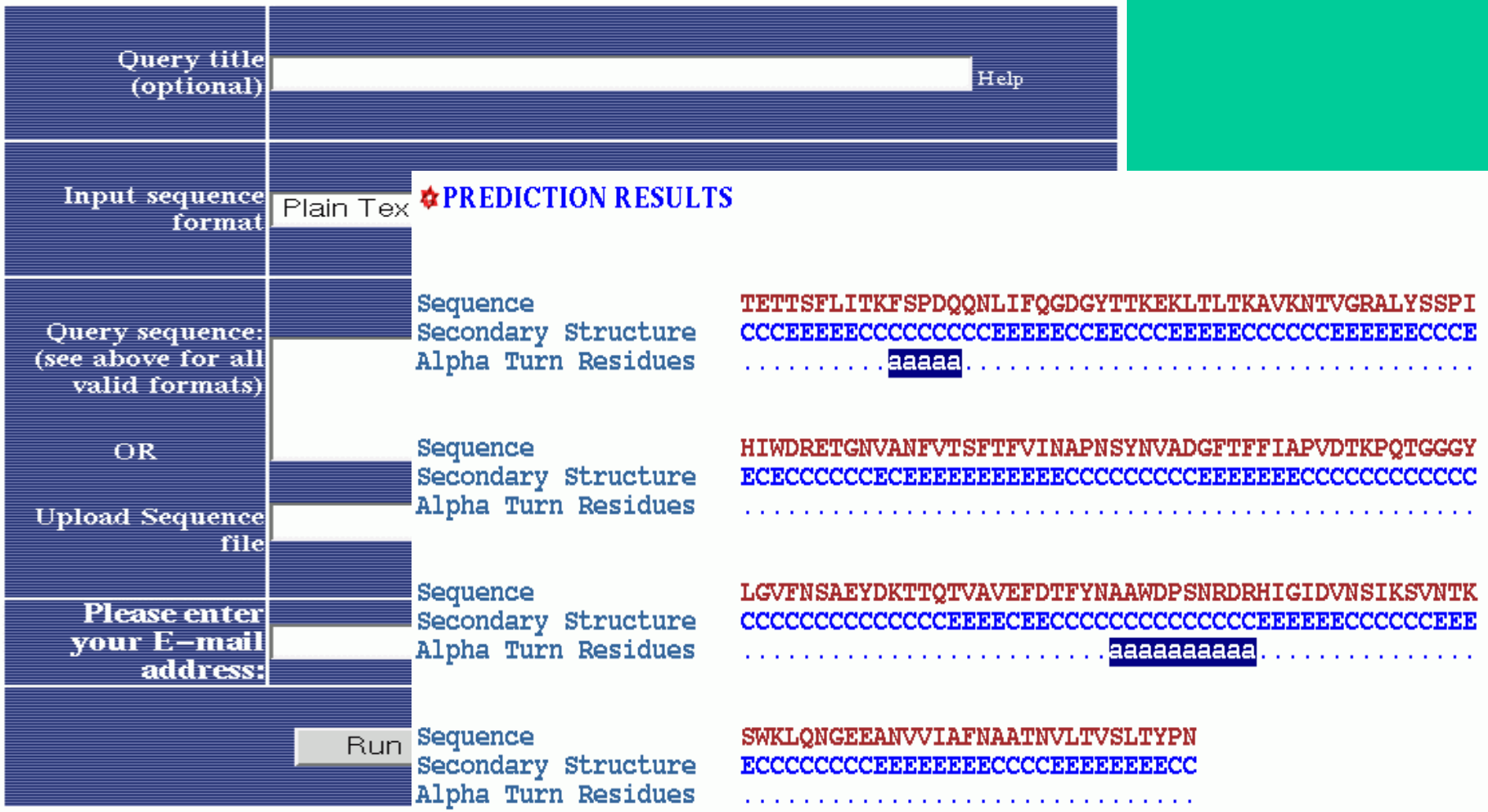
Harpreet Kaur and G P S Raghava (2003) A neural network based method for prediction of γ-turns in proteins from multiple sequence alignment. ***Protein Science*** 12, 923-929.



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* .



Attempt to Predict Tertiary Structure of bioactive Peptides

Contribution of β-turns in tertiary structure

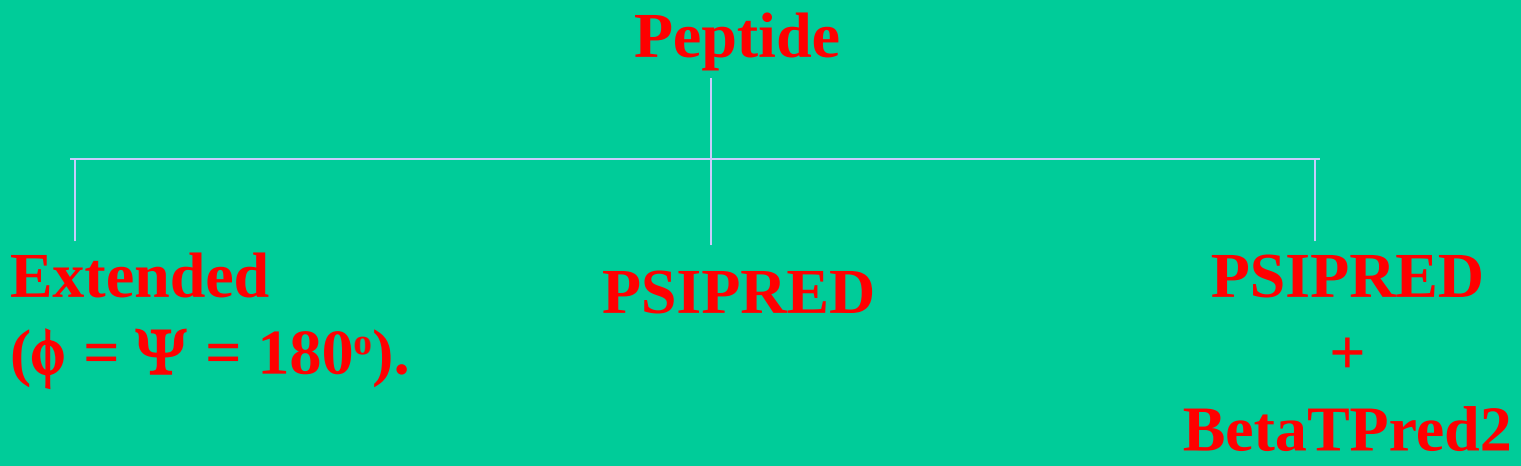
- 3D structures of 77 biologically active peptides have been selected from PDB and other databases
- Bioactive peptides having only natural amino acids and are linear with length varying between 9-20 residues.

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

1) First model has no secondary structure ($\phi = \Psi = 180^{o}$) information
2) Regular Secondary structure states predicted by PSIPRED.
3) Regular and irregular (turns) secondary structure predicted by BetaTPred2.

Steps Involved

- Dihedral angles assigned based on secondary structure
- Side chain angles based on dihedral angles using rotamer library
- Building tertiary structure from internal coordinates
- Energy Minimization using AMBER



Prediction of Tertiary Structure of bioactive Peptides

Root Mean Square Deviation has been calculated…….

PepStr: Prediction of Peptide structure of bioactive peptides

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.

Benchmarking of PepStr

Thomas et al.(2006) Proteins 65:889 compare Robetta, PepLook and Pepstr

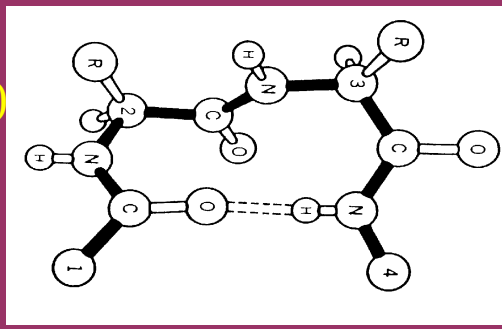
PepLook and Pepstr overperform best method Robetta of protein structure prediction

Structure close to NMR structures

Secondary structure particularly turn predicted better by Pepstr

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.

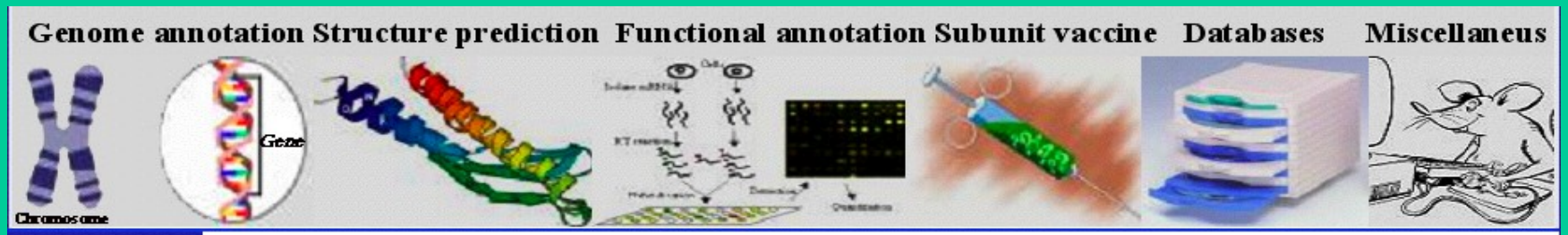


Protein Structure Prediction

- 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



Personal | Research | Web Server | Services offered | Group members

n Jubilee Awards (2009)" Dr. G.P.S. Raghava elected "Fellow of the Indian A

Links: CRDD site ; [Home & profile at Google]; Wikipedia; UAMS Site; Linkedin Site

Webserver

Genome annotation

Structure prediction

Functional annotation

vaccine design

Databases

Miscellaneous

Feedback

Subject	Web Servers Developed by Dr G P S Raghava's group Standalone Package GPSR at IMTECH and at SourceForge					
Protein Structure Prediction	APSSP2	PROCLASS	PSA	RPFOLD	BTEVAL	BetatPred2
	BetatPred	CHpredict	AR_NHPred	TBBPred	BetaTurns	BhairPred
	OXBench	StruComp	GammaPred	AlphaPred	PepStr	SarPred
Target for Subunit Vaccine	PROPRED1	nHLApred	MMBpred	Pcleavage	TAPpred	CTLpred
	PROPRED	HLADR4pred	MHC2pred	MHC	MHCbench	AbAg
	BCEpred	ABCpred	BCIPep	HLApred	MHCBN	HaptenDB
	HLA_affi	FDR4	Tmhcpred			
Genome Annotation	FTG	GWBLAST	GWFASTA	EGPred	SVMgene	SRF
	MyPattern	GeneBench	FTGPred	CDpred	DNAbinder	Polyapred
	Pprint	ECGpred	SGpred			
Functional Annotation	AC2Dgel	NRpred	GPCRpred	GPCRsclass	ESLpred	PSLpred
	BTXpred	Mitpred	SRTpred	Oxypred	VGIchan	HSLpred
	DNAsize	GSTpred	Mango	LGEpred	NTXpred	VICMpred
	ALGpred	PseaPred	RSL-Pred	AntiBP	COPid	siRNAPred
	ESLPred2	ISSpred	CyclinPred	CytoPred	ChemoPred	PFMpred
	TBPred	ProPrint	plPred	PseaPred2	RNAPred	AntiBP2
Databases	MHCBN	BCIPEP	HaptenDB	RFSB	FSGP	PRRDB