Indian J. Applied & Pure Bio. Vol. 38(3), 1308-1314 (2023). **ISSN: 0970-2091**
A Web of Science Journal

Analysis of the secondary structural elements of a Ribovirus coiled-coil protein using the DSSP assignment method

[*]**P. Thenmozhi Kanaga and S. Arul Mugilan**

P.G. & Research Department of Physics, Kamarajar Government Arts College, Surandai-627859 (India), Affiliated to Manonmanium Sundaranar University, Abishakapatti, Tirunelveli-627012 (India)
*Register number 21111062132005
[*]Corresponding Author E-mail: **thenmozhikanaga@gmail.com**

Abstract

One of the most important problems in computational biology is the assignment of the secondary structure of proteins, which has applications in medicine and biotechnology. **DSSP, STRIDE, XTLsstr, PSEA, SEGNO, SECSTR, P-CURVE** and **KAKSI** are some of the secondary structure assignment techniques that are now available. Using the DSSP assignment approach, we investigate which secondary structural elements are present in the ribovirus coiled-coil protein. Amino acid residues are used to generate the deviation parameter values, which are then normalized. With these parameters, the secondary structural element in the ribovirus coiled-coil protein can be predicted. The graphical representation shows the preferential and non-preferential amino acid residues in the secondary structural element of ribovirus coiled-coil protein.

Key words : Coiled Coil protein, Dictionary of secondary structure of proteins (DSSP), Deviation parameter value, Drug designing, secondary structure elements.

Proteins are polypeptide structures made up of one or more extended chains of amino acid residues. It is possible to recognize a protein at every level of its structure. Primary, secondary, tertiary, and quaternary structures of each protein are all present in at least one instance. When comparing all other structures of proteins, secondary structures play an important role in numerous applications of structural biology[21]. To assign the secondary structure of the protein, many approaches such as DSSP, STRIDE, PSEA, XTLSSTR, and SECSTR are available. These strategies, based on multiple methods, logically produce results that may vary slightly[11,19]. Acceptable Cα -spacing measurements and dihedral angles are defined by KAKSI before an assignment is made[8,22]. STRIDE is a piece of software that is similar to DSSP. It employs hydrogen-bond patterns in a manner quite similar to that of DSSP, however, the concept of hydrogen-bonds differs slightly. The secondary structures assigned by STRIDE also take into account (/) angles[5]. DEFINE

uses only Cα coordinates to compare Cα distances with distances in represented secondary structure segments. In addition, super secondary structures are described[17]. The global peptide axis is generated by the P-CURVE technique, which is based on setting helical parameters for peptide units[18]. PSEA only considers Cα atoms. This involves parameters for distance and angle[12]. With the help of estimated distances and angles from the backbone geometry, XTLSSTR was created to assign secondary structures "in the same way a person assigns structure visually". It is focused on interactions between amides[10]. The assignments performed by SEGNO are based on Cα coordinates, Phi/ Psi angles and an angle-distance hydrogen bond[13]. To locate and explore the rare pi-helices, SECSTR was created[4]. Secondary structure assignments are classified into eight states: alpha helix, beta sheet, beta turn, bend, pi-helix, 3/10-helix, isolated beta-bridge, and random structure. In times of need, secondary structure assignments have been divided into three classes (H for -helix, b for strand, and c for coil) as follows: XTLSSTR: (G,g,H,h) = H, (E,e) = b, others (T,N,P,p,-) = c; DSSP, STRIDE and SECSTR: (H,G,I) = H, (E,b) = b, others (S,T,blank) = c [8]. In this article, we use the DSSP database to perform a structural study of the ribovirus coiled-coil protein.DSSP, which uses reasonably secure criteria, is the most commonly used method for assigning secondary structure WOLFGANG KABSCH and CHRISTIAN SANDER described in a previous literature review "Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features"[6,7,9,15,16]. The main goals of the current work are to determine the amino acid residue distribution and the normalized deviation parameter value for each secondary structural elements. A graphical representation is also available to describe the amino acid position in the coiled-coil protein. Arul Mugilan and Veluraja have discussed the deviation parameter value for alpha helix, beta sheet and random structures in previous work[14]. However, in our work, we calculate the normalized deviation parameter value for alpha helix, beta sheet, beta turn, pi-helix, 3/10-helix, isolated beta-bridge and random structure. The structural study of the ribovirus coiled-coil protein is used by experimental biologists for drug development[1,3].

The coiled-coil structure of the ribovirus consists of 488 proteins. The first step was to obtain protein IDs in coiled-coil structure from the protein database. The DSSP files were retrieved from the DSSP database for the relevant coiled-coil structure of the protein in the second step. From the DSSP database we acquired 478 DSSP files. The DSSP database does not have the remaining hundred file details. The current study includes 478 proteins. The third step was to use the PYTHON programming language to predict the number of amino acid residues present in the secondary structure elements represented in the coiled-coil structure. The results are taken from Microsoft Excel. In the fourth phase, the amino acid residues are used to calculate the bias parameter value, and then these values are normalized. The fifth step was to create a plot between amino acid residues and the normalized deviation parameter value. The graph represents the result of the structural analysis of the ribovirus coiled-coil protein. Flow chart of the present work is given in Fig. 1.

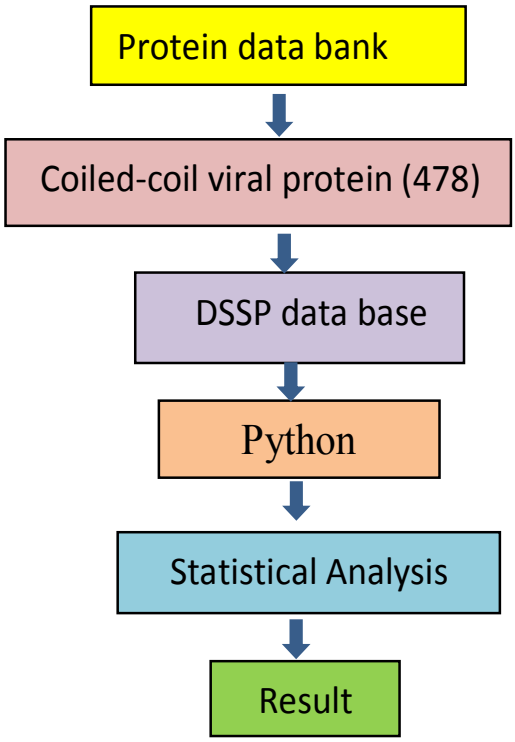

Fig. 1. Flow chart

Deviation parameter value calculation :

The deviation parameter value for coiled-coil proteins can be found using the formula described here[2,14];

$$DP = \frac{Oserved - Expected}{Expected} \times 100$$

The deviation parameter values are then normalized using the following formula[20].

$$N = \frac{X - Minimum}{Maximum - Minimum}$$

The number of residues of the secondary structural elements of the selected protein (478) was also determined from the DSSP output using the Python program. The distribution of amino acids in the ribovirus coiled-coil protein is shown in Table1. The table gives a total of 5, 75,043 amino acid residues. There are 23.2% alpha helices, 27% beta sheets, 7.5% beta turn, 0.2% pi-helices, 1.6% 3/10-helices, 2.5% isolated beta-bridges, and 38% random structure in the total number of amino acid residues. In addition, the ribovirus coiled-coil protein does not have a bend structure. This is verified by the output of the DSSP files. Table-2 shows the normalized deviation parameter value for the secondary structure elements. S.A Mugilan and K. Veluraja have predicted the existence of Ala (alpha helix), Val (beta strand) and Pro (random structure) with the highest deviation parameter value and Pro (alpha helix, beta sheet) with the lowest[14]. Other secondary structure elements (beta turn, pi-helix, 3/10-helix and isolated beta-bridge) do not have the deviation parameter value in the previous literature review. Secondary structural elements of the deviation parameter values are determined using DSSP assignment in this study. The maximum normalized deviation parameter value is held by Leu (Alpha helix), Phe (Beta sheet), Gly (Beta turn), Ala (Pi-helix), Pro (3/10-helix), Trp (Isolated beta-bridge), and Cys (random structure). The minimum normalized deviation parameter value has been contributed by Pro (Alpha helix, Beta sheet), Trp (Beta turn), Cys, Met, Pro (Pi-helix), Ile (3/10-helix, Random Structure), and Cys (Isolated beta-bridge. Also, the normalized deviation parameter values of the amino acid residues differ slightly from previously published data[14].

Table-1. The amino acid residues distribution of ribovirus coiled-coil protein

Amino acid	Total number of amino acid residues	Alpha helix(H)	Beta sheet(E)	Beta turn(T)	Pi-helix(I)	3/10 -helix(G)	Isolated beta-bridge residue(B)	Random structure
A	32071	8047	10731	2632	351	302	282	9726
R	25269	7787	5520	2323	13	708	1083	7835
N	45967	10718	6362	6116	37	869	669	21196
D	29750	8781	3872	2206	20	1267	204	13400
C	731	55	79	5	0	3	1	588
E	38002	13144	8442	2889	4	660	455	12408
Q	23371	8081	5888	1715	28	503	646	6510
G	46324	3209	10811	8984	185	174	259	22702
H	14233	2034	6164	245	8	516	1036	4230
I	37097	9751	15467	972	197	90	2198	8422
L	43355	17177	12767	2004	27	551	741	10088
K	35047	9783	8430	2110	12	564	726	13422
M	10046	3158	2929	557	0	158	46	3198
F	19984	1587	9171	1402	192	588	557	6487
P	21366	409	1895	3083	0	1021	49	14909
S	42326	7858	9844	3736	8	780	1056	19044
T	45896	7518	10861	1065	4	164	1263	25021
W	10553	2082	3891	59	5	33	1204	3279
Y	21975	5167	8469	474	8	127	164	7566
V	31680	6927	13396	502	17	207	1911	8720
Total	575043	133273	154989	43079	1116	9285	14550	218751

Graphical representation

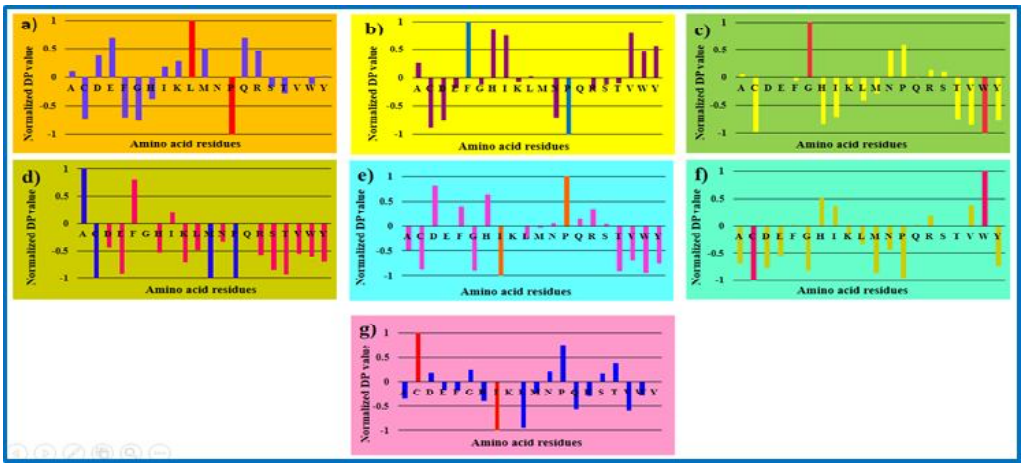

Fig. 2. Graphical representation for normalized deviation parameter value a) alpha helix b) beta sheet c) beta turn d) pi helix e) 3/10 helix e) isolated beta bridge f) random structure

Table-2. Normalized deviation parameter value for ribovirus coiled coil protein

Amino acid	Alpha-helix	Beta-sheet	Beta-turn	Pi-helix	3/10-Helix	Isolated beta-bridge	Random structure
A	0.108846	0.257188	0.05127	1	-0.4885	-0.68513	-0.34991
C	-0.71883	-0.88129	-0.98178	-1	-0.87723	-1	1
D	0.380259	-0.74603	0	-0.43887	0.829119	-0.76722	0.160056
E	0.691354	-0.18249	0	-0.91214	0	-0.55027	-0.15171
F	-0.69791	1	-0.05827	0.807681	0.396336	0.004084	-0.1679
G	-0.74874	-0.11368	1	0	-0.90268	-0.82091	0.254146
H	-0.37966	0.845606	-0.83048	-0.53085	0.620877	0.522915	-0.40166
I	0.182081	0.749142	-0.69938	0.189443	-1	0.366531	-1
K	0.281991	-0.06985	-0.20342	-0.71421	0	-0.17956	0
L	1	0.01743	-0.40735	-0.48019	-0.24758	-0.33322	-0.95176
M	0.497996	0	-0.27284	-1	-0.02671	-0.86383	-0.2214
N	0	-0.69548	0.483646	-0.32815	0.05054	-0.44083	0.185432
P	-1	-1	0.578991	-1	1	-0.96076	0.747069
Q	0.690703	0	-0.01123	0	0.13659	0.001413	-0.56065
R	0.460028	-0.20513	0.134914	-0.57059	0.350147	0.177195	-0.29194
S	-0.16541	-0.1186	0.103845	-0.84224	0.034876	0	0.158908
T	-0.27492	-0.09368	-0.74309	-0.92725	-0.91606	0	0.384892
V	0	0.784513	-0.85042	-0.5521	-0.69941	0.378913	-0.58877
W	-0.10709	0.460996	-1	-0.60453	-0.94871	1	-0.28636
Y	0.012044	0.560675	-0.76693	-0.69613	-0.75464	-0.74153	0

The graph shows the relationship between amino acid residues and the value of the normalized deviation parameter. The positive side, where Leu was detected at higher concentrations and showed a greater propensity to form an alpha helix, is shown in Fig. 2a. On the negative side, Pro was more common and showed less tendency. Phe is more likely to form the beta sheet and is observed more frequently on the positive side of Fig. 2b. Pro was more prominent and tended to have less slope, which was problematic. On the positive side of Fig. 2c, Gly was found more frequently and has a higher probability of generating beta turns. As a result, Trp were more widespread and tended to be less negative. Ala were seen more frequently and tended to form pi-helixes more frequently in Fig. 2d. On the other hand, Cys, Met and Pro were more ubiquitous and had a lesser chance. Pro were identified more frequently and have a higher likelihood to form 3/10-helix in Fig. 2e positive side. On the negative side, Ile were more prevalent and had a smaller propensities. Trp were more generally observed in the Fig. 2f positive side, which also has a greater propensity to generate

isolated beta-bridges. So as on the negative side Cys were existed more and have small such tendency. Fig. 2g's positive side region, which includes Cys was more prominent and has a higher propensity to form random structures. On the negative side, Ile were more widespread and had a lower tendency.

The coiled-coil protein structure of ribovirus was studied using the DSSP database. The result shows the distribution of amino acid residues in secondary structural elements. In addition, the normalized deviation parameter values as well as the preferential and non-preferential amino acid residues of the secondary structural elements are determined from the plot. The present work provides information about the ribovirus coiled-coil protein, which is used by experimental biologists in drug development.

I would like to acknowledge my professor, Dr. S. Arul Mugilan, for his support during this work.

References :

1. Ainsley A. Mcfarlane, L. George Orriss, and Jorg Stetefeld, (2009) *European Journal of Pharmacology, 625:* 101–107.
2. Arul Mugilan, Jemimah Sherlyn and Preethi Jennifer, (2014). *Trends in Bioinformatics., 7*(1): 1-6
3. Bruce Yu. Y., (2002). *Advanced Drug Delivery Reviews., 54:* 1113–1129.
4. Fodje MN, and S. Al-Karadaghi, (2002). *Protein Eng., 15*(5): 353-358.
5. Frishman D, and P. Argos, (1995). *Proteins., 23*(4): 566-579.
6. Hooft R.W.W., C. Sander, M. Scharf and G Vriend, (1996) *iscb.org., 12:* 525-529.
7. Jan Zacharias and Ernst-Walter Knapp, (2014) *J. Chem. Inf. Model., 54:* 2166-2179.
8. Juliette Martin, Guillaume Letellier, Antoine Marin, Jean-François Taly, Alexandre G De Brevern and Jean-François Gibrat, (2005) *BMC Structural Biology., 5:* 17.
9. Kabsch W, Sander, (1983). *Biopolymers., 22*(12): 2577-637.
10. King SM and WC Johnson (1999) *Proteins 3*(35): 313-320.
11. Klose D. P., B. A Wallace and Janes Robert W. (2010) *Bioinformatics., 26*(20): 2624–2625.
12. Labesse G, N Colloc'h, J Pothier, and JP Mornon, and P-Sea (1997) *Comput Appl Biosci., 13*(3): 291-5.
13. Maria Vittoria Cubellis, Fabien Cailliez and Simon C Lovell, (2005) *BMC Bioinformatics., 6*(Suppl 4): S8.
14. Mugilan S.A. and Veluraja, (2000) *Indian Acad. Sci., 25:* 81–91.
15. Robbie P. Joosten, A.H. Tim TE Beek, Elmar Krieger, Maarten L. Hekkelman, Rob W.W. Hooft, Reinhard Schneider, Chris Sander5 and Gert Vriend, (2011) *Nucleic Acids Research., 39:* D411–D419.
16. Sammy, Khalife, Therese Malliavin and Leo Liberti, (2021) Secondary structure assignment of proteins in the absence of sequence information, *Bioinformatics Advances., 1 :* 1–8.

17. Richards FM, and CE Kundrot, (1988) *Proteins., 3*(2): 71-84.
18. Sklenar H, C Etchebest, and R Lavery, (1989) *Proteins., 6:* 46-60.
19. Sonya M. King and W. Curtis Johnson, (1999) *PROTEINS: Structure, Function, and Genetics., 35:* 313–320.
20. Susan Costantini, Giovanni Colonna, and Angelo M. Facchiano (2006) *Biochemical and Biophysical Research Communications., 342:* 441–451.
21. Terrence Sanvictores, and Fabiola Farci, (2023) *Primary Protein Structure, Biochemistry.*
22. Yuan Zhanga, and Celeste Sagui, (2015) *Journal of Molecular Graphics and Modelling., 55:* 72–84.