VOL.4NO.2September2015 - ISSN 2233 – 1859

Operations Research Society in Bosnia and Herzegovina

Southeast Europe Journal of Soft Computing

Available online: http://.scjournal.ius.edu.ba

IUS Soft Computing Research Group

Secondary Structure and Neighbor Preferences of Amino Acids

Betul Akcesme, Mehmet Can
International University of Sarajevo,
Faculty of Engineering and Natural Sciences, HrasnickaCesta 15, Ilidža 71210 Sarajevo, Bosnia and Herzegovina
betul.cicek@yahoo.com
mcan@ius.edu.ba

Article Info

Article history:
Article received on17 Jul. 2015
Received in revised form on 17 Aug. 2015

Keywords:
Protein Structure, Protein Structure Prediction, Propensity Tables, Secondary Protein Structure

Abstract

The mystery of the relation between amino acid sequences and folding of the proteins started to fascinate researchers starting from 1960'ies. When three-dimensional structures of globular proteins were first obtained by X-ray crystallography, there was no obvious relation foundbetween amino acid sequence and conformation. The ability of globular proteins refold from their denatured, time-random coils in the absence of other biological material, led some scientists to believe in that all the information for the native, biologically active conformation is contained within the amino acid sequence. In 1970'ies Anfinsen postulated that the native structure of a protein depends only on the amino acid sequence and on the conditions of solution, and not on the kinetic folding pathway. During that decade protein folding code was seen as a sum of many small interactions. But the key idea was that the primary sequence encoded secondary structures, which then encoded tertiary structures. In this article the claim that primary sequence encodes the secondary structure will be tested by the propensity of amino acids to helix-sheet-coil conformations.

1. INTRODUCTION

Genetic variation of protein coding genes is one of the main components of genetic diversity. Too many studies have been done in order to understand how proteins evolve and gain different functions by amino acid substitutions. Mostly two approaches have been applied to study the pattern of amino acid substitutions: empirical and parametric. First one is based on the comparative analysis of amino acid sequences (Dayhoff and Barker 1972; Dayhoff, Schwartz, and Orcutt 1978; Dayhoff, Barker, and Hunt 1983). The second,with the objective of building a realistic model of amino acid substitutions, is initialized by studies on the relationship between amino acid dissimilarities and substitution patterns (Zuckerkandl and Pauling 1965; Sneath 1966; Epstein 1967; Clarke 1970; Grantham 1974; Miyata, Miyazawa, and Yasunaga 1979; Kimura 1983; Xia and Li 1998).

Although the reverse is supported by observations, both of these two approaches assume that amino acid substitutions at different amino acid sites are independent of each other. It is problematic because function of protein directly depends on its three dimensional structure and amino acids in a peptide cannot be thought independent form each other in order to have particular protein conformation.

Tendency of each amino acid being in a certain secondary structure varies among amino acids. Using a bunch of proteins at hand, Ala and Glu are considered as wellα-helix formers, whereasGly and Pro tend to disrupt the α-helix structure. Similarly,Glu and Pro are poor β-sheet

formers, whereas Ile and Val are good (Chou and Fasman 1974a, 1978b; Branden and Tooze 1998). According to outcomes of those empirical studies, conformational parameters have been proposed to predict secondary structure of proteins (Chou and Fasman 1978a).Chou-Fasman conformational parameters were driven from a data set containing low number of proteins. However, now PDB database contains around 100,000 protein with known structures so that study of Chou-Fasman should be renewed. We aim to update results of the propensity of the 20 amino acids found in helices, sheets and coils based on enlarged datasets.

Understanding neighboring effect of amino acids can give idea about amino acid dissimilarities. Grantham's distance and Miyata's distance are two indices of amino acid dissimilarity which are based on the volume, the polarity, and the chemical property of the side chain, and first two amino acid properties, respectively (Grantham 1974; Miyata, Miyazawa, and Yasunaga 1979),.

Even though, 134 properties of amino acids were listed (Sneath1966), 10 of them were studied in detail (Xia and Li 1998. 10 amino acid properties, the chemical composition of the side chain, two polarity measures, hydropathy, isoelectric point, volume, aromaticity, aliphaticity, hydrogenation, and hydroxythiolation, have been shown to affect on the evolution of the genetic code, the amino acid composition of proteins, and the pattern of nonsynonymous substitutions. Which amino acid properties should be applied to prepare an index of amino acid dissimilarity remained as an essential question? Both Grantham's and Miyata's distances were constructed arbitrary. It may cause the rise of old controversies among scientists (Kimura 1983; Gillespie 1991). Kimura proposed that nonsynonymous substitutions occur between similar amino acids and increasing dissimilarity between amino acids cause the decrease ofthe substitution rate. Whereas,Gillespie proposed that the most frequent nonsynonymous substitutions were not between the chemically most similar amino acids, but instead were between amino acids with a Miyata's distance near 1. However, amino acids with aMiyata's distance near 1 may be actually more similar to each other but, because inappropriate choice of amino acid properties they were considered as dissimilar.

XuhuaXia, and ZhengXie (Xia, and Xie 2002) concluded that amino acids interact with neighboring amino acids, to generate protein structures. They studied the pattern of association and repulsion of amino acids based on 24,748 protein-coding genes from human, 11,321 from mouse, and 15,028 from Escherichia coli, and documented the pattern of neighbor preference of aminoacids. They have found that all amino acids have different preferences for neighbors. They also analyzed 7,342 proteins with knownsecondary structure and estimated the propensity of the 20 amino acids occurring in three of the major secondarystructures, i.e., helices, sheets, and turns. They claim that in spite of the existence of a number of intriguing association andrepulsion patterns, much of the neighbor preference can be explained by the propensity ofthe amino acids in forming different secondary structures.

Theyfurther observed that amino acids having similar setsof neighbors substituting each other more frequently than those having very different sets of neighbors. Based on those findings, they concluded that the similarity in neighbor preference among amino acids is significantly correlated with thenumber of amino acid substitutions in both mitochondrial and nuclear genes

One of the three aims of this paper is to generate an updated propensity table of the 20 amino acids occurring in the three groups of secondary structure elements: helices, sheets, and coils, by using high number of know protein structures. The second aim is to estimate the genomic pattern of neighbor preference for the 20 amino acids by the huge amount of available protein data and interpret the neighbor preference with reference to protein secondary structures. The third is to incorporate the differences in neighbor preference between amino acids into a new formulation of amino acid dissimilarity index.

2. HISTORICAL BACKGROUND

For five decades, it has been believed that protein function, regulation, and interactions can be learned from their structure (Chou et. al 2009; Chou, 2004), which motivates development of novel methods for the prediction of the protein structure. These predictions concern various levels and aspects of the protein structure including the tertiary structure (Bujnicki, 2006; Floudas, 2007), solvent accessibility, depth, flexibility and packing of residues (Kurgan, et. al. 2008), and secondary structure (Rost 2003; Dehzangi et. al. 2014).

In contrast to the tertiary structure that describes position of each of the protein's atoms, the secondary structure simplifies the protein structure to a set of spatially local folding patterns that include a-helices, b-strands and coils. The spatial distribution of these local patterns determines the overall, three-dimensional shape of proteins in which individual secondary structures interact with each other creating more complex structures such as parallel or antiparallel β-sheets, β-barrels, and others. In spite that final product is complex, protein structures can be categorized into a few structural classes depending on the amount, types and spatial distribution of the secondary structures found in their fold.

3. CHOU-FASMAN CONFORMATIONAL PARAMETERS

Using the criteria described in (Chou and Fasman 1978a) for the four conformational states, they delineated the number of amino acids in the α, β, coil,andβ-turn regions of 29 proteins (Chou, and Fasman,1978b). The average frequency for helices, β-sheets, and β-turns were respectively

$\langle f_\alpha \rangle = 0.38,\ \langle f_\beta \rangle = 0.20, and\ \langle f_t \rangle = 0.32.$ (1)

When the frequency of residuesin the α, β, and β-turn regions are divided by their respective averagefrequency, their conformational parameters are obtained:

$P_\alpha = \langle f_\alpha \rangle / \langle f_\alpha \rangle, P_\beta = f_\beta / \langle f_\beta \rangle, and\ P_t = f_t / \langle f_t \rangle.$ (2)

These conformational potentials areshown in Table 1.

Table 1 Conformational parameters for α-helical, β-sheet, and β-turn residues in 29 proteins (simplified from Chou and Fasman 1978b).

AA	Pa	AA	Pb	AA	Pt
Glu	1.51	Val	1.7	Asn	1.56
Met	1.45	Ile	1.6	Gly	1.56
Ala	1.42	Tyr	1.47	Pro	1.52
Leu	1.21	Phe	1.38	Asp	1.46
Lys	1.16	Trp	1.37	Ser	1.43
Phe	1.13	Leu	1.3	Cys	1.19
Gln	1.11	Cys	1.19	Tyr	1.14
Trp	1.08	Thr	1.19	Lys	1.01
Ile	1.08	Gln	1.1	Gln	0.98
Val	1.06	Met	1.05	Thr	0.96
Asp	1.01	Arg	0.93	Trp	0.96
His	1.	Asn	0.89	Arg	0.95
Arg	0.98	His	0.87	His	0.95
Thr	0.83	Ala	0.83	Glu	0.74
Ser	0.77	Ser	0.75	Ala	0.66
Cys	0.7	Gly	0.75	Met	0.6
Tyr	0.69	Lys	0.74	Phe	0.6
Asn	0.67	Pro	0.54	Leu	0.59
Pro	0.57	Asp	0.55	Val	0.5
Gly	0.57	Glu	0.37	Ile	0.47

It shouldbe noted that all five charged residues (Arg, Asp, Gln, His, Lys) are unfavorable for β formation with $P_\beta <$ 1.00, while three of them (Asp, His, and Arg) are helical indifferentwith $P_\alpha \cong$ 1.00. On the other hand, a-breaking residues (Pro, Gly, and Asn)are strong β-turn formers with $P_t >$ 1.50, while β-formers are generallyfound infrequently in bend regions.

After forty years, using all proteins of the PDB database we modify this table as follows.

Table 2 Conformational parameters for α-helical, β-sheet, and β-turn calculated from 2,066,0981residues in 80,592 proteins.

AA	Pa	AA	Pb	AA	Pt
Leu	2.38	Val	2.62	Leu	2.38
Ala	2.21	Leu	2.08	Ala	2.21
His	1.79	Ile	1.87	His	1.79
Lys	1.32	Thr	1.39	Lys	1.32
Val	1.24	Ala	1.22	Val	1.24
Ile	1.2	Phe	1.14	Ile	1.2
Arg	1.18	Ser	1.1	Arg	1.18
Asp	1.01	Tyr	1.05	Asp	1.01
Ser	0.99	Lys	0.96	Ser	0.99
Gln	0.94	Gly	0.94	Gln	0.94
Thr	0.84	His	0.9	Thr	0.84
Phe	0.82	Arg	0.88	Phe	0.82
Gly	0.71	Asp	0.63	Gly	0.71
Tyr	0.68	Gln	0.6	Tyr	0.68
Asn	0.66	Asu	0.54	Asn	0.66
Met	0.53	Glu	0.46	Met	0.53
Pro	0.49	Met	0.43	Pro	0.49
Glu	0.42	Pro	0.4	Glu	0.42
Trp	0.31	Cys	0.4	Trp	0.31
Cys	0.23	Trp	0.38	Cys	0.23

Dramatic changes observed in

Glu: P_α: 1.51 → 0.42, Met: P_α: 1.45 → 0.53,
Cys: P_α: 0.70 → 0.23, Trp: P_β: 1.37 → 0.38,
Cys: P_β: 1.19 → 0.40, Met: P_β: 1.05 → 0.43,
Asn: P_t: 1.56 → 0.66, Gly: P_t: 1.56 → 0.71,
Pro: P_t: 1.52 → 0.49, Cys: P_t: 1.19 → 0.23.

4. XIA, AND XIE PROBABILITIES OF AMINO ACIDS OCCURRING IN HELICES, SHEETS, AND COILS

Xia, and Xie 2002 retrieved 7, 342 proteins with known structures from the PDB database (Berman et al. 2000), extracted helices, sheets, and coils according to the PDB Format Description, Version 2.2, and counted the frequency distribution of amino acids in each of the three structure categories.

Table 3 Frequency distribution of amino acids in helices, sheets, computed from Xia, and Xie 2002

amino	H	S	C
Ala	0.62	0.2	0.18
Arg	0.7	0.28	0.02
Asu	0.66	0.3	0.04
Asp	0.68	0.27	0.04
Cys	0.53	0.45	0.03
Gln	0.7	0.28	0.02
Glu	0.73	0.25	0.02
Gly	0.56	0.38	0.06
His	0.6	0.37	0.03
Ile	0.53	0.46	0.01
Leu	0.67	0.32	0.02
Lys	0.7	0.28	0.02
Met	0.67	0.3	0.02
Phe	0.58	0.41	0.01
Pro	0.59	0.35	0.06
Ser	0.6	0.37	0.04
Thr	0.54	0.42	0.03
Trp	0.58	0.4	0.02
Tyr	0.56	0.42	0.02
Val	0.51	0.48	0.01

Today at PDB database there around 200,000 proteins with predicted secondary structure. We eliminated duplications, and short proteins of length less than 30 residues. In Table 2we listed the probabilities of amino acids to be in α-helical, β-sheet, and β-turn conformations calculated from 20,660,981residues in 80,592 proteins.The comparison of probabilities in Table 1, and Table 2, dramatic changes is observed in many amino acids and conformations.

Table 4Probabilities of amino acids to reside in α-helical, β-sheet, and β-turn conformations calculated from 20,660,981residues in 80,592 proteins.

amino	H	S	C
Ala	0.48	0.18	0.34
Arrg	0.4	0.22	0.38
Asu	0.26	0.15	0.59
Asp	0.31	0.13	0.56
Cys	0.26	0.32	0.42
Gln	0.42	0.2	0.38
Glu	0.31	0.24	0.45
Gly	0.16	0.16	0.68
His	0.47	0.17	0.36
Ile	0.34	0.41	0.25
Leu	0.43	0.28	0.29
Lys	0.39	0.2	0.41
Met	0.42	0.26	0.33
Phe	0.33	0.35	0.32
Pro	0.18	0.1	0.72
Ser	0.27	0.21	0.52
Thr	0.24	0.3	0.45
Trp	0.35	0.34	0.32
Tyr	0.32	0.35	0.33
Val	0.29	0.44	0.26

When table is normalized column wise, the distribution of conformal states along amino acids are obtained. Table 3 shows these probabilities. For example the probabilities in the first column, are probabilities of helix conformations being at a certain amino acid.

Table 5 Probabilities of conformations being at a given amino acid.

amino	H	S	C
Ala	0.12	0.06	0.07
Arrg	0.06	0.04	0.04
Asu	0.03	0.03	0.06
Asp	0.05	0.03	0.08
Cys	0.01	0.02	0.02
Gln	0.05	0.03	0.03
Glu	0.02	0.02	0.02
Gly	0.04	0.05	0.12
His	0.09	0.04	0.05
Ile	0.06	0.09	0.03
Leu	0.11	0.1	0.06
Lys	0.07	0.05	0.06
Met	0.02	0.02	0.01
Phe	0.04	0.06	0.03
Pro	0.02	0.02	0.08
Ser	0.05	0.06	0.08
Thr	0.04	0.07	0.06
Trp	0.02	0.02	0.01
Tyr	0.04	0.05	0.03
Val	0.06	0.13	0.04

Xia, and Xiein (Xia, and Xie, 2002) also calculated propensities of amino acids occurring in one of the three structure categories. They do this as follows: Let N_{Tot} be the total number of amino acids in the three structure categories; N_i (where i = 1, 2, . . . , 20 corresponding to the 20 amino acids) be the number of amino acid *i* found in all three structure categories; N_h, N_s, and N_t be the number of amino acids found in helices, sheets, and coils, respectively; and $N_{h,i}$, $N_{s,i}$, and $N_{t,i}$be the number of amino acids in helices, sheets, and coils, respectively. If amino acids occur equally likely in the three secondary structures, then the expected numbers of $N_{h,i}$, $N_{s,i}$, and $N_{t,i}$are, respectively,

$$E(N_{h,i}) = \frac{N_h N_i}{N_{Tot}}, E(N_{s,i}) = \frac{N_s N_i}{N_{Tot}}, E(N_{t,i}) = \frac{N_t N_i}{N_{Tot}} \quad (3)$$

The propensity of amino acid i occurring in helices is defined as

$$P_{h,i} = \frac{N_{h,i} - E(N_{h,i})}{N_i}, P_{s,i} = \frac{N_{s,i} - E(N_{s,i})}{N_i}, P_{t,i} = \frac{N_{t,i} - E(N_{t,i})}{N_i} \quad (4)$$

$P_{h,i}$ measures how strongly an amino acid is associated with one particular secondary structure and is independent of sample size. Xia, and Xiein (Xia, and Xie, 2002) retrieved only 7,342 proteins instead of all proteins in the PDB database.They claimed that$P_{h,i}$, $P_{s,i}$, and $P_{t,i}$ values would be stabilized after analyzing just 3,000 protein structures. However our research proved the reverse.

Table 6 Propensities of the amino acids to occur in secondary structures. In this Table, P_h and P_s are strongly and negatively correlated.

aa	PH	PS	PC
Ala	0.1046	□0.1001	□0.0044
Arrg	0.0727	□0.0643	□0.0085
Asu	0.0257	□0.0439	0.0181
Asp	0.0532	□0.0699	0.0167
Cys	□0.105	0.1025	0.0025
Gln	0.0727	□0.0675	□0.0053
Glu	0.0984	□0.0942	□0.0041
Gly	□0.0686	0.0335	0.0351
His	□0.0315	0.0312	0.0003
Ile	□0.1006	0.1142	□0.0136
Leu	0.038	□0.0272	□0.0108
Lys	0.0637	□0.0616	□0.0021
Met	0.0427	□0.038	□0.0047
Phe	□0.0554	0.0668	□0.0114
Pro	□0.038	0.0041	0.0338
Ser	□0.033	0.0227	0.0104
Thr	□0.0866	0.0817	0.0049
Trp	□0.0506	0.0579	□0.0074
Tyr	□0.0741	0.0806	□0.0065
Val	□0.1228	0.1343	0.0115

However the propensity calculations obtained from the largest database available, 20,660,981residues in 80,592 proteins showed that the symmetry in the above table cannot be generalized.In Table 5we depicted the propensities of amino acids calculated through (3-4). Comparing probabilities in Table 4, and Table 5, dramatic changes is observed in many amino acids and conformations.

Table 7 Propensities of the amino acids to occur in secondary structures obtained from 20,660,981residues. In this Table, strong and negative correlation of the P_h and $P_s$valuesare not observed.

aa	PH	PS	PC
Ala	0.1422	-0.0595	-0.0827
Arrg	0.0653	-0.0206	-0.0447
Asu	-0.0743	-0.0905	0.1648
Asp	-0.0267	-0.1115	0.1383
Cys	-0.0776	0.0789	-0.0013
Gln	0.0821	-0.0426	-0.0395
Glu	-0.026	0.0011	0.0249
Gly	-0.1722	-0.0879	0.2602
His	0.1394	-0.0718	-0.0676
Ile	0.0107	0.1647	-0.1754
Leu	0.101	0.0337	-0.1347
Lys	0.0537	-0.042	-0.0117
Met	0.0826	0.0114	-0.094
Phe	-0.0017	0.1026	-0.1009
Pro	-0.1578	-0.1412	0.299
Ser	-0.0662	-0.032	0.0982
Thr	-0.0906	0.06	0.0306
Trp	0.0121	0.0934	-0.1055
Tyr	-0.017	0.1083	-0.0913
Val	-0.0427	0.2012	-0.1585

5. NEIGHBOR PREFERENCE IN AMINO ACIDS

There are 400 possible amino acid doublets, with 20 amino acids. Let $N_{ij},\ i,j = 1,2,.20$ corresponding to the 20 amino acids be the number of amino acid pairs, with amino acid *j* following amino acid *i*. For example, $N_{Ala,Arg}$ is the number of Ala-Arg pairs in all sequences; $N_{Arg,Ala}$ is the number of Arg-Ala pairs in all sequences, and so on.
The N_{ij} values apparently depend on amino acid usage. If amino acid *j* is very abundant, then obviously N_{ij} and N_{ji} will be large, too. If amino acid i does not have any neighbor preference, then the expected value for N_{ij} is

$$E(N_{ij}) = P_j \sum_{j=1}^{20} N_{ij} \quad (5)$$

wherePj is the frequency of amino acid *j*. It seems that certain amino acids to be neighbors more likely than expected from random association. For instance, good α-helix formers should be more likely to be neighbors, as should β-sheet formers.

Whether the 20 N_{ij} values for amino acid *i* deviate significantly from the expectation of random association can be tested by a chi-square goodness-of-fit test with

$$\chi_i^2 = \sum_{j=1}^{20} \frac{\left(N_{ij} - E(N_{ij})\right)^2}{E(N_{ij})} \quad (6)$$

The degree of freedom associated with χ^2 is 19. SinceP_j is not calculated from the 20 N_{ij} valuesχ^2 calculated as 19 rather than 18. The strength of the neighbor preference can be measured with following formula:

$$SP_i = \sqrt{\chi_i^2 / \sum_{j=1}^{20} N_{ij}} \quad (7)$$

Since x^2value depends on the sample size we should avoid to use x^2directly for measuring the strength of preference. A more abundant amino acid tends to yield a large x^2 value than a less abundant amino acid. In contrast, SP_i is independent of sample size and can therefore facilitate comparisons among amino acids. SP_icannot tell about amino acid tendency for being in certain position since it takes only positive values. Xia, and Xie (Xia, and Xie, 2000) also use the following index (I_{ij}) to measure the preference of amino acid *i* for amino acid *j*:

$$I_{ij} = \left(N_{ij} - E(N_{ij})\right) / E(N_{ij}) \quad (8)$$

Apparently, I_{ij} will be positive if amino acid i has amino acid j as its neighbor more frequently than expected, and negative if amino acid i has amino acid j as its neighbor less frequently than expected.N_{ij} may differ from N_{ji}, i.e., amino acid i may have different preferences for amino acids that go before it and those that go after it.

Table 8.SP_istrength of neighbor preference values from Equation (7).χ^2test shows that, there is significant difference between the random, and observed frequencies.

amino	X2 After	SPI After	SPI Before
Ala	622.521	0.1	0.1079
Arg	300.538	0.0873	0.0759
Asu	368.91	0.1057	0.1066
Asp	473.815	0.1024	0.0979
Cys	96.9579	0.095	0.1162
Gln	244.8	0.0899	0.1026
Glu	23205.2	1.0095	0.9978
Gly	415.732	0.085	0.086
His	728.053	0.1179	0.1204
Ile	203.969	0.0682	0.0901
Leu	511.858	0.0843	0.0638
Lys	407.911	0.0956	0.1106
Met	267.982	0.1189	0.1475
Phe	307.695	0.0993	0.1131
Pro	487.612	0.116	0.1243
Ser	542.727	0.1042	0.125
Thr	486.341	0.1062	0.094
Trp	164.106	0.1263	0.1463
Tyr	341.854	0.1122	0.0985
Val	383.397	0.0846	0.0687
Chi	30551.4	DF 19	Prob 0.00

6. RESULTS AND DISCUSSION

Although seemingly different amino acids have some association with particular secondary structures, this association is found to be dependent on protein families considered. Xia, and Xiein (Xia, and Xie, 2002) observed that Ala and Gluare found most frequently in helices, Val, Cys, and Ile found most frequently in sheets and Gly and Profound most frequently in coils (table 1).

When the propensity investigated a larger data set, it is found that at least the positions of Glu, Cys, Gly, and Pro changed dramatically:

Glu: $P_\alpha: 1.51 \rightarrow 0.42$, Cys: $P_\beta: 1.19 \rightarrow 0.40$,
Gly: $P_t: 1.56 \rightarrow 0.71$, Pro: $P_t: 1.52 \rightarrow 0.49$.

On the other hand, while in Table 4 propensities of the amino acids to occur in secondary structures, P_h and P_s are seen strongly and negatively correlated, when database enlarged in Table 5, propensities of the amino acids to occur in secondary structures obtained from 20,660,981residues, strong and negative correlation of the P_h and $P_s$valuesare not observable any more.

Neighbor preference in amino acids prevails in even larger databases. For each of the 20 amino acids, the N_{ij} values deviate highly significantly from $E(N_{ij})$ also in thelargerdataset, with P = 0.0000.

REFERENCES

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, And P. E. Bourne. (2000) The Protein Data Bank. Nucleic Acids Res. 28:235– 242.

Branden, C., AndTooze, J. (1998) Introduction To Protein Structure. Garland Publishing, Inc., New York.

Bujnicki, J.M. (2006) Protein-structure prediction by recombination of fragments.Chembiochem , 7(1):19–27.

Chou, K.C. (2004) Structural bioinformatics and its impact to biomedical science.Curr Med Chem, 11(21):05–34.

Chou, P.Y., AndFasman, G. D. (1974a) Conformational Parameters For Amino Acids In Helical, Beta-Sheet, And Random Coil Regions Calculated From Proteins. Biochemistry 13:211– 222.

Chou, K.C., Wei, D., Du, Q. Sirois, S., and Zhong, W. (2006) Progress in computational approach to drug development against SARS.Curr Med Chem, 13(32):63–70.

Chou, P.Y., AndFasman, G. D. (1978a) Empirical Predictions Of Protein Conformation. Annu. Rev. Biochem. 47:251–276.

Chou, P.Y., AndFasman, G. D. (1978b) Prediction Of The Secondary Structure Of Proteins From Their Amino Acid Sequence. Adv. Enzymol. Relat. Areas Mol. Biol. 47:45–148.

Clarke, B. (1970) Selective Constraints On Amino-Acid Substitutions During The Evolution Of Proteins. Nature 228:159–160.

Dayhoff, M. O., Schwartz, R. M., and Orcutt. B. C. (1978) A Model Of Evolutionary Change In Protein. Pp. 345–352 In M. O. Dayhoff, Ed. Atlas Of Protein Sequence And Structure. Natl. Biomed.Res. Found., Silver Spring, Md.

Dayhoff, M. O., and Barker W.C. (1972) Mechanisms and Molecular Evolution: Examples. pp. 41–45 in M. O. Dayhoff, ed. Atlas of Protein Sequence And Structure. Natl. Bio- Med. Res. Found., Washington, D.C.

Dayhoff, M. O., Barker, W. C., And Hunt L. T. (1983) Establishing Homologies In Protein Sequences. Methods Enzymol. 91:524–545.

Dehzangi, A., Paliwal, K., Lyons, J., Sharma, A., and Sattar, A. (2014) Proposing a highly accurate protein structural class predictor using segmentation-based features, The Twelfth Asia Pacific Bioinformatics Conference (APBC 2014) Shanghai, China. 17-19 January 2014.

Epstein, C. J. (1967) Non-Randomness Of Amino-Acid Changes In The Evolution Of Homologous Proteins. Nature 215:355– 359.

Floudas, C.A. (2007) Computational methods in protein structure prediction. BiotechnolBioeng, 97(2):207–213.

Gillespie, J. H. (1991) The Causes Of Molecular Evolution. Oxford University Press, Oxford.

Grantham, R. (1974) Amino Acid Difference Formula To Help Explain Protein Evolution. Science 185:862–864.

Kimura, M. (1983) The Neutral Theory Of Molecular Evolution. Cambridge University Press, Cambridge,United Kingdom.

Kurgan, L.A., Cios, K.J., Zhang, H., Zhang, T., Chen, K., Shen, S., and Ruan, J. (2008) Sequence-based methods for real value predictions of protein structure. Current Bioinformatics, 3(3):183–196.

Miyata, T., Miyazawa, S., AndYasunaga, T. (1979). Two Types of Amino Acid Substitutions In Protein Evolution. J. Mol. Evol. 12:219–236.

Rost, B. (2003) Prediction in 1D: secondary structure, membrane helices, and accessibility. Methods Biochem Anal, 44:559–587.

Singh, J., And Thornton, J. M. (1992) Atlas Of Protein Side- Chain Interactions. IRL Press, Oxford.

Sneath, P. H. A. (1966) Relations Between Chemical Structure & Biological Activity In Peptides. J. Theor. Biol. 12:157– 195.

Xia, X., And Li, W.H. (1998) What Amino Acid Properties Affect Protein Evolution? J. Mol. Evol. 47:557–564.

Zuckerkandl, E., And Pauling, L. (1965). Evolutionary Divergence And Convergence In Proteins. Pp. 97–166 In V. Bryson And H. J. Vogel, Eds. Evolving Genes And Proteins. Academic Press, New York.

Xia, X., and Xie, Z. (2002) Protein Structure, Neighbor Effect, and a New Index of Amino Acid Dissimilarities, Mol. Biol. Evol. 19(1):58–67.