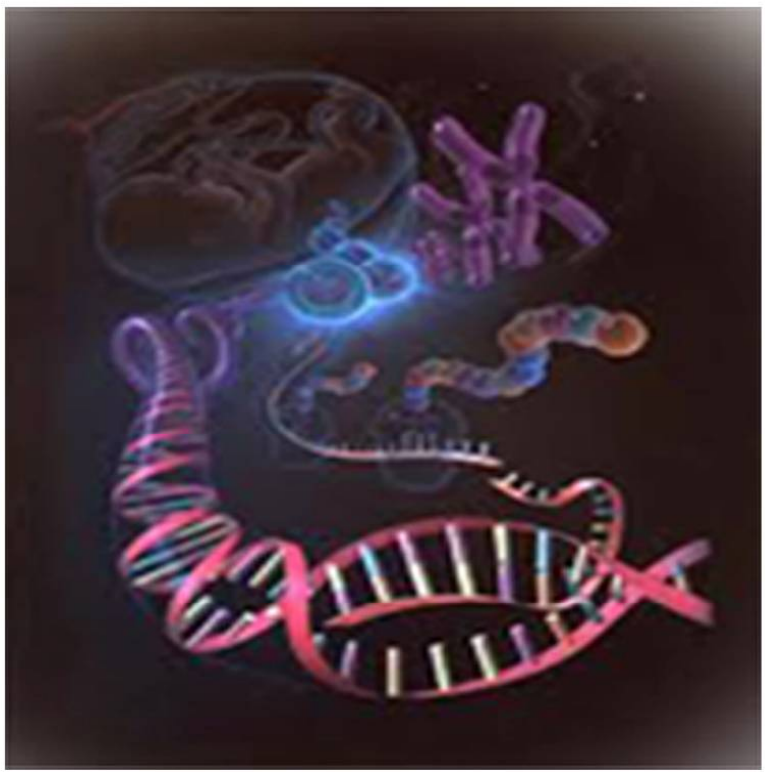
IJLBPR

ISSN 2250 - 3137
Vo1. I, No.1, January 2012

International Journal of

Life Sciences Biotechnology and Pharma Research



International Journal of Life Sciences Biotechnology and Pharma Research
Hyderabad, INDIA

www.ijlbpr.com

Email : editorijlbpr@gmail.com or editor@ijlbpr.com

International Journal of Life Sciences Biotechnology and Pharma Research

ISSN 2250-3137 www.ijlbpr.com
Vol.1, Issue. 1, January 2012

Research Paper

PROTEIN SECONDARY STRUCTURE PREDICTION: AN APPLICATION OF CHOU-FASMAN ALGORITHM IN A HYPOTHETICAL PROTEIN OF SARS VIRUS

DSVGK Kaladhar[1*]

*Corresponding Author: ***DSVGK Kaladhar,*** ✉ *dr.dowluru@gmail.com*

Chou-Fasman algorithm is an empirical algorithm developed for the prediction of protein secondary structure. Implementation and interpretation of the secondary structure of protein has been done using C programming and the output of the result has been predicted good results compared with SOPMA, PSI Pred and Chou-Fasman v1.1 servers. The predicted protein confirmed good accuracy to PSSP results from C programming of the query protein compared to PDB.

Keywords: Chou-Fasman algorithm, C programming, PSSP

INTRODUCTION

The Chou-Fasman is an empirical algorithm (Chou and Fasman, 1978) for the prediction of protein secondary structure originally developed by Robert S. Chao and Gerald D. Fasman in 1978. The method is based on analyses of the relative frequencies of each amino acid in alpha helices, beta sheets, and turns based on known protein structures solved with X-ray crystallography (Nick and Martin, 1998; Avijit and Robert, 1995; Catherine et al., 1994). From these frequencies a set of probability parameters were derived for the appearance of each amino acid in each secondary structure type, and these parameters are used to predict the probability that a given sequence of amino acids would form a helix, a beta strand, or a turn in a protein.

The original Chou-Fasman parameters found some high tendencies among individual amino acids to prefer one type of secondary structure over others (Jack and Russell, 1982). Alanine, glutamate, leucine, and methionine were identified as helix formers, while proline and glycine, due to the unique conformational properties of their peptide bonds, commonly end a helix (Floare *et al.*, 2009).

A protein sequence with amino acids a1a2 a3a4…and is taken as a query sequence. The secondary structure prediction problem is to predict whether each amino acid is in α-helix, a

[1] Department of Bioinformatics, GIS, GITAM University, Visakhapatnam, Andhra Pradesh, 530045, India.

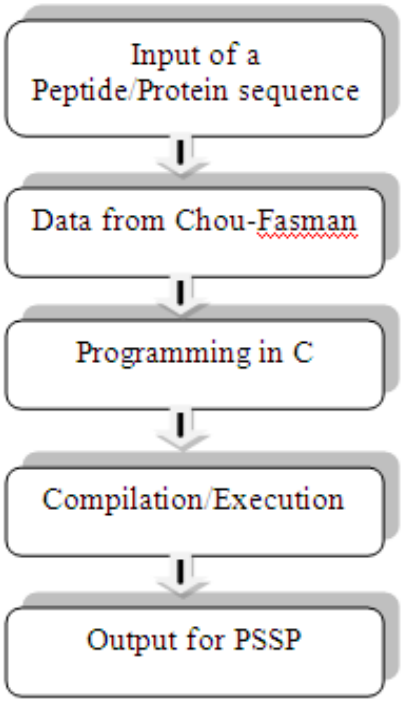
β-sheet, or neither (i.e coil) (Ning and Terrence, 1988). The original Chou-Fasman parameters were derived from a very small and non-representative sample of protein structures that were known at the time of their original work. These original parameters have since been shown to be unpredictable and have been updated from a current dataset, along with implementations to the initial algorithm.

METHODOLOGY

C Programming for PSSP

A program consists of a number of statements, functions, and file handlings etc which are usually executed in sequence. Programs can be much more powerful if we can control the order in which statements are run.

The C programming has been written based on the Chou-Fasman algorithm for the prediction of protein secondary structure.



Chou-Fasman Algorithm

The Chou-Fasman method predicts helices and strands in a similar fashion, first searching linearly through the sequence for a "nucleation" region of high helix or strand probability and then extending the region until a subsequent four-residue window carries a probability of less than 1.

Step 1: Calculate propensities from a set of solved structures. For all 20 amino acids i,

calculate these propensities by:

$$\frac{Pr[i \mid \beta - sheet]}{Pr[i]} \quad \frac{Pr[i \mid \alpha - helix]}{Pr[i]}$$

$$\frac{Pr[i \mid other]}{Pr[i]}$$

Step 2: identify a bend at residue number j,

Step 3: calculate the following value (Table 1): p(t)=f(j)*f(j+1)*f(j+2)*f(j+3)

where f(j), f(j+1), f(j+2) and f(j+3) are bend frequencies in the four positions on the beta turn.

Step 4: If the average value for P(turn)>1.00 in the tetrapeptide where P(turn) is the conformational parameter for ß-turn ; and

Step 5: The averages for the tetrapeptide obey the inequality P(helix)<P(turn)>P(sheet), then a ß-turn is predicted at that location where P(helix) and P(sheet) are the conformational parameters for helix and sheet respectively.

Step 6: If Helex or sheet are not predicted, provide as 'C'. If Helix is predicted, provide as 'H'. If sheet is predicted, provide as 'B'.

SOPMA, PSI PRED and Chou-Fasman v1.1 servers

SOPMA - Self Optimized Prediction Method via ExPASY tools (http://npsa-pbil.ibcp.fr/cgi-bin/

Table 1: Conformational parameters and positional frequencies for helix, β-sheet and β-turn residues

Name	P(a)	P(b)	P(turn)
Alanine	1.42	0.83	0.66
Arginine	0.98	0.93	0.95
Aspartic acid	1.01	0.54	1.46
Asparagine	0.67	0.89	1.56
Cysteine	0.70	1.19	1.19
Glumatic acid	1.51	0.37	0.74
Glutamine	1.11	1.10	0.98
Glycine	0.57	0.75	1.56
Histidine	1.00	0.87	0.95
Isoleucine	1.08	1.60	0.47
Leucine	1.21	1.30	0.59
Lysine	1.14	0.74	1.01
Methionine	1.45	1.05	0.60
Phenylalanine	1.13	1.38	0.60
Proline	0.57	0.55	1.52
Serine	0.77	0.75	1.43
Threonine	0.83	1.19	0.96
Tryptophan	1.08	1.37	0.96
Tyrosine	0.69	1.47	1.14
Valine	1.06	1.70	0.50

Note: P(a), P(b) and P(turn) are conformational parameters of helix, ß-sheet and ß-turns.

npsa_automat.pl?page=npsa_sopma.html), PSIPred v3.0 using low mask complexity regions as filtering options (http://bioinf.cs.ucl.ac.uk/psipred/) and Secondary Structure Prediction by Chou-Fasman, GOR and Neural Network (ver. 1.1) server (http://cib.cf.ocha.ac.jp/bitool/MIX/) are the online tools that predict the secondary structure type for each residue in an amino acid sequence. The servers takes as input a sequence consisting of one-letter amino acid codes (A C D E F G H I K L M N P Q R S T V W Y) (NOTE: B and Z are not recognized as valid amino acid codes) or three-letter amino acid codes separated by spaces (ALA CYS ASP GLU PHE GLY HIS ILE LYS LEU MET ASN PRO GLN ARG SER THR VAL TRP TYR). The output is a secondary structure prediction for each position in the sequence. The predicted type will be either: 'H', a helix element; 'E', or 'B' a beta strand element, or 'C', a turn element.

RESULTS

C programming for protein secondary structure prediction (PSSP) was implemented and written based if chou-fasman algorithm and the comparison of result with protein sequence 9 from SARS genome (NC_004718) is predicted (Table 2). The CDS predicted from 27273 to 27641 with /db_xref="GI:29836501" is a product of SARS hypothetical protein sars7a from Annotated file of NCBI.

SOPMA, PSIPred, and Chou-Fasman server are the online tools that predict the protein secondary structure type for each residue in an amino acid sequence (Table 3, Figure 1 to 3). The predicted protein confirmed to be having good accuracy to PSSP results from C programming of the query protein by comparing with PDB structure (Figure 3). The Translated sequence provided is:

provide/translation="MKIILFLTLIVFTSCELY
HYQECVRGTTVLLKEPCPSGTYEGNSPFHPLADNKFA
LTCTSTHFAFACADGTRHTYQLRARSVSPKLFIRQ
EEVQQELYSPLFLIVAALVFLILCFTIKRKTE"

Table 2: Result from Executed C Program

chou fasman algrithm:
copyright-c:

no	A.A	<pa>	<pb>	<pc>	HELIX	BETA	COIL	PSSP:
1	'M'	1.19	1.25	1.16	'H'	'B'	'C'	'B'
2	'K'	1.13	1.31	1.13	'H'	'B'	'C'	'B'
3	'I'	1.12	1.47	1.19	'H'	'B'	'C'	'B'
4	'I'	1.16	1.39	0.92	'H'	'B'	'-'	'B'
5	'L'	1.10	1.29	0.77	'H'	'B'	'-'	'B'
6	'F'	1.10	1.29	0.77	'H'	'B'	'-'	'B'
7	'L'	1.08	1.35	0.87	'H'	'B'	'-'	'B'
8	'T'	1.04	1.45	0.99	'H'	'B'	'-'	'B'
9	'L'	1.12	1.50	1.05	'H'	'B'	'C'	'B'
10	'I'	1.02	1.47	1.17	'H'	'B'	'C'	'B'
11	'V'	0.95	1.25	1.13	'-'	'B'	'C'	'B'
12	'F'	0.86	1.13	1.13	'-'	'B'	'C'	'B'
13	'T'	0.95	0.88	1.22	'-'	'-'	'C'	'C'
14	'S'	1.05	0.90	1.10	'H'	'-'	'C'	'H'
15	'C'	1.03	1.08	0.87	'H'	'B'	'-'	'B'
16	'E'	1.10	1.00	0.88	'H'	'B'	'-'	'B'
17	'L'	0.90	1.28	0.61	'-'	'B'	'-'	'B'
18	'Y'	0.87	1.23	0.64	'-'	'B'	'-'	'B'
19	'H'	1.08	0.95	0.91	'H'	'-'	'-'	'H'
20	'Y'	1.00	1.03	0.90	'H'	'B'	'-'	'B'
21	'Q'	1.10	1.09	1.02	'H'	'B'	'C'	'B'
22	'E'	1.06	1.05	1.25	'H'	'B'	'C'	'B'
23	'C'	0.83	1.14	1.04	'-'	'B'	'C'	'B'
24	'V'	0.86	1.14	1.04	'-'	'B'	'C'	'B'
25	'R'	0.80	1.02	1.04	'-'	'B'	'C'	'B'
26	'G'	0.82	1.21	0.91	'-'	'B'	'-'	'B'
27	'T'	0.98	1.35	0.84	'-'	'B'	'-'	'B'
28	'T'	1.08	1.37	0.71	'H'	'B'	'-'	'B'
29	'V'	1.16	1.26	0.71	'H'	'B'	'-'	'B'
30	'L'	1.27	0.93	0.86	'H'	'-'	'-'	'H'
31	'L'	1.11	0.74	0.89	'H'	'-'	'-'	'H'
32	'K'	0.99	0.71	1.01	'-'	'-'	'C'	'C'
33	'E'	0.84	0.67	0.93	'-'	'-'	'-'	'C'
34	'P'	0.65	0.76	0.89	'-'	'-'	'-'	'C'
35	'C'	0.65	0.81	0.93	'-'	'-'	'-'	'C'
36	'P'	0.68	0.81	0.93	'-'	'-'	'-'	'C'
37	'S'	0.71	1.04	0.91	'-'	'B'	'-'	'B'
38	'G'	0.90	0.95	0.94	'-'	'-'	'-'	'C'
39	'T'	0.90	0.95	0.94	'-'	'-'	'-'	'C'
40	'Y'	0.86	0.87	0.95	'-'	'-'	'-'	'C'
41	'E'	0.88	0.69	1.18	'-'	'-'	'C'	'C'
42	'G'	0.64	0.73	0.94	'-'	'-'	'-'	'C'
43	'N'	0.78	0.89	1.06	'-'	'-'	'C'	'C'
44	'S'	0.87	0.89	1.05	'-'	'-'	'C'	'C'
45	'P'	0.82	0.84	0.84	'-'	'-'	'-'	'C'
46	'F'	0.98	1.02	0.81	'-'	'B'	'-'	'B'
47	'H'	1.05	0.89	0.68	'H'	'-'	'-'	'H'
48	'P'	1.05	0.80	0.80	'H'	'-'	'-'	'H'
49	'L'	1.08	0.89	0.90	'H'	'-'	'-'	'H'
50	'A'	1.06	0.75	1.02	'H'	'-'	'C'	'H'
51	'D'	0.99	0.89	1.15	'-'	'-'	'C'	'C'
52	'N'	1.10	0.96	0.95	'H'	'-'	'-'	'H'
53	'K'	1.23	1.06	0.82	'H'	'B'	'-'	'B'
54	'F'	1.15	1.17	0.82	'H'	'B'	'-'	'B'
55	'A'	1.04	1.13	0.76	'H'	'B'	'-'	'B'
56	'L'	0.89	1.22	0.83	'-'	'B'	'-'	'B'
57	'T'	0.78	1.08	1.07	'-'	'B'	'C'	'B'
58	'C'	0.78	1.08	1.07	'-'	'B'	'C'	'B'
59	'T'	0.86	1.00	1.08	'-'	'-'	'C'	'C'

Table 2 (Cont.)

no	A.A	<pa>	<pb>	<pc>	HELIX	BETA	COIL	PSSP:
60	'S'	0.93	1.05	1.14	'-'	'B'	'C'	'B'
61	'T'	1.10	1.07	0.95	'H'	'B'	'-'	'B'
62	'H'	1.17	1.12	1.01	'H'	'B'	'C'	'B'
63	'F'	1.27	1.11	0.93	'H'	'B'	'-'	'B'
64	'A'	1.17	1.06	0.87	'H'	'B'	'-'	'B'
65	'F'	1.17	1.06	0.87	'H'	'B'	'-'	'B'
66	'A'	1.14	0.85	0.93	'H'	'-'	'-'	'H'
67	'C'	0.92	0.83	0.95	'-'	'-'	'-'	'C'
68	'A'	0.96	0.83	0.96	'-'	'-'	'-'	'C'
69	'D'	0.85	0.85	1.17	'-'	'-'	'C'	'C'
70	'G'	0.85	0.94	1.05	'-'	'-'	'C'	'C'
71	'T'	0.91	1.05	1.11	'-'	'B'	'C'	'B'
72	'R'	0.88	1.12	0.99	'-'	'B'	'-'	'B'
73	'H'	0.91	1.16	0.76	'-'	'B'	'-'	'B'
74	'T'	0.96	1.27	0.63	'-'	'B'	'-'	'B'
75	'Y'	1.00	1.20	0.77	'-'	'B'	'-'	'B'
76	'Q'	1.18	1.04	0.81	'H'	'B'	'-'	'B'
77	'L'	1.15	1.00	1.04	'H'	'-'	'C'	'H'
78	'R'	1.04	0.86	1.28	'H'	'-'	'C'	'H'
79	'A'	1.06	1.05	1.14	'H'	'B'	'C'	'B'
80	'R'	0.89	1.03	1.33	'-'	'B'	'C'	'B'
81	'S'	0.79	0.94	1.10	'-'	'-'	'C'	'C'
82	'V'	0.89	0.94	0.98	'-'	'-'	'-'	'C'
83	'S'	0.93	0.83	0.86	'-'	'-'	'-'	'C'
84	'P'	1.02	0.99	0.80	'H'	'-'	'-'	'H'
85	'K'	1.14	1.25	1.04	'H'	'B'	'C'	'B'
86	'L'	1.10	1.30	1.18	'H'	'B'	'C'	'B'
87	'F'	1.08	1.25	1.22	'H'	'B'	'C'	'B'
88	'I'	1.17	1.00	1.31	'H'	'-'	'C'	'H'
89	'R'	1.28	0.69	1.31	'H'	'-'	'C'	'H'
90	'Q'	1.30	0.88	1.17	'H'	'-'	'C'	'H'
91	'E'	1.30	0.88	1.17	'H'	'-'	'C'	'H'
92	'E'	1.20	1.07	0.93	'H'	'B'	'-'	'B'
93	'V'	1.20	1.07	0.93	'H'	'B'	'-'	'B'
94	'Q'	1.24	0.97	0.81	'H'	'-'	'-'	'H'
95	'Q'	1.13	1.06	0.78	'H'	'B'	'-'	'B'
96	'E'	1.05	0.97	0.99	'H'	'-'	'-'	'H'
97	'L'	0.81	1.02	0.75	'-'	'B'	'-'	'B'
98	'Y'	0.81	1.02	0.75	'-'	'B'	'-'	'B'
99	'S'	0.92	1.00	0.92	'-'	'-'	'-'	'C'
100	'P'	1.03	1.13	0.68	'H'	'B'	'-'	'B'
101	'L'	1.16	1.39	0.92	'H'	'B'	'-'	'B'
102	'F'	1.12	1.49	1.04	'H'	'B'	'C'	'B'
103	'L'	1.19	1.36	0.91	'H'	'B'	'-'	'B'
104	'I'	1.25	1.24	0.96	'H'	'B'	'-'	'B'
105	'V'	1.28	1.16	0.69	'H'	'B'	'-'	'B'
106	'A'	1.28	1.16	0.69	'H'	'B'	'-'	'B'
107	'A'	1.21	1.30	0.82	'H'	'B'	'-'	'B'
108	'L'	1.15	1.42	0.77	'H'	'B'	'-'	'B'
109	'V'	1.12	1.50	1.05	'H'	'B'	'C'	'B'
110	'F'	1.16	1.39	0.92	'H'	'B'	'-'	'B'
111	'L'	1.05	1.35	0.86	'H'	'B'	'-'	'B'
112	'I'	1.03	1.37	1.04	'H'	'B'	'C'	'B'
113	'L'	0.97	1.26	0.89	'-'	'B'	'-'	'B'
114	'C'	0.93	1.34	1.16	'-'	'B'	'C'	'B'
115	'F'	1.05	1.23	1.16	'H'	'B'	'C'	'B'
116	'T'	1.01	1.12	1.25	'H'	'B'	'C'	'B'
117	'I'	1.10	1.00	1.24	'H'	'B'	'C'	'B'
118	'K'	1.03	0.90	1.10	'H'	'-'	'C'	'H'
119	'R'	1.12	0.81	1.25	'H'	'-'	'C'	'H'
120	'K'	0.88	0.58	0.87	'-'	'-'	'-'	'C'
121	'T'	0.58	0.39	0.63	'-'	'-'	'-'	'C'
122	'E'	0.38	0.09	0.39	'-'	'-'	'-'	'C'

Figure 1: Sequence, Modeled Structure and Secondary Structure Of Gene 9 From SARS Genome

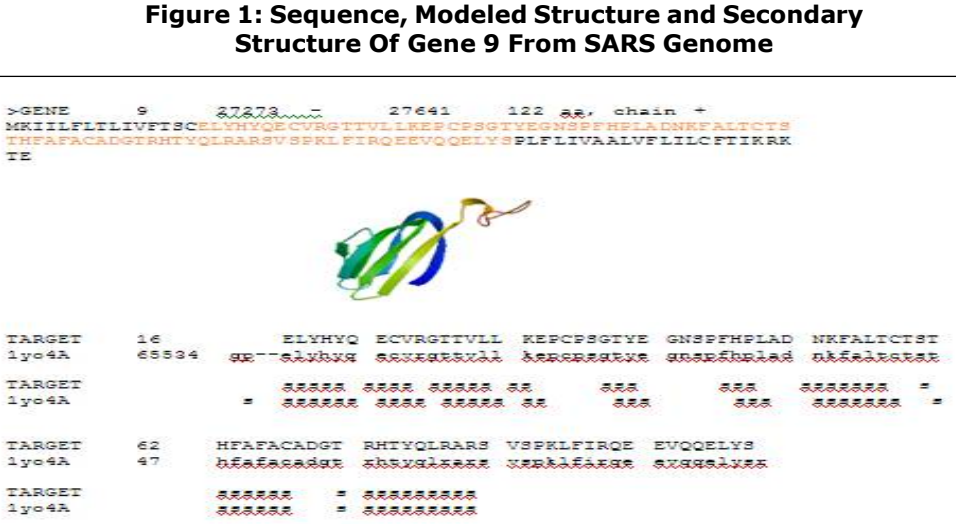


Figure 2: Graph from Chou Fasman Prediction Server

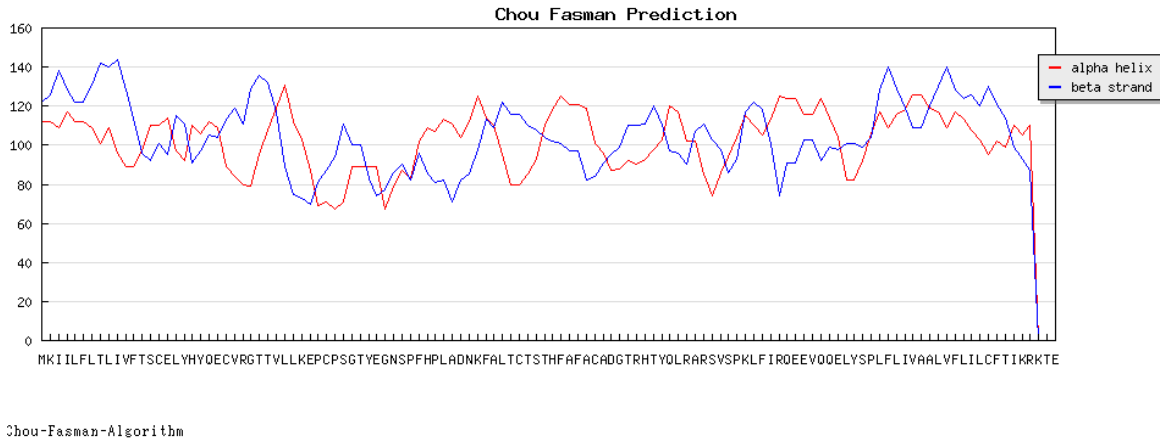


Figure 3: Comparative Result with Other Online Prediction Servers

Protein Secondary Structure prediction

SEQUENCE	MKIILFLTLIVFTSCELYHYQECVRGTTVLLKEPCPSGTYEGNSPFHPLADNKFALTCTS
SOPMA	HHHHHEEEEEEHHHHHHHHHHHHHTTEEEEECCCCCCCCCCCCCCCCCCCTTEEEEECC
PSI PRED	CEEEEEEHHHHHHCCCCCEEEEECCCEEEEECCCCCCCCCCCCCCCCCCCCCCEEEEECC
CHOU-FASMAN	EEEEEEEEEEEECCCCEEHHHHEEEEEEHHHHCCCCEEECCCCCCCHHHHHHHHHHEEEEE
C PROGRAM	BBBBBBBBBBBBCHBBBBHBBBBBBBBBBHHCCCCCBCCCCCCCCBHHHHCHBBBBBBCB
PDB	BBBBBBBBBBBBBBBBB BBBB BBBB BBBBBB BB

SEQUENCE	THFAFACADGTRHTYQLRARSVSPKLFIRQEEVQQELYSPLFLIVAALVFLILCFTIKRKTE
SOPMA	CEEEEEETTTCCEEEEEEHTTCCTTEEECCCCCCHHHCCHHHHHHHHHHHHHHHHCCCTTCH
PSI PRED	CEEEEEEECCCCEEEEEEECCCCCCCCCCHHHHHHHCCCHHHHHHHHHHHHHHHHHEECCCCC
CHOU-FASMAN	HHHHHHHCCCEEEEECCCEEECCCEEEHHHHHHHHHHEECCEEEECCEEEEEEEEECCCCCC
C PROGRAM	BBBBBHCCCCBBBBBBHHBBCCCHBBBHHHHBBHBHBBCBBBBBBBBBBBBBBBBBBHHCCC
PDB	BBBBBBB B BBBBBBBBB

Table 3: Result from Chou Fasman v1.1 Prediction Server

	Pa	Pb	Pt	a	b	<Pa>	<Pb>	<Pt>	<pt>	St	
1	M	132	101	60	H	I	112 *	122 *	63	5.69e-06	E
2	K	121	73	101	h	b	112 *	126 *	63	1.70e-06	E
3	I	99	157	47	i	H	109 *	138 *	53	3.42e-06	E
4	I	99	157	47	i	H	117 *	128 *	56	4.89e-06	E
5	L	130	117	59	H	h	112 *	122 *	68	7.11e-06	E
6	F	111	123	60	h	h	112 *	122 *	68	6.71e-06	E
7	L	130	117	59	H	h	109 *	131 *	65	1.33e-05	E
8	T	78	133	96	i	h	101 *	142 *	63	1.48e-06	E
9	L	130	117	59	H	h	109 *	140 *	54	3.77e-06	E
10	I	99	157	47	i	H	96	144 *	63	1.06e-05	E
11	V	97	164	50	i	H	89	128 *	87	1.75e-05	E
12	F	111	123	60	h	h	89	114 *	104	1.02e-04 *	E
13	T	78	133	96	i	h	97	96	108	8.95e-05 *	C
14	S	72	94	143	b	i	110 *	92	98	3.43e-05	C
15	C	95	107	119	i	h	110 *	101 *	91	4.02e-05	C
16	E	144	51	74	H	B	114 *	95	85	8.62e-06	C
17	L	130	117	59	H	h	97	115 *	95	4.61e-05	E
18	Y	73	131	114	b	h	92	111 *	105	4.31e-05	E
19	H	112	83	95	h	i	110 *	91	95	2.15e-05	H
20	Y	73	131	114	b	h	106 *	97	101	7.92e-05 *	H
21	Q	112	100	98	h	I	112 *	105 *	85	2.75e-05	H
22	E	144	51	74	H	B	109 *	104 *	84	7.06e-06	H
23	C	95	107	119	i	h	89	113 *	105	1.08e-04 *	E
24	V	97	164	50	i	H	84	119 *	99	9.86e-05 *	E
25	R	100	94	95	I	i	80	111 *	110	3.06e-05	E
26	G	64	87	156	B	i	79	129 *	99	3.80e-05	E
27	T	78	133	96	i	h	95	136 *	75	1.82e-05	E
28	T	78	133	96	i	h	108 *	132 *	66	1.04e-05	E
29	V	97	164	50	i	H	119 *	117 *	67	5.30e-06	H
30	L	130	117	59	H	h	131 *	89	73	7.03e-06	H
31	L	130	117	59	H	h	112 *	75	96	3.67e-05	H
32	K	121	73	101	h	b	103 *	73	111	1.44e-05	H
33	E	144	51	74	H	B	87	70	124	1.34e-04 *	C
34	P	55	62	152	B	B	69	81	141	1.95e-05	C
35	C	95	107	119	i	h	71	87	142	8.52e-04 *	C
36	P	55	62	152	B	B	67	94	136	2.13e-04 *	C
37	S	72	94	143	b	i	71	111 *	127	8.29e-05 *	E
38	G	64	87	156	B	i	89	100 *	110	8.04e-05 *	E
39	T	78	133	96	i	h	89	100 *	110	6.54e-05	E
40	Y	73	131	114	b	h	89	83	125	8.51c-05 *	C
41	E	144	51	74	H	B	89	74	132	9.64e-05 *	C
42	G	64	87	156	B	i	67	77	151	7.20e-05	C
43	N	78	66	156	i	b	79	86	127	4.95e-05	C
44	S	72	94	143	b	i	87	90	112	1.27e-04 *	C
45	P	55	62	152	B	B	83	82	114	2.64e-05	C
46	F	111	123	60	h	h	102 *	96	91	6.60e-06	H
47	H	112	83	95	h	i	109 *	85	93	8.80e-05 *	H
48	P	55	62	152	B	B	107 *	81	105	7.23e-06	H
49	L	130	117	59	H	h	113 *	82	106	7.55e-05 *	H
50	A	139	79	66	H	i	111 *	71	117	1.20e-04 *	H
51	D	106	66	146	h	b	104 *	82	115	5.71e-05	H
52	N	78	66	156	i	b	112 *	85	95	6.98e-05	H
53	K	121	73	101	h	b	125 *	98	71	5.52e-06	H
54	F	111	123	60	h	h	114 *	113 *	70	1.28e-05	H
55	A	139	79	66	H	i	110 *	109 *	85	1.25e-05	H
56	L	130	117	59	H	h	95	122 *	92	6.09e-05	E
57	T	78	133	96	i	h	80	116 *	113	3.14e-05	E
58	C	95	107	119	i	h	80	116 *	113	1.59e-04 *	E
59	T	78	133	96	i	h	85	110 *	107	4.20e-05	E
60	S	72	94	143	b	i	93	108 *	98	7.83e-05 *	E
61	T	78	133	96	i	h	110 *	104 *	79	1.52e-05	H
62	H	112	83	95	h	i	118 *	102 *	70	1.31e-05	H
63	F	111	123	60	h	h	125 *	101 *	63	1.69e-05	H
64	A	139	79	66	H	i	121 *	97	77	1.10e-05	H
65	F	111	123	60	h	h	121 *	97	77	3.04e-05	H
66	A	139	79	66	H	i	119 *	82	99	9.02e-06	H
67	C	95	107	119	i	h	101 *	84	121	3.08e-04 *	H
68	A	139	79	66	H	i	96	91	116	9.91e-05 *	C
69	D	106	66	146	h	b	87	95	123	6.90e-05	C

Table 3 (Cont.)

	Pa	Pb	Pt	a	b	<Pa>	<Pb>	<Pt>	<pt>	St	
70	G	64	87	156	B	i	88	99	110	5.89e-05	C
71	T	78	133	96	i	h	92	110 *	95	6.70e-05	E
72	R	100	94	95	I	i	90	110 *	100	2.67e-05	E
73	H	112	83	95	h	i	93	111 *	100	1.69e-04 *	E
74	T	78	133	96	i	h	98	120 *	91	1.45e-05	E
75	Y	73	131	114	b	h	103 *	110 *	91	2.46e-05	E
76	Q	112	100	98	h	I	120 *	97	79	1.06e-05	C
77	L	130	117	59	H	h	117 *	96	78	1.92e-05	C
78	R	100	94	95	I	i	102 *	90	99	5.58e-05	C
79	A	139	79	66	H	i	102 *	107 *	88	4.21e-05	E
80	R	100	94	95	I	i	85	111 *	107	2.89e-05	E
81	S	72	94	143	b	i	74	103 *	122	4.90e-05	E
82	V	97	164	50	i	H	86	98	111	2.78e-05	C
83	S	72	94	143	b	i	94	86	113	1.82e-04 *	C
84	P	55	62	152	B	B	104 *	93	93	2.74e-05	C
85	K	121	73	101	h	b	115 *	117 *	66	5.01e-06	E
86	L	130	117	59	H	h	110 *	122 *	65	2.76e-06	E
87	F	111	123	60	h	h	105 *	118 *	75	1.95e-05	E
88	I	99	157	47	i	H	113 *	100 *	78	1.08e-05	H
89	R	100	94	95	I	i	125 *	74	85	3.38e-05	H
90	Q	112	100	98	h	I	124 *	91	74	1.81e-05	H
91	E	144	51	74	H	B	124 *	91	74	9.22e-06	H
92	E	144	51	74	H	B	116 *	103 *	80	9.75e-06	H
93	V	97	164	50	i	H	116 *	103 *	80	1.44e-05	H
94	Q	112	100	98	h	I	124 *	92	82	3.91e-05	H
95	Q	112	100	98	h	I	114 *	99	86	2.00e-05	H
96	E	144	51	74	H	B	104 *	98	97	1.69e-05	H
97	L	130	117	59	H	h	82	101 *	117	3.37e-05	E
98	Y	73	131	114	b	h	82	101 *	117	2.71e-05	E
99	S	72	94	143	b	i	92	99	103	8.45e-05 *	C
100	P	55	62	152	B	B	106 *	104 *	82	1.16e-05	C
101	L	130	117	59	H	h	117 *	128 *	56	5.04e-06	E
102	F	111	123	60	h	h	109 *	140 *	54	1.02e-06	E
103	L	130	117	59	H	h	116 *	129 *	55	3.37e-06	E
104	I	99	157	47	i	H	118 *	119 *	57	4.19e-06	E
105	V	97	164	50	i	H	126 *	109 *	60	1.15e-05	C
106	A	139	79	66	H	i	126 *	109 *	60	8.70e-06	C
107	A	139	79	66	H	i	119 *	120 *	58	2.73e-06	E
108	L	130	117	59	H	h	117 *	130 *	57	1.33e-05	E
109	V	97	164	50	i	H	109 *	140 *	54	5.12e-06	E
110	F	111	123	60	h	h	117 *	128 *	56	1.34e-06	E
111	L	130	117	59	H	h	113 *	124 *	71	9.56e-06	E
112	I	99	157	47	i	H	108 *	126 *	71	8.18e-06	E
113	L	130	117	59	H	h	103 *	120 *	83	1.66e-05	E
114	C	95	107	119	i	h	95	130 *	80	2.22e-05	E
115	F	111	123	60	h	h	102 *	121 *	76	7.87e-06	E
116	T	78	133	96	i	h	99	114 *	84	1.79e-05	E
117	I	99	157	47	i	H	110 *	99	86	4.65e-05	C
118	K	121	73	101	h	b	105 *	93	98	3.32e-05	C
119	R	100	94	95	I	i	110 *	87	91	3.35e-05	C
120	K	121	73	101	h	b	0	0	0	0.00e+00	C
121	T	78	133	96	i	h	0	0	0	0.00e+00	C
122	E	144	51	74	H	B	0	0	0	0.00e+00	C

DISCUSSION

Before any X-ray or NMR structure was known for the family, the prediction of protein secondary structure from an aligned family of proteins has been highlighted by several accurate predictions. New computational techniques that apply Artificial intelligence machine learning and discriminate analysis show promise as alternatives to neural networks (Geoffrey, 1995).

Successful secondary structure prediction provides a starting point for direct tertiary structure modeling and provides necessary information for protein folding resides completely within the primary structure. Although the development of

advanced molecular biology laboratory techniques such as X-ray crystallography and NMR *in silico* prediction methods will narrow the gap between available sequences and structures (Nageswara et al., 2010).

Methods for protein secondary structure prediction provide information that is useful both in *ab initio* structure prediction and as additional restraints for fold recognition algorithms. Many approaches have been devised for predicting the secondary structure from the protein sequence such simple linear statistics, evolutionary trees, physicochemical properties, linear discrimination, machine learning, neural networks, k-way nearest neighbors, simple residue substitution matrices and combinations of different methods with consensus approaches (James and Geoffrey, 2000).

CONCLUSION

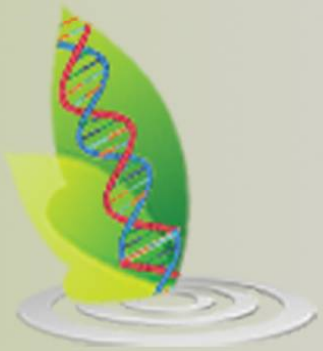
The C program predicted good accuracy compared with SOPMA, PSI PRED and Chou-Fasman v1.1 servers. Further implementation for the prediction of three dimensional structures of the proteins should be done.

ACKNOWLEDGMENT

The author would like to thank GITAM University for providing lab facility and access to e-journals to carry out the research.

REFERENCES

1. Avijit C and Robert L B (1995), "Stability of á-Helices", *Advances in Protein Chemistry*, Vol. 46, pp. 141-176.
2. Catherine K S, Jane M W and Lynne R (1994), "A Thermodynamic Scale for the beta-Sheet Forming Tendencies of the Amino Acids", *Biochemistry*, Vol. 33, pp. 5510-5517.
3. Chou P Y and Fasman G D (1978), "Empirical Predictions of Protein Conformation", *Annual Review of Biochemistry*, Vol. 47, pp. 251-276 .
4. Floare C G, Bogdan M, Horovitz O, Mocanu A and Tomoaia-Cotisel M (2009), "Analysis of the Secondary Structure of a Protein's N-Terminal", *J. Phys., Conf. Ser.*, Vol. 182, pp. 012008.
5. Geoffrey J B (1995), "Protein Secondary Structure Prediction", *Current Opinion in Structural Biology*, Vol. 5, pp. 372-376.
6. Jack K and Russell F D (1982), "A Simple Method For Displaying The Hydropathic Character Of A Protein", *Journal of Molecular Biology*, Vol. 157, pp. 105-132.
7. James A C and Geoffrey J B (2000), "Application of Multiple Sequence Alignment Profiles to Improve Protein Secondary Structure Prediction", *PROTEINS: Structure, Function, and Genetics*, Vol. 40, pp. 502-511.
8. Nageswara R P V, Uma D T, Kaladhar D S V G K, Sridhar G R and Allam A R (2010), "Protein Secondary Structure Prediction Using Pattern Recognition Neural Network", *International Journal of Engineering Science and Technology*, Vol. 2, pp. 1752-1757.
9. Nick C P and Martin J S (1998), "A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins", *Biophysical Journal*, Vol. 75, pp. 422-427.
10. Ning Q and Terrence J S (1988), "Predicting the Secondary Structure of Globular Proteins Using Neural Network Models", *Journal of Molecular Biology*, Vol. 202, pp. 865-884.



International Journal Life Sciences Biotechnology and Pharma Research
Hyderabad, INDIA. Ph: +91-09441351700, 09059645577
E-mail: editorijlbpr@gmail.com or editor@ijlbpr.com
Website: www.ijlbpr.com