Advances in Bioresearch
Adv. Biores., Vol 13 (3) May 2022: 174-181
©2022 Society of Education, India
Print ISSN 0976-4585; Online ISSN 2277-1573
Journal's URL:http://www.soeagra.com/abr.html
CODEN: ABRDC3
DOI: 10.15515/abr.0976-4585.13.3.174181

Advances in Bioresearch

ORIGINAL ARTICLE

Template Selection for CASP Targets: Role of Profile Based Alignment and Physico-Chemical Property Conservation

Konda Mani Saravanan and Kishore Kumar Meenakshi Sundaram*
Research and Publication Wing, Bharath Institute of Higher Education and Research, Chennai – 600 073, Tamilnadu, India.
Corresponding Email author: marinemeenasundar@bharathuniv.ac.in

ABSTRACT
For the past four decades, extensive research has been carried out for the development of the methods of protein structure prediction. Development of hybrid methods and novel algorithms including different parameters has contributed to a large extent for the progress of protein structure prediction. In spite of the development of several structure prediction methods with better accuracy, it is not still clear how the one-dimensional amino acid sequence of a protein codes for the three-dimensional structure. The high complex nature of sequence-structure relationship is due to the interplay between physics and evolution and hence the problem must be viewed from a physico-chemical perspective. Further, one of the most important factors influencing the ability to predict accurate models is the extent of structural conservation between target and template. Considering the above facts, we have made a retrospective analysis of the earlier CASP targets and their templates by using physico-chemical properties correlation coefficient of the amino acid residues which were utilized by Argos (1987) in his sensitive sequence comparison algorithm. For most of the targets and templates in all four structural classes, a reasonable correlation coefficient is observed for any one of the five properties. Also, the profile based alignment between target and template was better than the substitution matrix based alignment. The results discussed here point to the need for the development of novel algorithms by incorporating profile alignment and physico-chemical correlation coefficients to select the template or a fold from fold library during the comparative modeling or threading procedures.
Keywords: *Profile alignment; Physico-chemical properties; secondary structure prediction; CASP targets and templates; Protein Folding*

Received 20.03.2022 Revised 11.04.2022 Accepted 14.05.2022

How to cite this article:
K M Saravanan and K K M Sundaram. Template Selection for CASP Targets: Role of Profile Based Alignment and Physico-Chemical Property Conservation, Saudi Arabia. Adv. Biores. Vol 13 [3] May 2022. 174-181

INTRODUCTION

The number of protein sequences in databases has increased in an exponential manner in the recent past decades due to the advances in genome sequencing technology. However, the majority of sequences are not provided with functional information because of the slow and expensive nature of the experimental procedures for structure determination. One way to tackle this problem is to compare the sequences with those of known protein structures available in Protein Data Bank [1]. The main concern of the structural biologists at present is to bridge the gap between sequence and structure knowledge, often termed as sequence-structure gap. It is the main factor driving the need for prediction of protein structure.

If one tries to assign two dihedrals with two possible positions for each dihedral for a given protein sequence with 100 residues, there will be 2^{200} conformations. Since the conformational space of even a very small protein is considerably large, the process of finding the correct one by a random search is not possible. However according to Anfinsen, proteins can fold to their native structures spontaneously without the intervention of any agent and therefore the protein fold is coded in the amino acid sequence itself [2]. Protein structure prediction is therefore a problem of much scientific interest and it is not still clear as to how structure is encoded in sequence.

Computer methods for protein analysis address this problem since they study the relationship between the amino acids sequence and structure. Since proteins have structural features which define functional

similarities, the need for structure estimation methods is high. Developments of novel algorithms including different parameters have contributed a large extent for the progress of protein structure prediction. Due to significant developments in the structure prediction field, processing and consequent analysis of predicted structures has become as a complex procedure [3].

Sequence similarity search is a crucial step in analyzing newly determined sequences. However, sequences of homologous proteins can diverge and there are most sensitive methods available to find the homologues. Modern secondary structure prediction methods utilize evolutionary information derived from multiple sequence alignment to provide better insights into the positional conservation of physico-chemical features such as hydrophobicity and hints at position of loops in the regions of insertions and deletions corresponding to gaps in the alignment [4]. The secondary structure prediction algorithm with different types of multiple sequence alignment profiles derived from the homologous sequences is shown to provide better accuracy than other alignment methods based on substitution matrices [5]. There are several consensus meta-servers such as the NPS web server [6] and JPred server [7] that returns predictions from several secondary structure prediction methods and provide a consensus secondary structure using a neural network, thereby improving the average accuracy of prediction. In general, most of the secondary structure prediction methods predict the secondary structures of all-alpha proteins more accurately than other classes [8].

In our earlier work, we used a structural descriptor known as Long Range Order (LRO) [9] to characterize the native fold of the homologous family of proteins [10]. In a recent work, we have shown that the presence of unusual combination of amino acid residues in CASP8 targets (T0498 and T0499) at the secondary structural element positions may lead the secondary structure prediction methods not to predict the structural states accurately [11]. Further, we have shown that the influence of certain biophysical properties such as hydrophobic residues, hydrophilic residues, difference in secondary structural propensities of surrounding residues and difference in cooperative long range interactions in identical octapeptides adopting different conformations [12].

CASP is a community wide structure prediction experiment that runs every two years to assess the quality of structure prediction methods developed by various research groups. The CASP experiment challenges prediction teams to submit structural models for a set of sequences whose structures have been recently solved experimentally but have not yet been published. From the series of CASP experiments, it is known that a correct protein fold prediction can be obtained by one method but not by the others [13]. It has also been observed that no method can reliably distinguish between weak hits (beyond a threshold score) and wrong hits and that often a correct model is found among the top hits of the method. From such and other observations many human expert predictors realized that in order to produce better predictions, the results from a number of independent methods need to be analyzed. Another successful practice observed in previous CASP was to build hybrid models from fragments. Automated meta-predictors using this approach have also been developed.

There are several lessons from previous CASP experiments such as the need for an analytical approach to find that what is the success and failure behind a prediction method and to identify which prediction method has greater accuracy and what are all the significant parameters which made the method to predict the structure successfully. Pair wise sequence identity between target and template is not an effective parameter for describing the difficulty of a target. One of the most important factors influencing the ability to predict accurate models is the extent of structural conservation between target and template. Hence in the present work, we have made a retrospective analysis of previous CASP targets and their templates by using alignment programs and five physico-chemical properties used by Argos [14] in sequence comparison algorithm.

MATERIAL AND METHODS

DATA SET

The targets released in CASP competitions from CASP1 to CASP8 form the source of our present study. The target sequence was subjected to BLAST search[15] against PDB database [1] and the resulting hit with 100% identity PDB ID was assigned to the particular target. From the available targets, we have grouped the 66 fold recognition/comparative modelling category targets for our analysis. We have also obtained the templates for 66 targets grouped according to their structural class which is used to recognize the fold from the Prediction center website. Among several templates for a particular target, the template which gave good model has been considered.

ALIGNMENT OF TARGETS AND TEMPLATES

The targets and templates were subjected to alignment by using STRETCHER [16] and Fold and Function Alignment Server (FFAS) [17]. Stretcher calculates an optimal global alignment between target and

template by using a modification of the classic dynamic programming algorithm which uses linear space where as FFAS server utilizes information present in sequences of homologous proteins and performs profile-profile alignment. Aligned regions without gaps were considered for further computations.

COMPUTATION OF DAYHOFF/BLOSUM SCORE

To evaluate the extent of homology between CASP targets and their corresponding templates, a classical measure is to compute scores based on substitution matrices such as Dayhoff's PAM 250 mutation matrix [18]. This matrix expresses the relative weight with which each amino acid is replaced by another amino acid. Amino acids replace each other depending on their chemical similarity; for example a charged residue such as an aspartic acid will be replaced by glutamic acid with a similar charge and so on. We have made an analysis of aligned sequence of 66 pairs of targets and templates by using the BLOSUM 62 mutation data matrix [19]. When the pair of target and template sequence was compared, for each position the BLOSUM score was obtained and summed up. The pair with highest score will be closely related to each other.

COMPUTATION OF PHYSICO-CHEMICAL CORRELATION COEFFICIENT

We have made use of five kinds of physico-chemical properties of amino acid residues namely surrounding hydrophobicity [20], bulkiness [21], turn preference [22], antiparallel strand preference [23] and refractivity index [21] which were utilized by Argos [14] in his sequence comparison algorithm. The original values provided by different authors were normalized to a value of 1.0 and made positive by the addition of the most negative value [24]. These selective parameters according to Argos [14] are highly sensitive for structurally aligned amino acid residues. Hence we computed cross correlation coefficient by substituting sequence of numerical values which represents any one of the above physical or chemical property in the place of amino acid sequence of target and template sequences. Calculations of average correlation coefficients using a set of properties were found to improve the signal noise ratio and in our calculation average cross correlation coefficient were also computed [25].

A quantitative expression of homology between two amino acid sequences X and Y is obtained by computation of cross correlation coefficient described below. The coefficient C(j) at the jth residue of the sequence Y is expressed by comparing a sequence of N residues long, which starts at the uth residue and ends at the $(u+N)^{th}$ residue in the sequence X with the sequence Y from the jth residue to the $(u+N)^{th}$ residue.

$$C(j) = \frac{\sum_{i=1}^{N} (X(u+i-1) - \langle X \rangle)(Y(j+i-1) - \langle Y \rangle)}{\left[\left\{\sum_{i=1}^{N} (X(u+i-1) - \langle X \rangle)^2\right\}\left\{\sum_{i=1}^{N} (Y(j+i-1) - \langle\langle Y \rangle\rangle^2)\right\}\right]^{1/2}}$$

where

$$\langle X \rangle = \frac{1}{N}\left(\sum_{i=1}^{N} X(u+i-1)\right), \qquad \langle Y \rangle = \frac{1}{N}\left(\sum_{i=1}^{N} Y(j+i-1)\right)$$

Here X (u + i – 1) is the index value of an amino acid at the position (u + i – 1) in X and Y(j + i – 1) at the position (j + i – 1) in Y. The whole computation process have been carried out and automated by using an in house FORTRAN program in SUN ULTRA 40 M2 workstation.

RESULTS AND DISCUSSION

TARGET – TEMPLATE RELATIONSHIP:

Alignment of target and template remains a complex problem in the comparative modeling category and more importantly the quality of the alignments does not correlate well with the level of sequence identity approaching 40%. For the all alpha targets and templates, the BLOSUM and DAYHOFF scores are looking sensible. This is because of the less structural complexity of proteins in this structural class. But in case of all beta, alpha+beta and alpha/beta class targets and templates, the BLOSUM and DAYHOFF score are not evenly distributed. For the targets, T0130, T0468, T0414, T0497, T0152, T0138, T0135, T0168, T0189, T0471, T0482, T0466, T0462, T0474, T0473, T0421, T0413, T0420, T0400, T0435, T0479 and T0502 have negative scores which are given in Table 1 - 4. This clearly indicates that the scores or alignment based on the substitution matrices do not provide an optimal solution when the sequence identity is less than 40%. Improvements in comparative modeling could occur as the database of available targets continues to grow.

From our results of target-template alignment, FFAS server works pretty good than the other alignment algorithms based on the substitution matrices. The difference in percent identity between an optimal alignment (STRETCHER) and Profile-Profile alignment (FFAS) is given in Table 1 – 4. For example, the target T0177 has 28 identities out of 249 residues in the STRETCHER alignment where as it has 74 identities out of 240 residues in the profile based FFAS alignment. It has been pointed out earlier that there is a marked difference in residue identity/percent identity between two various alignments program and in the case of CASP targets and templates profile based alignment works much better than others [26].

The final quality of the model depends upon the selection of correct template which is someway related to the target. Hence we made a systematic analysis to explore the target – template relationship by using five important physico-chemical properties of the amino acid residues used by Argos. Since the sequences of homologous proteins can diverge beyond the point where their relationship cannot be recognized by pair wise sequence comparisons, the results suggest that the relationship between physico-chemical properties of amino acid residues between target and template is very crucial while selecting template for a target with very low sequence identity falling under comparative modeling category.

The physico-chemical correlation coefficients between targets and templates for each property are given in Table 1 - 4. In the case of all alpha targets and templates, the hydrophobicity and anti parallel turn preference property seem to be correlated well. This may be due to the similar packing angle preferences of helix-helix interactions [27] where as in the case of all beta targets and templates, the properties such as turn preference, bulkiness and anti parallel turn preference show sensible correlation coefficient than the other two properties. The abundance of beta turns in all beta proteins is a major reason for conservation of turn preference property which is reflected in the correlation coefficient between target and template.

The physical properties such as hydrophobicity and anti parallel turn preference were much conserved which is clearly reflected in the correlation coefficient between targets and templates of alpha+beta and alpha/beta classes. Sensible correlation of the bulkiness property was found between targets and templates of mixed structural class in the CASP experiment. This is due to the vital role of bulk property of an amino acid residue in packing organization of secondary structural elements in mixed class of proteins [28]. Interestingly, it is noted that very poor correlation coefficient compared to other properties was found for the refractive index property. For most of the targets and templates in all three structural classes, a reasonable correlation coefficient is observed for any one of the five properties.

In previous CASP experiments it has been demonstrated that predictors are rarely able to predict models that are closer to the target structure than the structure of the closest template [29]. The best performing groups used the same techniques with similar success as in previous years. The top three groups considered a range of templates and alignments from a range of sources before proceeding with the modeling step, though all did develop some new techniques to improve their predictions [30]. Sensitive methods such as physico-chemical correlation coefficient may be invoked to detect that similarity. Considering the above facts, we propose that profile based alignment and physico-chemical correlation coefficient may serve as powerful tools to select the template in order to model the structure for the targets in both comparative modeling and fold recognition category.

TABLE 1: Target – Template Relationship in all-α structural Class

			All Alpha Targets										
			STRETCHER		FFAS								
S. No	Target	Template	% ID	Identities	% ID	Identities	DAYHOFF	BLOSUM	Hydrophobicity	Turn Preference	Bulkiness	Refractive Index	Anti Parallel turn Preference
1	T0406	2QNL	6.60%	11(167)	14.00%	24(161)	20	47	0.37	0.33	0.16	0.14	0.25
2	T0456	3BHY	4.80%	14(291)	36.00%	100(275)	452	501	0.70	0.64	0.61	0.49	0.65
3	T0459	1LJ9	4.90%	7(144)	22.00%	19(86)	21	57	0.42	0.42	0.45	0.20	0.58
4	T0481	2F22	5.80%	9(154)	18.00%	28(149)	25	62	0.39	0.23	0.28	0.17	0.30
5	T0498	2J5Y	6.60%	4(61)	53.00%	24(45)	96	96	0.71	0.62	0.69	0.44	0.68
6	T9913	256B	5.60%	6(107)	No Significant Profile Alignment								

TABLE 2: Target – Template relationship in all-Beta Structural Class

All Beta Targets

S. No	Target	Template	STRETCHER % ID	STRETCHER Identities	FFAS % ID	FFAS Identities	DAYHOFF	BLOSUM	Hydrophobicity	Turn Preference	Bulkiness	Refractive Index	Anti Parallel turn Preference
1	T0130	1FA0	2.40%	13(537)	11.00%	12(104)	-32	-1	0.38	0.29	0.37	0.17	0.26
2	T0137	1PMP	25.60%	34(133)	42.00%	56(131)	264	280	0.60	0.68	0.61	0.59	0.66
3	T0190	1BZ8	7.10%	9(126)	29.00%	34(116)	113	134	0.55	0.43	0.52	0.51	0.57
4	T0392	2OCS	13.80%	15(109)	25.00%	22(87)	24	73	0.52	0.62	0.39	0.26	0.59
5	T0397	1X82	6.80%	13(190)	20.00%	10(50)	24	4	0.05	0.33	0.49	0.31	0.17
6	T0402	2I02	8.80%	13(148)	19.00%	26(134)	66	97	0.50	0.37	0.49	0.37	0.41
7	T0409	1H9M	8.30%	12(145)	21.00%	14(64)	7	31	0.23	0.41	0.35	0.31	0.34
8	T0412	1YSQ	6.20%	12(193)	18.00%	33(180)	31	46	0.30	0.32	0.41	0.17	0.33
9	T0414	2OA2	4.70%	7(148)	10.00%	14(132)	-29	-14	0.25	0.27	0.28	0.13	0.27
10	T0426	2NMX	36.40%	103(283)	59.00%	155(260)	762	866	0.77	0.74	0.82	0.81	0.78
11	T0468	1JB7	2.40%	12(495)	12.00%	13(103)	30	-6	0.25	0.14	0.16	0.20	0.33
12	T0488	2FE5	6.30%	6(95)	30.00%	29(94)	90	127	0.67	0.60	0.57	0.36	0.60
13	T0497	2I51	6.20%	12(195)	11.00%	19(164)	-11	-47	0.19	0.21	0.23	0.24	0.23
14	T9903	1AAJ	5.70%	6(105)	20.00%	3(15)	28	10	0.41	0.20	0.05	0.60	0.21
15	T9912	2MCM	10.70%	12(112)	No Significant Profile Alignment								
16	T0181	1FCP	1.60%	11(705)	No Significant Profile Alignment								
17	T0415	1C8C	3.70%	4(109)	No Significant Profile Alignment								

TABLE 3: Target – template relationship in α+β structural class

Alpha + Beta

S. No	Target	Template	STRETCHER % ID	STRETCHER Identities	FFAS % ID	FFAS Identities	DAYHOFF	BLOSUM	Hydrophobicity	Turn Preference	Bulkiness	Refractive Index	Anti Parallel turn Preference
1	T0135	1IPB	4.10%	9(217)	8.00%	9(103)	-37	-5	0.37	0.25	0.24	0.16	0.42
2	T0150	1CK2	19.80%	21(106)	33.00%	32(96)	97	139	0.52	0.45	0.62	0.48	0.53
3	T0152	1B6B	7.10%	15(210)	13.00%	23(171)	-34	-51	0.31	0.27	0.25	0.07	0.20
4	T0169	1QSN	4.30%	7(162)	14.00%	24(163)	5	47	0.41	0.37	0.44	0.26	0.44
5	T0192	1QSO	7.60%	13(171)	15.00%	25(160)	58	38	0.34	0.31	0.33	0.31	0.33
6	T0404	2J9C	11.80%	14(119)	18.00%	19(104)	29	31	0.25	0.32	0.40	0.21	0.34
7	T0451	1NWW	8.70%	13(149)	14.00%	19(133)	54	7	0.24	0.05	0.22	0.21	0.30
8	T0453	2R2Z	9.50%	9(95)	25.00%	24(93)	78	104	0.51	0.54	0.55	0.36	0.61
9	T0499	2IGD	47.50%	29(61)	62.00%	35(56)	161	167	0.76	0.60	0.56	0.65	0.77
10	T9901	10FV	5.80%	10(171)	No Significant Profile Alignment								
11	T0148	1AB8	6.40%	14(220)	No Significant Profile Alignment								
12	T0472	3BRC	5.10%	8(156)	No Significant Profile Alignment								

TABLE 4: Target – template relationship in α/β structural class

			Alpha/Beta										
			STRETCHER		FFAS								
S. No	Target	Template	% ID	Identities	% ID	Identities	DAYHOFF	BLOSUM	Hydrophobicity	Turn Preference	Bulkiness	Refractive Index	Anti Parallel turn Preference
1	T0138	1E6M	3.00%	4(135)	8.00%	11(128)	-27	-9	0.40	0.37	0.33	0.08	0.35
2	T0167	1JEO	20.60%	39(189)	33.00%	63(188)	214	304	0.61	0.61	0.59	0.49	0.63
3	T0168	1K56	5.80%	19(327)	8.00%	23(264)	-63	-65	0.24	0.19	0.20	0.12	0.16
4	T0177	1LFP	11.20%	28(249)	30.00%	74(240)	147	306	0.55	0.48	0.54	0.37	0.54
5	T0178	1JCJ	13.50%	35(260)	22.00%	56(244)	118	212	0.50	0.47	0.48	0.33	0.50
6	T0188	1EO1	12.00%	15(125)	27.00%	34(125)	90	126	0.52	0.45	0.48	0.41	0.54
7	T0189	1JXI	8.20%	26(319)	10.00%	29(268)	-41	-41	0.19	0.13	0.35	0.20	0.24
8	T0388	2F8A	10.60%	22(208)	29.00%	57(193)	239	253	0.50	0.41	0.55	0.48	0.57
9	T0389	1GMX	4.60%	7(153)	14.00%	20(134)	-30	52	0.37	0.42	0.47	0.38	0.41
10	T0400	1QST	3.70%	6(162)	10.00%	17(162)	-44	-5	0.39	0.30	0.37	0.06	0.24
11	T0411	1GMX	8.50%	12(141)	13.00%	16(120)	39	45	0.50	0.45	0.39	0.27	0.49
12	T0413	1GKL	6.60%	20(304)	10.00%	33(312)	-2	-45	0.27	0.21	0.20	0.21	0.20
13	T0420	2PRV	6.30%	12(189)	12.00%	12(93)	1	-10	0.42	0.37	0.33	0.02	0.45
14	T0421	1NN5	6.30%	19(300)	8.00%	20(228)	-44	-31	0.31	0.30	0.29	0.18	0.32
15	T0432	2RI7	3.40%	6(174)	24.00%	27(110)	73	89	0.45	0.58	0.25	0.17	0.45
16	T0433	1G0O	11.00%	31(283)	17.00%	42(246)	88	137	0.38	0.33	0.47	0.29	0.41
17	T0435	1ZKK	4.80%	8(167)	10.00%	16(148)	-67	-42	0.24	0.17	0.28	0.19	0.33
18	T0461	1P60	11.10%	21(189)	20.00%	35(167)	83	120	0.46	0.35	0.42	0.25	0.46
19	T0462	3B79	3.20%	5(154)	12.00%	5(40)	-5	5	0.39	0.39	0.23	-0.13	0.41
20	T0466	1X54	3.00%	13(434)	6.00%	6(91)	-21	-24	0.45	0.30	0.24	0.02	0.37
21	T0471	1F38	6.20%	12(192)	13.00%	15(112)	-24	19	0.34	0.36	0.36	0.13	0.46
22	T0473	2E65	3.80%	9(235)	12.00%	8(64)	3	-5	0.35	0.29	0.39	0.29	0.24
23	T0474	2GPE	5.00%	4(80)	10.00%	5(46)	-5	3	0.45	0.37	0.29	0.16	0.38
24	T0479	1VH5	10.10%	15(148)	20.00%	26(127)	-14	49	0.35	0.34	0.24	0.20	0.32
25	T0482	7AHL	3.10%	9(293)	12.00%	14(111)	-14	0	0.24	0.22	0.31	0.29	0.44
26	T0486	2J5I	7.50%	22(292)	21.00%	59(271)	87	174	0.39	0.39	0.50	0.40	0.36
27	T0492	2DTR	3.50%	8(226)	18.00%	11(60)	9	14	0.40	0.53	0.50	0.29	0.34
28	T0502	1XRG	3.70%	11(294)	15.00%	17(107)	-5	-2	0.14	0.29	0.22	0.04	0.36
29	T0157	1L6Y	4.30%	14(323)	No Significant Profile Alignment								
30	T0469	1ZR6	2.00%	10(503)	No Significant Profile Alignment								
31	T0393	2F1K	6.40%	18(281)	No Significant Profile Alignment								

Target – Target ID
Template – Template PDB ID
STRETCHER - STRETCHER Alignment results
FFAS - FFAS Alignment results
DAYHOFF – Dayhoff score between target and template
BLOSUM – BLOSUM Dayhoff score between target and template
Hydrophobicity, Turn Preference, Bulkiness, Refractive Index and anti parallel turn preference – Correlation coefficient of five properties between target and template.

Without investigating protein molecular function and interactions, protein sequences are not valuable. In order to investigate molecular function and interactions, structure elucidations based on computational methods are the only possible alternative to the experimental investigation of the protein structure space. Assessing the reliability of structure prediction methods requires more attention by comparing computational models with the corresponding experimental structures. Classical physics based methods

can give a reasonable estimate of the similarity between a target sequence and the corresponding experimental template structure when no relationship was detected between them by using several statistical measures [31].

CONCLUSION

Identifying the best structural template for a target is still a big challenge. But from the extent of structural conservation between target and template structures, the quality of models produced by comparative modeling is determined by various factors, such as the ability to deduce the correct structural alignment with the template protein and the accuracy of the modeling step. Sensible physico-chemical correlation coefficients have been found between target and template. This allows us to propose that alternate sophisticated methods such as physico-chemical correlation coefficient may be invoked to detect the similarity between targets and templates. Hence, we propose from our analysis that profile-profile alignment and physico-chemical correlation coefficient between target sequence and template structure may serve as a powerful tool to select the template in order to model the structure for the targets in both comparative modeling and fold recognition category.

REFERENCES

1. Berman H, Henrick K, Nakamura H, Markley JL. (2007). The worldwide Protein Data Bank (wwPDB): Ensuring a single uniform archive of PDB data. Nucleic Acids Res; 35: D301– 303.
2. Anfinsen CB. (1973). Principles that govern the folding of protein chains. Science; 181: 223-230.
3. Zemla A, Venclovas C, Fidelis K, Rost B. (1999). A modified definition of SOV, a segment-based measure for protein secondary structure prediction assessment Proteins; 34: 220–223.
4. Levin JM, Pascarella S, Argos P, Garnier J. (1993). Quantification of secondary structure prediction improvement using multiple alignments. Protein Eng; 6: 849–854.
5. Cuff JA, Barton GJ. (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins; 40: 502-511.
6. Combet C, Blanchet C, Geourjon C, Deleage G. (2000). NPS@: network protein sequence analysis. Trends Biochem Sci; 25: 147–150.
7. Cole C, Barber JD, Barton GJ. (2008). The Jpred3 secondary structure prediction server Nucleic Acids Res 35: W197-W201.
8. Gromiha MM, Selvaraj S. (1998). Protein secondary structure prediction in different structural classes. Protein Eng; 11: 249-251.
9. Gromiha MM, Selvaraj S. (2001). Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: Application of long-range order to folding rate prediction. J Mol Biol; 310: 27-32.
10. Saravanan KM, Selvaraj S. (2009). Analysis and visualization of long-range contacts in homologous families of proteins. The Open Struct Biol; 3: 104-125.
11. Saravanan KM, Harihar B, Saranya N, Selvaraj S.(2010). Sequence and Structural analysis of two designed proteins with 88% identity adopting different folds. Protein Eng Des & Sel; 23: 911-918.
12. Saravanan KM, Selvaraj S. (2011). Search for identical octapeptides in unrelated proteins: Structural plasticity revisited. Biopolymers 98(1):11-26. doi: 10.1002/bip.21676.
13. Cymmerman I, Feder M, Pawlowski M, Kurowski MA, Bujnicki JM. (2004). Computational Methods for Protein Structure Prediction and Fold Recognition. In Bujnicki JM, editor. Nucleic Acids and Molecular Biology, Springer Verlag 1-21: DOI: 10.1007/978-3-540-74268-5_1.
14. Argos P. (1987). A sensitive procedure to compare amino acid sequences. J Mol Biol; 193: 385 – 396.
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol 1990; 215: 403-410.
16. Myers EW, Miller W. (1988). Optimal alignments in linear space. Comput Applic Biosci; 4: 11-17.
17. Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A.(2005). FFAS03: a server for profile-profile sequence alignments. Nucl Acids Res; 33: W284-W288.
18. Dayhoff MO, Schwartz RM, Orcutt BC. (1978). In Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Washington . p 345-352.
19. Henikoff S, Henikoff JG. (1992). Performance evaluation of amino acid substitution matrices. Proc Natl Acad Sci USA ; 89: 10915–10919.
20. Manavalan P, Ponnuswamy PK. (1978). Hydrophobic character of amino acid residues in globular proteins. Nature ; 275: 673-674.
21. Jones DD. (1975). Amino acid properties and side chain orientations in proteins: A cross correlation approach. J Theor Biol; 50: 167–183.
22. Palau J, Argos P, (1982). Puigdomenech P. Protein secondary structure : Studies on the limits of prediction accuracy. Int J Peptide Protein Res; 19: 394-401.
23. Lifson S, Sander C. (1979). Anti parallel and parallel β strands differ in amino acid residue preferences. Nature; 282: 109-111.

24. Ponnuswamy PK, Selvaraj S. (1992). Structural similarities in the repeat sequences of plasma apolipoproteins, A-I, A-IV and E. Protein Seq and Data Anal; 5: 47-56.
25. Kubota Y, Takahashi S, Nishikawa K, Ooi T. (1981). Homology in protein sequences expressed by correlation coefficients. J. Theor Biol; 91: 347-361.
26. Yona G, Levitt M. (2002). Within the Twilight Zone: A Sensitive Profile-Profile Comparison Tool Based on Information Theory. J Mol Biol; 315: 1257 – 1275.
27. Hespenheide BM, Kuhn LA. (2003). Discovery of a significant, nontopological preference for antiparallel alignment of helices with parallel regions in sheets. Protein Sci; 12: 1119-1125.
28. Argos P, Rossmann MG, Grau UM, Zuber H, Frank G, Tratschin JD. (1979). Thermal stability and protein structure. Biochemistry; 18: 5698-5703.
29. Tramontano A, Morea V. (2003). Assessment of homology-based predictions in CASP5. Proteins; 53: 352-68.
30. Tress M, Ezkurdia I, Grana O, Lopez G, Valencia A. Assessment of predictions submitted for the CASP6 comparative modeling category. Proteins 2005; 61: 27-45.
31. Moult J. (2005). A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. Curr Opin Struct Biol; 15: 285 – 289.