Clinical and Experimental Health Sciences

The Frequencies of Amino Acids in Secondary Structural Elements of Globular Proteins

Cevdet Nacar

Marmara University, School of Medicine, Department of Biophysics, İstanbul, Türkiye.

Correspondence Author: Cevdet Nacar
E-mail: cevnacar@marmara.edu.tr
Received: 19.01.2023 **Accepted:** 06.02.2023

ABSTRACT

Objective: The frequencies of amino acids in proteins for different structural levels have been determined by many studies. However, due to the different content of data sets, findings from these studies are inconsistent for some amino acids. This study aims to eliminate the contradictions in the findings of the studies by determining the frequencies of the amino acids in all structural level of globular proteins.

Methods: The frequencies of the amino acids in overall protein, in secondary structural elements (helix, sheet, coil) and in subtypes of secondary structural elements (α-, π-, and 3_{10}-helices, and first, parallel and anti-parallel strands) were calculated separately using a data set including 4.882 dissimilar globular peptides. The frequencies of the amino acids were calculated as the ratio of the total number of a specific residue in related structure to the total number of all residues in the related structure.

Results: The frequencies of residues determined in this study is partially in consistent with the other studies. The differences are probably due to the data set contents of the studies. The frequencies of the amino acids in subtypes of secondary structural elements were determined for the first time in this study.

Conclusion: Variations in the frequencies of PRO residue in 3_{10}-helix structure and of ILE, LEU, and VAL residues in strands of sheet structure are valuable findings for the improvement of secondary structure prediction methods, as they can be used as secondary structural elements markers.

Keywords: amino acid, globular protein, secondary structure

1. INTRODUCTION

The frequencies of amino acids in proteins could provide important information about the nature of the proteins and this information can be used in many application areas of proteomics, molecular biology and bioinformatics. Because of this, the frequencies of amino acids for both overall protein structures (1-18) and specific peptide structures (19-39) are constantly being investigated. Despite the partial consistency in the results of these studies, there are some issues related to the data set and methodology that need to be clarified and improved.

Firstly, data set used in the studies must be homogenous in regard of protein main class (i.e., globular, membrane, and fibrous proteins). It must include only one type of protein class. Otherwise, the information on amino acid frequencies will be diluted and will not reflect the true nature of the residue abundance, as each protein class has very different amino acid content depending on the different sequence and structural characteristics of the class. Likewise, findings from studies based on specific species (10, 13, 16, 17) are not conclusive in this regard. In order to attain a more qualified level of homogeneity, proteins of extremophile organisms should also be removed from the data set.

Secondly, studies also require layering of data set at the level of subtypes of secondary structural elements of proteins. That is, frequencies of the amino acids in proteins should be determined for subtypes of helix (such as, α-, π-, and 3_{10}-helices) and of sheet (such as, first, parallel and anti-parallel strands) structures. This approach will provide further information about the residue frequencies. As stated in previous paragraph, this layering method will also prevent the information about frequencies from being masked which caused by subtype heterogeneity.

Third issue concerns the similarity of the proteins in data set. Similar proteins mostly have similar structure and residue content. Thus, similar proteins change the weights of the residue numbers towards their residue numbers, resulting

DOI: 10.33808/clinexphealthsci.1239176

in bias in residue frequencies. Since proteins with less than 25% similarity considered dissimilar, the similarity of proteins in data set should not exceed this value. Because the whole genome of a species includes many similar proteins (e.g. in human genome, there are many similar G-Protein-Coupled-Receptors [GPCR] and Guanine Nucleotide Binding Proteins [G proteins]), this issue especially needs to be taken into account in studies using the whole genome of a particular species.

This study investigates the frequencies of amino acids in globular proteins at different structural levels and aims to contribute to resolve the inconsistencies in the findings obtained from different studies. For these purposes, a comprehensive and qualified data set which were obtained from Nacar (40) was used in this study.

2. METHODS

2.1. Protein Data Set

Protein chain data set was obtained from Nacar (40). It contains 4,882 protein chains, all of which are globular peptides. Also, it does not include any protein of extremophile organisms. Protein structure files were downloaded from Protein Data Bank (41). The lengths of peptide chains are longer than 80 residues. The sequence identity of any peptide pair is smaller than 25%, ensuring that the peptides are not similar.

2.2. Peptide Sequences

Amino acid sequences of peptides were obtained from the ATOM or HETATM entries of the Protein Data Bank (PDB) files of the peptides. Modified residues were replaced with original residues by referring to MODRES entry and expression tags were removed using SEQADV entry data. In this study, amino acids were represented using the three-letter-code system of IUPAC-IUB.

2.3. Secondary Structural Elements

The regions corresponding to secondary structural elements of peptides were determined according to the data deposited in HELIX and SHEET entries of the PDB files of the peptides. The regions other than helix and sheet were considered random coil or loop.

2.4. Frequencies of Amino Acids

The frequencies of amino acids were determined for overall peptides and, for each helix and sheet types and their subtypes (α-helix, 3_{10}-helix and first, parallel, anti-parallel strands) separately. The frequencies of the amino acids were calculated as the ratio of the total number of a specific residue in related structure to the total number of all residues in the related structure.

3. RESULTS

3.1. Distribution of Lengths of Peptides

Length distribution of the peptides were shown in Figure 1. Lengths of peptides have a great variety, but most of them lay in range of 80-to-500 residues, roughly. Peptides of 1,000 residues or longer were group into a single group as the last item of the figure.

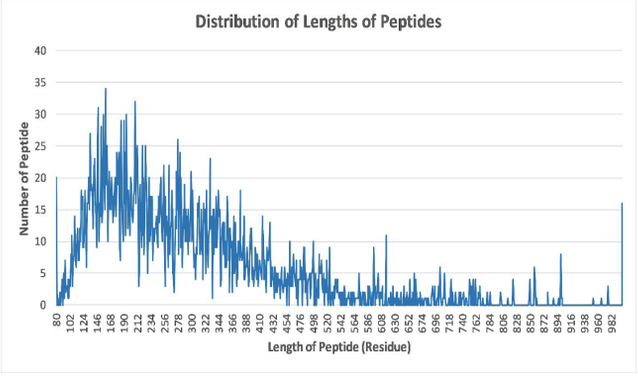

Figure 1. *Distribution of lengths of peptides. Most peptides are less than 500 residues in length. Peptides of 1.000 residues or longer were group into a single item.*

3.2. Distribution of Secondary Structural Elements

4.882 protein chains contain 1,419,498 amino acids. Of these amino acids, 613,334 (43.2 %) are located in the helix region, 344,676 (24.3 %) in the sheet region, and 461,488 (32.5 %) in the coil region. Also, percentages of secondary structural elements for each peptide of 4,882 chains were calculated and shown in Table 1 as mean value (Mean) and standard deviation (StdD). Despite the large standard deviation values, the latter values are in consistent with the former ones.

Table 1. *Distribution of secondary structural elements*

	Secondary structure	Mean (%)	StdD (%)
Helix	Overall	42.9	20.6
	α-Helix	37.8	20.6
	310-Helix	5.2	3.4
Sheet	Overall	22.6	13.9
	First Strand	4.0	3.3
	Parallel Strand	13.7	12.8
	Antiparallel Strand	4.8	5.4
	Coil	34.4	10.4

The dominant secondary structural element in globular proteins is helix structure and α-helix is the dominant subtype in helix. Second most abundant element is random coil. The less abundant one is the sheet structure. In sheet structure, parallel strand is the most abundant subtype.

Table 2. *Amino acid frequencies of proteins and secondary structural elements*

Amino acid	Protein (%)	Helix (%)			Sheet (%)				Coil (%)
	Overall	Overall	α	310	Overall	First	Parallel	Anti-Parallel	Overall
ALA	8.3	10.6	11.0	7.9	6.5	5.7	7.0	6.5	6.4
ARG	5.1	5.7	5.9	4.7	4.7	5.1	3.5	5.1	4.6
ASN	4.3	4.0	3.9	5.0	2.5	2.7	2.6	2.5	5.8
ASP	6.0	6.0	5.7	8.3	3.3	3.6	3.2	3.2	8.1
CYS	1.4	1.2	1.2	1.3	1.7	1.6	1.7	1.8	1.3
GLN	3.7	4.5	4.6	3.8	2.9	3.5	1.9	3.1	3.3
GLU	6.4	8.0	8.2	7.3	4.5	5.6	3.2	4.6	5.6
GLY	7.3	4.9	4.7	6.7	4.9	4.1	5.1	5.0	12.3
HIS	2.5	2.4	2.3	2.8	2.5	2.7	2.4	2.5	2.6
ILE	5.5	5.1	5.4	3.6	9.4	8.8	12.4	8.4	3.1
LEU	9.2	11.0	11.3	8.9	10.4	8.9	11.8	10.2	5.9
LYS	5.4	6.0	6.1	5.4	4.2	4.9	3.1	4.4	5.6
MET	2.1	2.5	2.6	1.9	2.2	2.0	2.4	2.2	1.6
PHE	4.2	4.0	3.9	4.3	6.0	5.9	5.9	6.1	3.0
PRO	4.8	2.8	2.2	6.9	2.0	2.6	1.4	2.0	9.6
SER	6.1	5.9	5.6	7.5	5.3	5.9	4.5	5.4	7.0
THR	5.6	4.7	4.8	4.4	6.7	7.1	6.0	6.8	5.9
TRP	1.5	1.5	1.5	1.9	2.0	1.9	1.5	2.2	1.2
TYR	3.6	3.4	3.4	3.7	5.1	4.9	4.2	5.5	2.7
VAL	7.0	5.6	5.8	3.8	13.2	12.5	16.1	12.3	4.2

3.3. Amino Acid Frequencies in Overall Proteins and Secondary Structural Elements

Frequencies of the amino acids are given in Table 2. Frequencies of amino acids were determined for overall protein, for secondary structure types (helix, sheet, and coil), and for subtypes of secondary structural elements (α-helix, 3_{10} – helix or first, parallel and anti-parallel strands of sheet), separately.

The most abundant amino acids in overall protein and overall helix structures are LEU and ALA residues. The most abundant residues in α-helix are LEU and ALA, while in 3_{10}-helix structure are LEU and ASP. The most abundant amino acids in sheet structure and in its subtypes are VAL and LEU residues. In parallel strand, there is also an abundance of ILE. In coil, the most abundant amino acids are GLY and PRO. Since the PRO residue is not abundant in helix and sheet structures, the abundance of it in coil is quite remarkable. However, this finding implies that PRO residue is excluded from the helix and sheet structures due to its inability to form backbone hydrogen bond.

The least abundant amino acids in overall protein, helix and α-helix structures are TRP and CYS residues. Besides those, MET is an another least abundant residue in 3_{10}-helix structure. Because PRO residue lacks of free amino group, it cannot form backbone hydrogen bond with carboxyl group of other amino acids. Therefore, it is not expected to be found in helix and sheet structures abundantly. However, it is noteworthy that PRO has been found 3.1 times more abundant is 3_{10}-helix structure. The least abundant amino acids in sheet structure are TRP, PRO and CYS. The least abundant residues in coil structure are CYS and TRP.

3.4. Abundance of Modified Residues

Of 1,419,498 amino acids, 4,192 are chemically modified and number of modified amino acids were presented in Table 3. While MET, ASN and CYS amino acids are the most modified residues, GLN, ILE and VAL residues have no modification.

Table 3. *Number of modified residues*

Amino Acid	#	Amino Acid	#
ALA	4	LEU	1
ARG	1	LYS	86
ASN	223	MET	3663
ASP	2	PHE	1
CYS	104	PRO	3
GLN	0	SER	27
GLU	22	THR	29
GLY	2	TRP	6
HIS	9	TYR	9
ILE	0	VAL	0

4. DISCUSSION

Frequencies of amino acids from nine studies (1-9) and this study tabulated in Table 4 for comparison. Because the frequencies were not given in quantitative values (given only in bar graphs), finding from other studies (10-12, 14-18) were not included in Table 4. Protein databases used by 1, 2, 3, 4, 6, 8, and 9 are UnitProtKB, NCBI+KEGG, PDB, PDB, SwissProt, OWL and SwissProt+PDB, respectively. Protein databases used by 5 and 7 were not specified. Amino acid frequencies in 4 were calculated from the *Table I* from Xia and Xie (9). While study 1 included only globular proteins, the other studies included various organisms/protein classes or they did not specify the protein class.

Table 4. *Comparison of frequencies of amino acids from different sources (%)*

	Nacar	1	2*	3	4	5	6	7	8	9SP	9PDB
ALA	8.3	7.8	11.1	8.3	11.3	8.2	7.6	7.4	7.5	7.9	7.7
ARG	5.1	4.7	4.1	4.7	4.9	5.2	5.2	4.2	5.2	5.4	4.9
ASN	4.3	4.1	5.0	4.8	3.6	4.4	4.4	4.4	4.6	4.1	4.6
ASP	6.0	5.1	5.0	5.9	4.9	5.3	5.3	5.9	5.2	5.4	5.8
CYS	1.4	1.6	0.4	1.5	1.5	1.1	1.6	3.3	1.8	1.5	1.7
GLN	3.7	3.8	5.9	3.7	3.8	3.6	3.9	3.7	4.1	4.0	4.0
GLU	6.4	7.1	5.9	6.1	6.5	6.5	6.5	5.8	6.3	6.7	6.7
GLY	7.3	6.1	16.3	7.9	5.5	6.9	6.9	7.4	7.1	7.0	7.2
HIS	2.5	2.3	1.9	2.2	2.2	2.1	2.2	2.9	2.2	2.3	2.4
ILE	5.5	5.5	3.5	5.5	6.1	6.8	5.9	3.8	5.5	5.9	5.6
LEU	9.2	10.8	5.7	8.4	9.8	10.1	9.5	7.6	9.1	9.7	8.7
LYS	5.4	5.4	2.1	5.8	5.9	6.0	6.0	7.2	5.8	5.9	6.4
MET	2.1	2.4	0.5	2.1	3.3	2.3	2.4	1.8	2.8	2.4	2.2
PHE	4.2	4.5	3.4	4.0	4.1	4.3	4.1	4.0	3.9	4.0	4.0
PRO	4.8	4.7	2.1	4.7	2.7	4.3	4.9	5.0	5.1	4.8	4.6
SER	6.1	7.7	11.8	6.1	5.2	6.5	7.1	8.1	7.4	6.8	6.2
THR	5.6	4.9	5.7	5.9	5.4	5.3	5.6	6.2	6.0	5.4	5.6
TRP	1.5	1.3	0.2	1.5	1.5	1.1	1.2	1.3	1.3	1.1	1.4
TYR	3.6	3.6	2.7	3.7	3.6	3.3	3.2	3.3	3.3	3.0	3.5
VAL	7.0	6.5	5.6	6.9	7.9	6.9	6.6	6.8	6.5	6.7	6.7

1 – Tripathi, Tripathi, Gupta 2014
2 – Moura, Savageau, Alves 2013
**ASN/ASP and GLN/GLU frequencies recalculated*
3 – Baud and Karlin 1999
4 – Xia and Xie 2002
5 – Itzkovitz and Alon 2022
6 – Varfolomeev, Uporov, Fedorov 2002
7 – King and Jukes 1969
8 – Trinquier and Sanejouand 1998
9 – Vacic, Uversky, Dunker, Lonardi 2007
SP: SwissProt
PDB: Protein Data Bank

Evaluating the Table 4 as a whole, it is observed that the frequencies of the ARG, ASN, ASP, GLU, HIS, PHE, THR, TYR, and VAL residues are nearly the same. The frequencies of remaining residues almost are the same except those all remaining residues of 2, ALA, GLY and PRO residues of 4, and CYS, ILE, LEU residues of 7.

Because study 1 based on globular proteins, its findings are comparable to this study. Despite the different data set sizes (study 1 included 557 peptides while Nacar (40) included 4,882 peptides), findings of these two studies are highly consistent, except for residues ASP, GLY, LEU and SER. The inconsistencies in frequencies for these residues are no more than 25%. Therefore, findings of this study contribute to the frequencies of ASP, GLY, LEU and SER amino acids in globular proteins.

The frequencies of amino acids in secondary structural elements of proteins from two studies (1, 9) and this study given in Table 5. Amino acid frequencies in 1 and 2 were calculated using data provided by Xia and Xie (9) and Baud and Karlin (1), respectively. Frequencies of residues in helix are almost completely consistent. But, in sheet and coil structures, inconsistency prevails. In sheet structure, the remarkable differences exist in frequencies of ILE, LEU and VAL residues. In the coil region of the peptide, most of the residues, except CYS, HIS, MET, PHE, TRP, TYR, and VAL, are conspicuous in regard of frequency changes, especially ALA, ASP, GLY, and PRO.

Table 5. *Comparison of amino acid frequencies of secondary structural elements from different sources*

	Overall (%)			Helix (%)			Sheet (%)			Coil (%)	
Amino Acid	Nacar	1	2	Nacar	1	2	Nacar	1	2	Nacar	1
ALA	8.3	11.3	8.3	10.6	11.2	11.7	6.5	6.8	4.2	6.4	46.4
ARG	5.1	4.9	4.7	5.7	5.6	5.7	4.7	4.1	2.8	4.6	2.0
ASN	4.3	3.6	4.8	4.0	3.9	3.8	2.5	3.2	1.9	5.8	3.7
ASP	6.0	4.9	5.9	6.0	5.4	5.3	3.3	4.0	2.0	8.1	4.8
CYS	1.4	1.5	1.5	1.2	1.3	1.1	1.7	2.0	1.3	1.3	1.0
GLN	3.7	3.8	3.7	4.5	4.3	4.7	2.9	3.1	1.9	3.3	1.8
GLU	6.4	6.5	6.1	8.0	7.7	8.6	4.5	4.8	2.9	5.6	3.3
GLY	7.3	5.5	7.9	4.9	5.0	4.0	4.9	6.1	3.3	12.3	7.6
HIS	2.5	2.2	2.2	2.4	2.1	2.0	2.5	2.4	1.5	2.6	1.3
ILE	5.5	6.1	5.5	5.1	5.2	5.6	9.4	8.3	6.3	3.1	1.7
LEU	9.2	9.8	8.4	11.0	10.5	11.1	10.4	9.2	6.3	5.9	3.4
LYS	5.4	5.9	5.8	6.0	6.6	6.6	4.2	4.9	3.0	5.6	3.2
MET	2.1	3.3	2.1	2.5	3.6	2.7	2.2	3.0	1.5	1.6	1.6
PHE	4.2	4.1	4.0	4.0	3.8	4.0	6.0	5.0	3.7	3.0	1.4
PRO	4.8	2.7	4.7	2.8	2.6	2.4	2.0	2.8	1.3	9.6	3.7
SER	6.1	5.2	6.1	5.9	5.1	5.0	5.3	5.7	3.3	7.0	4.3
THR	5.6	5.4	5.9	4.7	4.8	4.4	6.7	6.8	4.8	5.9	3.8
TRP	1.5	1.5	1.5	1.5	1.4	1.6	2.0	1.8	1.3	1.2	0.7
TYR	3.6	3.6	3.7	3.4	3.3	3.6	5.1	4.6	3.5	2.7	1.6
VAL	7.0	7.9	6.9	5.6	6.5	6.0	13.2	11.2	8.5	4.2	2.6

1 – Xia and Xie 2002
2 – Baud and Karlin 1999

Since the protein classes in the studies of 1 and 2 were not specified, these comparisons are valid to some extent. Despite this drawback, findings of this study are more reliable because of the comprehensive and qualified data set used in the study.

CYS and TRP amino acids are the least abundant residues in nearly all protein secondary structure. A biophysical explanation of this finding is probably based on functional and structural characteristics of the residues. The main function of CYS residue in protein is to establish disulfide bond between two CYS residues. This bond is an extremely

important biophysical determinant in stability of the peptide. However, the number of disulfide bonds in protein is very limited. Therefore, low frequency of the CYS in protein reflects this fact. TRP residue has an indole ring as side chain and this side chain remains very large compared to the side chains of the other amino acids. This spatial property of TRP cause difficulties in positioning of it in protein structure. This is the possible biophysical reason for the low frequency of TRP in the protein.

Despite there are few studies on the frequencies of amino acids in secondary structural elements of protein (1, 9), there is not any study on the frequencies of the residues in subtypes of secondary structural elements (i.e., α-, π-, 3_{10}-helix or first, parallel, anti-parallel strands of sheet structure). Therefore, residue frequencies in subtypes of secondary structural elements were determined for the first time in this study and presented in Table 2. Although the frequencies of most of the residues do not vary with subtype, it is worth discussing a few them. The frequency of PRO residue in 3_{10}-helix is 3,1 times higher than in α-helix. This remarkable increase is probably again related to its backbone hydrogen bond forming capability. 3_{10}-helix is a loose and short helix type. Its length generally spans several amino acids and backbone hydrogen bonds are formed between (n):(n+5) residues. Since the length is so short, most of the residues in 3_{10}-helix cannot form a backbone hydrogen bond due to the (n):(n+5) pattern restriction. Therefore, the reduction in the requirement of backbone hydrogen bond formation in 3_{10}-helix removes the biophysical barriers to PRO's presence in the helix and its frequency in 3_{10}-helix increases. The frequencies of ILE, LEU and VAL residues are higher in parallel strand than in first and anti-parallel strands. The frequency differences of the other residues are negligible. While the insights from the findings of this part are relatively minor, they are valuable for improving the understanding of the biophysical characteristics of the secondary structure of the protein. Besides, these findings may provide important information for secondary structure prediction algorithms. The differences in frequencies of these residues depending on subtypes can make them markers of subtypes and can be used in prediction algorithms.

5. CONCLUSION

The findings of this study are more conclusive than other studies due to the comprehensive and distinctive data set used in the study. The data set only includes globular proteins (excludes the extremophile proteins) with a similarity less than 25%. So, the results obtained from this study can eliminate the contradictions in amino acid frequencies in the scientific literature. In addition, the fact that the frequency variations of amino acids according to the subtypes of secondary structural elements were investigated for the first time in this study makes the findings of this study valuable in the field of research. Therefore, these findings may contribute the understanding of the structural features of protein secondary structure and to the improvement of the secondary structure prediction algorithms.

REFERENCES

[1] Baud F, Karlin S. Measures of residue density in protein structures. Proceedings of the National Academy of Sciences of the United States of America. 1999; 96(22): 12494-9.

[2] Itzkovitz S, Alon U. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Res. 2007; 17(4): 405-12.

[3] King JL, Jukes TH. Non-Darwinian evolution. Science. 1969; 164(3881): 788-98.

[4] Moura A, Savageau MA, Alves R. Relative Amino Acid Composition Signatures of Organisms and Environments. Plos One. 2013; 8(10).

[5] Trinquier G, Sanejouand YH. Which effective property of amino acids is best preserved by the genetic code? Protein Engineering. 1998; 11(3): 153-69.

[6] Tripathi V, Tripathi P, Gupta D. Statistical approach for lysosomal membrane proteins (LMPs) identification. Syst Synth Biol. 2014; 8(4): 313-9.

[7] Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. BMC Bioinformatics. 2007; 8: 211.

[8] Varfolomeev SD, Uporov IV, Fedorov EV. Bioinformatics and molecular modeling in chemical enzymology. Active sites of hydrolases. Biochemistry (Mosc). 2002; 67(10): 1099-108.

[9] Xia X, Xie Z. Protein structure, neighbor effect, and a new index of amino acid dissimilarities. Mol Biol Evol. 2002; 19(1): 58-67.

[10] Bogatyreva NS, Finkelstein AV, Galzitskaya OV. Trend of amino acid composition of proteins of different taxa. J Bioinform Comput Biol. 2006; 4(2): 597-608.

[11] Dyer KF. The Quiet Revolution: A New Synthesis of Biological Knowledge. Journal of Biological Education. 1971; 5: 15-24.

[12] Fagerlund A, Myrset AH, Kulseth MA. Construction and characterization of a 9-mer phage display pVIII-library with regulated peptide density. Appl Microbiol Biotechnol. 2008; 80(5): 925-36.

[13] Gaur RK. Amino acid frequency distribution among eukaryotic proteins. IIOAB Journal. 2014; 5(2): 6-11.

[14] Lehmann J. Genetic code degeneracy and amino acid frequency in proteomes. Grandcolas P, Maurel M-C, editors: Elsevier; 2018.

[15] Rao Y, Wang Z, Luo W, Sheng W, Zhang R, Chai X. Base composition is the primary factor responsible for the variation of amino acid usage in zebra finch (Taeniopygia guttata). PLoS One. 2018; 13(12): e0204796.

[16] Switzar L, Giera M, Niessen WM. Protein digestion: an overview of the available techniques and recent developments. J Proteome Res. 2013; 12(3): 1067-77.

[17] Tian L, Liu SJ, Wang S, Wang LS. Ligand-binding specificity and promiscuity of the main lignocellulolytic enzyme families as revealed by active-site architecture analysis. Sci Rep-Uk. 2016; 6.

[18] Tsuji J, Nydza R, Wolcott E, Mannor E, Moran B, Hesson G, et al. The frequencies of amino acids encoded by genomes that utilize standard and nonstandard genetic codes. Bios. 2010; 81(1): 22-31.

[19] Akashi H, Gojobori T. Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99(6): 3695-700.

[20] Berezovsky IN, Kilosanidze GT, Tumanyan VG, Kisselev LL. Amino acid composition of protein termini are biased in different manners. Protein Engineering. 1999; 12(1): 23-30.

[21] Bouziane H, Chouarfia A. Sequence – and structure-based prediction of amyloidogenic regions in proteins. Soft Comput. 2020; 24(5): 3285-308.

[22] Brooks DJ, Fresco JR, Lesk AM, Singh M. Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into the genetic code. Mol Biol Evol. 2002; 19(10): 1645-55.

[23] Brune D, Andrade-Navarro MA, Mier P. Proteome-wide comparison between the amino acid composition of domains and linkers. BMC Res Notes. 2018; 11(1): 117.

[24] Carugo O. Amino acid composition and protein dimension. Protein Sci. 2008; 17(12): 2187-91.

[25] dos Reis M, Yang ZH. Why Do More Divergent Sequences Produce Smaller Nonsynonymous/Synonymous Rate Ratios in Pairwise Sequence Comparisons? Genetics. 2013; 195(1): 195-204.

[26] Du MZ, Liu S, Zeng Z, Alemayehu LA, Wei W, Guo FB. Amino acid compositions contribute to the proteins' evolution under the influence of their abundances and genomic GC content. Sci Rep-Uk. 2018; 8.

[27] Flores SC, Lu LJ, Yang JL, Carriero N, Gerstein MB. Hinge Atlas: relating protein sequence to sites of structural flexibility. Bmc Bioinformatics. 2007; 8.

[28] Ganguli S, Datta A. Residue frequencies and conserved phylogenetic signatures in amino acid sequences of plant glutathione peroxidases, indicates habitat specific adaptation and dictates interactions with key ligands. American Journal of Bioinformatics Research. 2015; 5(1): 9-15.

[29] Gardini S, Cheli S, Baroni S, Di Lascio G, Mangiavacchi G, Micheletti N, et al. On Nature's Strategy for Assigning Genetic Code Multiplicity. Plos One. 2016; 11(2).

[30] Hormoz S. Amino acid composition of proteins reduces deleterious impact of mutations. Sci Rep. 2013; 3: 2919.

[31] Ilardo M, Bose R, Meringer M, Rasulev B, Grefenstette N, Stephenson J, et al. Adaptive Properties of the Genetically Encoded Amino Acid Alphabet Are Inherited from Its Subsets. Sci Rep. 2019; 9(1): 12468.

[32] Jackson EL, Ollikainen N, Covert AW, 3rd, Kortemme T, Wilke CO. Amino-acid site variability among natural and designed proteins. PeerJ. 2013; 1: e211.

[33] Karlin S, Brocchieri L, Bergman A, Mrazek J, Gentles AJ. Amino acid runs in eukaryotic proteomes and disease associations. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99(1): 333-8.

[34] Liu J, Bu CP, Wipfler B, Liang AP. Comparative Analysis of the Mitochondrial Genomes of Callitettixini Spittlebugs (Hemiptera: Cercopidae) Confirms the Overall High Evolutionary Speed of the AT-Rich Region but Reveals the Presence of Short Conservative Elements at the Tribal Level. Plos One. 2014; 9(10).

[35] Mbaye MN, Hou Q, Basu S, Teheux F, Pucci F, Rooman M. A comprehensive computational study of amino acid interactions in membrane proteins. Sci Rep. 2019; 9(1): 12043.

[36] McNair K, Ecale Zhou CL, Souza B, Malfatti S, Edwards RA. Utilizing Amino Acid Composition and Entropy of Potential Open Reading Frames to Identify Protein-Coding Genes. Microorganisms. 2021; 9(1).

[37] Tekaia F, Yeramian E, Dujon B. Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis. Gene. 2002; 297(1-2): 51-60.

[38] Wang HC, Li K, Susko E, Roger AJ. A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny. BMC Evol Biol. 2008; 8: 331.

[39] Zalucki YM, Power PM, Jennings MP. Selection for efficient translation initiation biases codon usage at second amino acid position in secretory proteins. Nucleic Acids Res. 2007; 35(17): 5748-54.

[40] Nacar C. Propensities of Amino Acid Pairings in Secondary Structure of Globular Proteins. Protein J. 2020; 39(1): 21-32.

[41] Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nat Struct Biol. 2003; 10(12): 980.

How to cite this article: Nacar C. The Frequencies of Amino Acids in Secondary Structural Elements of Globular Proteins . Clin Exp Health Sci 2023; 13: 261-266. DOI: 10.33808/clinexphealthsci.1239176