IJBPAS, February, 2022, 11(2): 518-525

ISSN: 2277–4998

International Journal of Biology, Pharmacy
and Allied Sciences (IJBPAS)
'A Bridge Between Laboratory and Reader'

www.ijbpas.com

COMPREHENSIVE ANALYSIS OF CORONA VIRUS PROTEINS AT ATOMIC LEVEL AND THEIR ASSESSMENT WITH THE PROTEINS OF OTHER SINGLE STRANDED RNA VIRUSES

JOHRI P

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus,
Malhaur, Gomti Nagar Extension, Lucknow, Uttar Pradesh, India

***Corresponding Author: Parul Johri: E Mail: pjohri@lko.amity.edu; +91 9838144680**

Received 15th March 2021; Revised 16th April 2021; Accepted 30th May 2021; Available online 1st Feb. 2022

https://doi.org/10.31032/IJBPAS/2022/11.2.5863

ABSTRACT

The only difference between viral protein is the way their amino acids are being aligned in the sequence. Although all the proteins are made up of same amino acids but there is a difference in their arrangements. In order to study this difference, we need to compare the sequence not only at their amino acid but also at their atomic level. Atom is the lowest level of protein sequence next to electrons. Sequence analysis at atom level can reveal insights for their biased atom content levels. In this study, various single stranded viruses are compared at their atomic level, considering carbon as the main element and their preferred atomic percentage has been calculated using in silico tools. After comparison of protein at their atom level one can concluded about their likelihood for attaining a particular percentage, which could be further utilized for protein classification analysis.

Keywords: Atomic, Carbon, Protein, In silico, Algorithm

INTRODUCTION

The folding of a protein totally depends on its primary structure. Once a researcher identifies or isolates a novel protein, the next step is to identify the protein. In order to identify the protein, one need to have the primary sequence of the protein which could be achieved by protein sequencing techniques. Many features, parameters and properties could be derived from the protein primary sequence **[1, 2]**. The protein sequence derived after protein sequencing is just a single letter amino acid code for all the amino acids present inthat protein. We can compare the sequences of these proteins to find similarities or difference and use these similarities and differences for inferring function, three dimensional structures or binding site analysis. There are many parameters which could be calculated form the primary sequence of the proteins, one of the easiest being the frequency of any amino acid. A lot more thing could be done with protein sequence like sequence comparison global or local, amino acid compositions, calculation of their molecular weights, pI values, secondary and tertiary structure prediction and so on. All these analyses basically compare one amino acid with another. In the present work we tried to study the protein sequences of various single stranded RNA viruses at their atomic levels. As all the amino acids are composed of only five atoms – carbon, nitrogen, hydrogen, sulphur and oxygen. These atoms constitute the basic building material for all the amino acids and hence the protein. As these atoms are the lowest level of sequences, we analyzed the proteins of human SARS corona virus and compared it with Human Respiratory syncytical virus (HRSV), Human metapneumovirus (HMPV), Human rhinovirus and Influenza virus A and C **[3-6]**. All these single stranded RNA viruses are responsible for common cold and throat infection. Amino acid composition of protein varies significantly among the various taxa. Protein sequences can vary greatly in their content of nitrogen, sulphur and carbon atoms and this elemental composition variation in the sequences can be influenced by the process of natural selection **[7-11]**. Understanding the difference in protein sequence at atomic level could be very useful in comparing sequence from different origin.

MATERIALS AND METHODS

The protein sequences from corona virus, influenza a, influenza c, human Syncytical virus, human metapneumovirus and human rhinovirus were derived and downloaded from the NCBI database in fasta sequence formats. The downloaded sequences were subjected to protparam analysis on expasy server for the atomic content calculation. The frequency and consecutively the percentage of all the atoms were recoded and saved in excel for further analysis.

Severe acute respiratory syndrome corona virus is species of corona virus that infect human, bats and certain other mammals. The virus has 15 proteins in total, their accession id and total atomic count with the percentage calculation is listed in **Table 1**.

Influenza A virus causes common viral infection that can be deadly especially in high risk groups that are young children and older adults. Influenza can be type A, B and C. The atomic content of the proteins of Influenza A and C is given in **Table 2 and Table 3**.

Human respiratory syncytial virus (HRSV) also known as human orthopneumovirus, is the principal viral agent of serious pediatric respiratory disease globally. It also is accountable for significant morbidity and mortality in the aging and in severely immunocompromised persons. HRSV is an encased non-segmented negative-strand RNA virus. The atomic composition of the proteins of this virus is listed in **Table 4**.

HMPV Human metapneumovirus (HMPV) is the source for upper and lower respiratory disease in people of all ages, particularly among young children, older adults, and people with deteriorated immune systems. Revealed in 2001, HMPV is in the paramyxovirus family lengthwise with respiratory syncytial virus (RSV).It is a respiratory viral pathogen that grounds spectrum of illness that array from asymptomatic contamination to bronchiolitis. The proteins of this virus were also subjected to atom level analysis and the results are detailed in **Table 5**.

The rhinovirus is the utmost communal viral infectious mediator in humans and is the chief cause of the common cold. Rhinovirus infection flourishes in temperatures of 33–35 °C, the temperatures found in the nose **[12-14]**. Rhinoviruses fit to the genus Enterovirus in the family Picornaviridae. The atomic content analysis of this protein is detailed in **Table 6**.

Table 1: Atomic content of Corona Virus

Protein name	accession id	Total atoms	carbon	Hydrogen	Nitrogen	Oxygen	Sulphur	carbon %	Hydrogen%	Nitrogen %	Oxygen %	sulphur %
Protein 3a	P59632	4342	1428	2161	343	397	13	32.8	49.7	7.8	9.1	0.2
Nucleoprotein	P59595	6393	1985	3150	618	633	7	31.04	49.2	9.6	9.9	0.1
Non-structural protein 3b	P59633	2596	794	1331	221	210	13	28.8	51.2	8.6	8.08	0.5
Replicase protein 1a	POC6U8	68155	21746	34030	5656	6468	255	31.9	49.9	8.2	9.4	0.3
Protein non-structural 7b	Q7TFA1	765	258	392	48	63	4	33.7	51.2	6.2	8.2	0.5
protein non-structural 8a	Q7TFAO	614	185	318	52	52	7	30.1	51.7	8.4	8.4	1.1
Protein 9b	P59636	1533	472	778	130	148	5	30.7	50.7	8.4	9.6	0.3
Non-structural protein6	P59634	1080	349	550	82	96	3	32.3	50.9	7.5	8.8	0.2
Non-structural protein 8b	Q80H93	1337	425	667	125	113	7	31.7	49.8	9.3	8.4	0.5
Replicase polyprotein 1b	POC6X7	110491	35380	55002	9262	10437	410	32.02	49.7	8.3	9.4	0.3
Protein 7a	P59635	1976	639	995	161	174	7	32.3	50.3	8.1	8.8	0.3
Uncharacterized protein 14	P7T267	1123	356	573	93	96	5	31.7	51.02	8.2	8.5	0.4
Membrane protein	P59596	3577	1155	1809	303	300	10	32.2	50.5	8.4	8.3	0.2
Envelope small membrane potein	P59637	1212	388	625	89	106	4	32.01	51.5	7.3	8.7	0.3
spike glycoprotein	P59594	19384	6252	9593	1609	1871	59	32.2	49.4	8.3	9.6	0.3

Table 2: Atomic Composition of protein in Influenza A Virus

Protein name of influenza a virus	accession id	Total atoms	carbon	Hydrogen	Nitrogen	Oxygen	Sulphur	carbon %	Hydrogen%	Nitrogen %	Oxygen %	sulphur %
Max protein 2	P06821	1552	493	776	134	144	5	31.7	50	8.6	9.2	0.3
PB2-S1	P0DOG6	8202	2541	4139	741	748	33	31.1	50.4	9.03	9.1	0.4
polymerase basic protein 2	P03428	12177	3785	6145	1089	1117	41	31.1	50.4	8.9	9.1	0.3
nuclear export protein	P03508	2020	635	1012	174	192	7	31.4	50.09	8.6	9.5	0.3
polymerase acidic protein	P03433	11547	3672	5750	980	1105	40	31.8	49.7	8.4	9.5	0.3
protein PA-X	P0CK64	4133	1303	2071	359	384	16	31.5	50.1	8.6	9.2	0.3
Nucleoprotein	P03466	7808	2426	3874	732	745	31	31.07	49.6	9.3	9.5	0.4
matrix protein 1	P03485	3934	1211	1987	353	366	17	30.7	50.5	8.9	9.3	0.4
Hemagglutin	P03452	8824	2810	4371	765	851	27	31.8	49.5	9.3	9.6	0.3
RNA directed RNA polymerase catalytic subunit	PO3431	12109	3803	6047	1075	1134	50	31.4	49.9	8.8	8.9	0.4
non structural protein 1	P03496	3658	1131	1846	326	344	11	30.9	50.4	8.9	9.4	0.3
Neuraminidase	P03468	6919	2215	3400	606	671	27	32.01	49.1	8.7	9.6	0.3
protein PB1-F2	P0C0UI	1486	467	749	141	124	5	31.4	50.4	9.4	8.3	0.3

Table 3: Atomic Composition of protein in Influenza C Virus

Protein name of influenza c	Accession no.	Total atom	Carbon	Hydrogen	nitrogen	Oxygen	sulphur	carbon%	hydrogen%	nitrogen%	oxygen %	sulphur%
Polyprotein p42	Q617B9	5883	1828	2972	504	550	29	31.07	50.5	8.56	9.34	0.5
Nuclear export protein	Q01640	2963	907	1516	244	284	12	30.6	51.1	8.2	9.5	0.4
Nucleoprotein	Q617C0	8999	2793	4532	824	827	23	31.03	50.3	9.1	9.1	0.2
polymerase acidic protein	Q617C2	11505	3660	5762	972	1066	45	31.8	50.08	8.4	9.2	0.3
polymerase basic protein 2	Q617C4	12410	3897	6246	1082	1146	39	31.4	50.3	8.7	9.2	0.3
hemagglutinin esterase fusion glycoprotein	P68762	10087	3200	5034	852	970	31	31.7	49.9	8.4	9.6	0.3
RNA directed RNA polymerase catalytic subunit	Q617C3	12122	3807	6105	1029	1126	55	31.4	50.03	8.4	9.2	0.5
Non structuralprotein 1	Q01639	3877	1222	1940	330	364	21	31.5	50.03	8.5	9.3	0.4

Table 4: Atomic Composition of protein in Human Syncytical Virus

Protein name HRS virus	Accession no.	Total atom	Carbon	Hydrogen	nitrogen	Oxygen	sulphur	carbon%	hydrogen %	nitrogen %	oxygen %	sulphur%
matrix protein	P0DOE6	4095	1287	2083	327	383	13	31.4	50.8	7.9	9.3	0.3
non structural protein 1	P0DOE8	2189	687	1103	179	209	11	31.3	50.3	8.1	9.5	0.5
small hyrdrophobic protein	P0DOE4	1083	351	553	83	93	4	32.4	51.1	7.6	8.4	0.3
nucleoprotein	O12395	6139	1931	3085	527	579	17	31.4	50.2	8.5	9.4	0.2
non structural protein 2	O12405	2088	671	1052	172	172	7	32.1	50.3	8.2	8.9	0.3
matrix m2	Q76TK9	3124	964	1572	282	282	8	30.8	50.3	9.02	9.5	0.2
RNA directed RNA poly L	Q76TK7	35544	11372	17895	2967	3224	86	31.9	50.3	8.3	9.07	0.2
major surface glycoprotein	O09634	4666	1436	2368	396	459	7	30.7	50.7	8.4	9.8	0.1
fusion glycoprotein	O09635	9008	2803	4572	746	861	26	31.1	50.7	8.2	9.5	0.2
Phosphoprotein	O09633	3764	1162	1862	320	413	7	30.8	49.4	8.5	10.9	0.1

Table 5: Atomic Composition of protein in Human Metapneumo Virus

Protein name of hMP virus	Accession no.	Total atom	Carbon	Hydrogen	nitrogen	Oxygen	sulphur	carbon%	hydrogen%	nitrogen%	oxygen %	sulphur%
Matrix M2-1	Q6WB97	2984	922	1494	274	287	7	30.8	50.1	9.1	9.6	0.2
Matrix protein M2	Q6WB96	1161	370	589	89	107	6	31.8	50.1	7.6	9.2	0.5
RNA directed RNA polymerase l	Q6WB95	32720	10379	16500	2766	2980	95	31.7	50.4	8.4	9.1	0.2
Fusion glycoprotein Fo	Q6WB98	8278	2583	4178	702	792	23	31.2	50.4	8.4	9.5	0.2
Matrix protein	Q6WB99	3950	1251	2012	306	370	11	31.6	50.9	7.7	9.3	0.2
Phosphoprotein	Q8B9Q8	4232	1417	2341	381	484	9	30.5	50.5	8.2	10.4	0.1
small hydrophoic protein	Q6WB95	2903	908	1460	244	280	11	31.2	50.2	8.4	9.6	0.3
major surface glucoprotein g	Q6WB94	3294	990	1638	303	355	5	30.05	49.7	9.2	10.7	0.1
Nucleoprotein	Q6WBA1	6152	1926	3092	526	596	12	31.3	50.2	8.5	9.6	0.1

Table 6: Atomic Composition of protein in Human Rhino Virus

Protein name of human rhinovirus	Accession no.	Total atom	Carbon	Hydrogen	nitrogen	Oxygen	Sulphur	carbon%	hydrogen%	nitrogen%	oxygen %	sulphur%
Genome Polyprotein	P07210	33783	10807	16818	2856	3210	92	31.9	49.7	8.4	9.4	0.2

RESULTS AND DISCUSSION

The proteins taken from Sars Coronavirus, Influenza (A and C), preferred to have a carbon percentage between 28.8 to 33.7The proteins taken from Human respiratory Syncytical virus, Human metapneumovirus, Human rhinovirus preferred to have a carbon percentage between 30.05 to 32.4. The present study shows the role of carbon content in viral protein analysis. Carbon being a major element in our body plays an important role in sequence analysis **(Table 7)**. Proteins have a significant level of carbon to maintain their virulence and this could be used for better understanding of their function and stability. Apart from carbon other atoms also need to be taken in account.

Table 7: The Maximum and Minimum Carbon percentage in various organisms

S.NO	ORGANISM	MAX CARBON %	MIN CARBON %
1	SARS CORONAVIRUS	33.7	28.8
2	INFLUENA A	32.01	30.7
3	INFLUENZA B	31.8	30.8
4	INFLUENZA C	31.8	30.6
5	HUMAN SYNCYTICAL VIRUS	32.4	30.7
6	HUMAN METAPNEUMOVIRUS	31.8	30.5
7	HUMAN RHINOVIRUS	31.9	31.9

CONCLUSION

Although all the proteins are made up of same twenty amino acids, the only difference lies in their position. Our nature tends to maintain a fixed level of atoms in a protein in order to main their functionality and nature. The above study on various single stranded RNA viral proteins concludes that the carbon composition is bracketed between a very closed interval for various species and is maintained across the families.

Acknowledgement

Authors are grateful to Dr. A. K. Chauhan, Founder President & Mr. Aseem Chauhan, Chancellor Amity University Haryana & Chairperson Amity Lucknow for providing necessary facilities and support. We also extend our gratitude to Prof. (Dr.) Sunil Dhaneshwar, Pro Vice Chancellor, Amity University, Uttar Pradesh Lucknow Campus for constant support and encouragements.

REFERENCES

[1] Johri, P. and Gokhale, M. A New Perspective for Sequence Analysis – Carbon content, 2013. Research and Review: Journal of Computational Biology, 2: 1-6.

[2] Rajasekaran, E., Akila, K. and Vennila, J.J, Carbon Contents of H1N1 Proteins, 2011. International Conference on Biosciences, Biochemistry and Bioinformatics, 5: 27-29.

[3] Akila, K. and Rajasekaran, E. What

Might be the Difference in Viral Proteins? 2009. International Journal of Bioinformatics Research, 1: 1-3.

[4] Rajasekaran, E., Asha, J. and Klaus, H. Magnitude of Thymine In Different Frames of Messenger RNAs, 2012. International Journal of Bioinformatics Research, 4: 273-275.

[5] Senthil, R., Sathish, S., Vennila, J.J. and Rajasekaran, E. Prediction of Ligand Binding Site in Globular Proteins, 2011. Journal of Advance Bioinformatics Application and Research, 2: 98-99.

[6] Akila, K., Kaliaperumal, R. and Rajasekaran, E.Carbon Distribution to Toxic Effect in Toxin Proteins, 2012. Bioinformation Discussion at the Interface of Physical and Biological Science, 8: 720-721.

[7] Rajasekaran, E. and Vijayasarathy, M. CARBANA: Carbon Analysis Program for Protein Sequences, 2011.Bioinformtaion, 5: 455-457.

[8] Rajasekaran, E. CARd: Carbon Distribution Analysis Program for Protein Sequences, 2012. Bio-information, 8: 508-512.

[9] Chase, M.W. and Fay, M.F. Barcoding of Plants and Fungi, 2009. Science, 325: 682-683.

[10] Nsimama, P.D., Mamboya, A.F., Amri, E. and Rajasekaran, E. Corelation Between The Mutated Colour Tunings and Carbon Distributions in Luciferase Bioluminescence, 2012. Journal of Computational Intelligence In Bioinformatics, 5: 105-112.

[11] Rajasekaran, E., Rajadurai, M., Vinobha, C.S. and Senthil, R. Are The Proteins Being Hydrated During Evolution? 2008. Journal of Computational Intelligence In Bioinformatics, 1: 115-118.

[12] Akila, K., Sneha, N. and Rajasekaran, E. Study of Carbon Distribution at Protein Regions of Disorder, 2012. Journal of Bioscience, Biochemistry and Bioinformatics, 2: 58-60.

[13] Johri, P., Trivedi, M. and Siddiqui, M.H, and Gokhale, M. A Study On The Presence and Distribution of Carbon Percentage in and Around the Sites of Glycosylation for Eukaryotic Proteins, 2016. Journal of Chemical and Pharmaceutical Research, 8: 52-63.

[14] Johri, P., Trivedi, M., Singh, A. and Siddiqui, M.H. Towards The Atomic Level Protein Sequence Analysis, 2016. Journal of Chemical and Pharmaceutical Research, 8: 204-207.