Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2015, 7(6):290-294

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Insights into the physiochemical characterization and classification of antiviral and viral proteins

Sajitha Lulu S.[a], Thabitha A.[a], Madhusmita Rout[a], Mohana Priya A.[a] and Vino S.[*b]

[a]*Bioinformatics Division, School of Bio Sciences and Technology, VIT University, Vellore, Tamilnadu, India*
[b]*Medical Biotechnology Division, School of Bio Sciences and Technology, VIT University, Vellore, Tamilnadu, India*

ABSTRACT

Physiochemical properties of amino acid features have frequently been used in reliable statistical method for classification of antiviral proteins. Antiviral proteins were protecting host cells from the entry of highly pathogenic viruses or deadly viruses. Viral recognization and host immunity against viral entry is still emerging research field in recent years. So we focused more on antiviral protein and its specificity. The protein structural features like amino acid composition ratio, hydrophobicity, hydrophilicity, positive charge, and negative charge calculated for viral and antiviral protein PROFEAT online tool. On the basis of physical properties, analysis was done with Pearson 2-tailed correlation using SPSS 22.0 software. Our proposed method will provide insight into important features and characteristics to identify novel antiviral proteins which inhibits viral replication during viral infection.

Keywords: Viral reorganization, Hydrophobicity, Hydrophilicity, Pearson 2-tailed correlation, Viral replication.

INTRODUCTION

Antiviral proteins are shown to be active against several life threatening viruses such as, HIV, influenza virus, dengue virus, West Nile virus, herpes simplex virus, cytomegalovirus, polio virus. Almost half a century our efforts focus on antiviral research [1-3]. To overcome viral effects in host so many drugs were available like acyclovir for herpes simplex virus (HSV) infection, and azidothymidine (AZT), stavudine and efavirenz for human immunodeficiency virus (HIV) infection [4-9]. This can be a cost effective and time-consuming process and side effects due to severe toxicity [10]. Certain antiviral peptides or proteins having lower toxicity, like enfuvirtide used for the treatment of HIV viral infection [11]. Collection of Antiviral peptides was compiled into databases such as, AVPred and HIPdb which are highly effective against specific viruses with experimental validation [12, 13].

Recent studies showed that several antiviral therapeutics agents are known to prevent the binding of viruses to the host cells either by interrupting the signalling process [14, 15] or inhibiting the replication of viruses [16–18] which involves DNA polymerase, reverse transcriptase and integrase [19]. Very few researches have been done in predicting and examining the antiviral proteins. In addition, they explored two different approaches such as amino acid composition, and physicochemical features to predict effective antiviral proteins [17]. Their results confirmed that, based on the physiochemical properties, support vector machine (SVM) approach and random forests (RF) were a prominent method to identify antiviral proteins. Physicochemical properties of amino acids are useful resources to identify antiviral proteins.

To understand antiviral protein specificity and activity is essential for their role of primary structure. In recent years to predicting the structural feature of proteins several computational approaches have been developed. Moreover, the physicochemical properties of amino acids in protein have been used to classify the viral and antiviral protein effectively.

EXPERIMENTAL SECTION

Dataset construction:

Sequences of Antiviral proteins were collected from Uniprot database using keywords like 'antiviral protein', 'viral protein'. The obtained data validated with literature and dataset created for both viral and antiviral proteins. Datasets organized into several columns like Uniprot ID, protein name, PDB accession numbers, resolution, sequence, sequence length, source organism, and reference (Table 1 and Table 2).

Table 1. Anti-Viral Proteins and its features

S.No	Uniprot ID	Protein name	Sequence length	Source organism	Reference
1.	Q7Z2W4	Zinc finger CCCH-type antiviral protein 1	902	*Homo sapiens*	Wang X., 2012
2.	Q02793	Antiviral protein SKI8	397	*Saccharomyces cerevisiae*	Arora C., 2004
3.	Q7Z434	Mitochondrial antiviral-signaling protein	540	*Homo sapiens*	Seth R.B., 2005
4.	P35207	Antiviral helicase SKI2	1,287	*Saccharomyces cerevisiae*	Widner W.R., 1993
5.	P10297	Antiviral protein I	313	*Phytolacca americana*	Monzingo A.F., 1993
6.	P21326	Antiviral protein MAP	278	*Mirabilis jalapa*	Habuka N., 1989
7.	Q8WXG1	Radical S-adenosyl methionine domain-containing protein 2	361	*Homo sapiens*	Rivieccio M.A., 2006
8.	P09922	Interferon-induced GTP-binding protein Mx1	631	*Mus musculus*	Tumpey T.M., 2007
9.	Q9N0Y2	Interferon-induced GTP-binding protein Mx2	711	*Canis familiaris*	Nakamura T., 2005
10.	Q8IUX4	DNA dC->dU-editing enzyme APOBEC-3F	373	*Homo sapiens*	Wiegand H.L., 2004
11.	Q9UII4	E3 ISG15--protein ligase HERC5	1024	*Homo sapiens*	Tang Y., 2010
12.	O95786	Probable ATP-dependent RNA helicase DDX58	925	*Homo sapiens*	Miyashita M., 2011
13.	O14879	Interferon-induced protein with tetratricopeptide repeats 3	490	*Homo sapiens*	Schmeisser H.,2010
14.	Q8IYM9	E3 ubiquitin-protein ligase TRIM22	498	*Homo sapiens*	Eldin P., 2009
15.	Q61190	Interleukin-10 receptor subunit beta	349	*Mus musculus*	Trost M., 2009
16.	Q15646	2'-5'-oligoadenylate synthase-like protein	514	*Homo sapiens*	Marques J., 2008
17.	Q8CBB9	Radical S-adenosyl methionine domain-containing protein 2	362	*Mus musculus*	Zhang Y., 2007
18.	Q9R0T8	Inhibitor of nuclear factor kappa-B kinase subunit epsilon	717	*Mus musculus*	Hemmi H., 2004
19.	Q8R5F7	Interferon-induced helicase C domain-containing protein 1	1025	*Mus musculus*	Kato H., 2006
20.	Q93VT9	60S ribosomal protein L10-1	220	*Arabidopsis thaliana*	Carvalho C.M., 2008

Table 2. Viral protein and its features

S.No	Uniprot ID	Protein name	Sequence length	Source organism	Reference
1.	P69718	Protein Rev	116	*Homo sapiens*	Crowl R., 1985
2.	Q98325	Viral CASP8 and FADD-like apoptosis regulator	241	*Molluscum contagiosum virus subtype 1*	Thome M., 1997
3.	P03303	Genome polyprotein	2179	*Human rhinovirus 14*	Stanway G., 1984
4.	P69726	Protein Vpr	96	*Human immunodeficiency virus*	Re F., 1995
5.	P00519	Tyrosine-protein kinase ABL1	1130	*Homo sapiens*	Yuan Z M., 1997
6.	P04615	Protein Rev	100	*Human immunodeficiency virus*	Guyader M., 1987
7.	P08392	Major viral transcription factor ICP4	1298	*Human herpesvirus 1*	Papavassiliou A.G., 1991
8.	P07242	Late transcription elongation factor H5	203	*Vaccinia virus*	Costa D., 2010
9.	P17765	Genome polyprotein	3056	*Bean yellow mosaic virus*	Hammond J., 2003
10.	P34015	Cytokine response-modifying protein B	349	*Variola virus*	Alejo A., 2006
11.	P03301	Genome polyprotein	2209	*Poliovirus type 1*	Nomoto A., 1982
12.	F5HIC6	Viral IRF3-like protein	566	*Human herpesvirus 8 type P*	Glenn M., 1999
13.	Q91QZ3	RNA replication polyprotein	1962	*Citrus leaf blotch virus*	Vives M C., 2001
14.	Q9YJW3	Genome polyprotein 1	2410	*Barley yellow mosaic virus*	Chen J., 1999
15.	F5HEZ4	Viral FLICE protein	188	*Human herpesvirus 8 type P*	Glenn M., 1999
16.	P14349	Gag polyprotein	648	*Human spumaretrovirus*	Maurer B.,1988
17.	P24937	Pre-protein VI	250	*Human adenovirus C serotype 5*	Chrocobzek J., 1992
18.	P03132	Protein Rep68	536	*Adeno-associated virus 2*	Srivastava A., 1983
19.	P03385	Envelope glycoprotein	665	*Moloney murine leukemia virus*	Wu S.R., 2008
20.	P18541	RING finger protein Z	90	*Lymphocytic choriomeningitis virus*	Cornu T.I., 2002

Physiochemical property prediction:

The viral and antiviral proteins were classified by obtaining hydrophilic and hydrophobic amino acid ratio, negative and positive charge amino acid ratio, number of helix, sheet and coil present on each individual protein. These all characteristics were obtained by using PROFEAT online tool. Simple python scripts can also be used to calculate the composition of amino acids, hydrophobic, hydrophilic, charge residues and the length of protein sequence. Graph Pad prism 6 has been used for graph generation.

Statistical method:

Pearson's Correlation coefficient helps to analyze the relation between two variables. In that the degree of the variables are linearly related in a sample. Pearson correlation coefficient ranges in value from -1.00 to +1.00. In Pearson's Correlation (2-tailed) significant values is less than 0.05.

RESULTS AND DISCUSSION

Viruses are the most dominant infectious agent in humans. Major outbreaks caused by various new viruses have been reported from different parts of the world in recent years such as SARS, Ebola, Swine flu, Dengue, Cholera and Dysentery to name a few. These diseases cause mostly mortality or DALY's (Disability adjusted life years) and

constitute for second highest total death worldwide. Hence, our aim is to classify the viral and antiviral proteins using physiochemical properties of amino acid.

Feature importance of amino acid composition:
In this study five physicochemical properties for amino acid were examined. Each property contains 20 numerical values to represent 20 amino acids. The feature of a protein sequence is described mainly by the number of positively and negatively charged amino acids and composition of amino acids present. Amino acid composition is the ratio of each amino acid in a peptide. The ratio of an amino acid with type of protein (X) was calculated

$$AAC(X) \sim N/L$$

Where N is the number of the amino acid with type and L is the length of protein X.

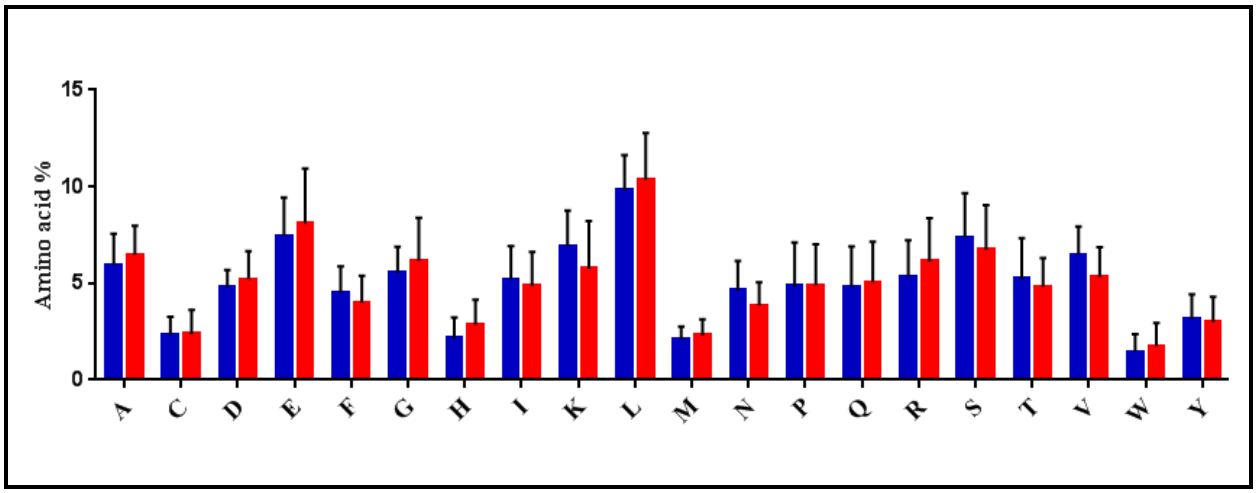

Figure 1. Distribution of amino acid composition of antiviral and viral proteins
The blue bars represent 20 antiviral protein composition and red bars represent 20 viral protein composition

The statistical distribution of amino acid composition ratio of antiviral protein is present more compare to viral protein except asparagine, serine, threonine, valine, tyrosine, and lysine (figure1).

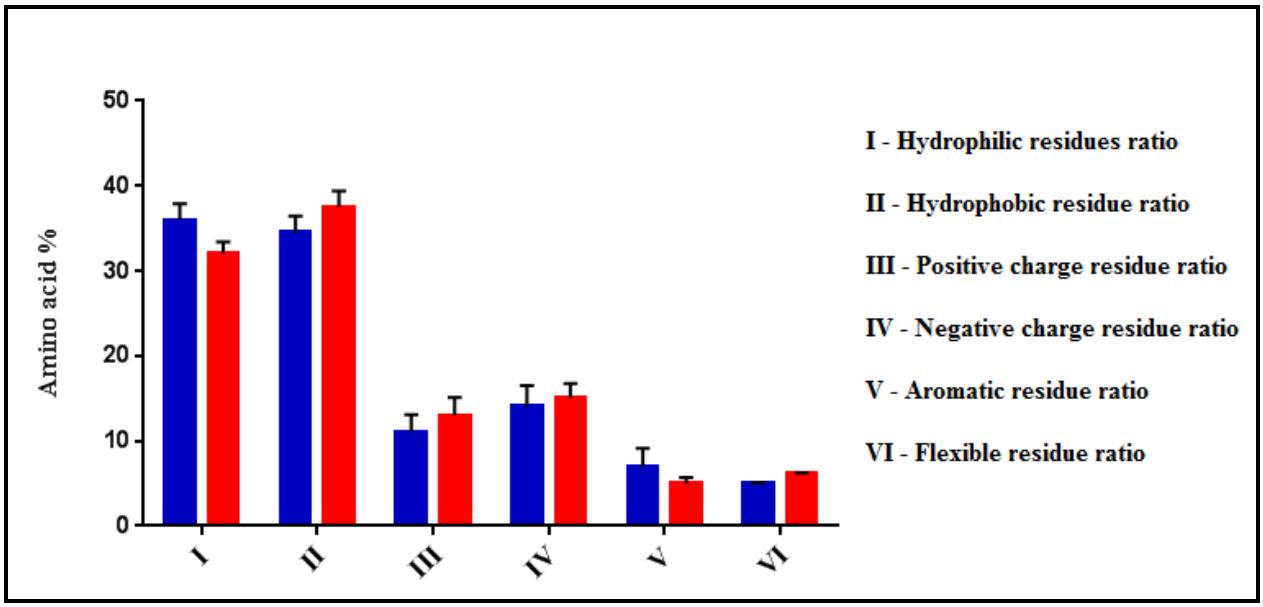

Figure 2. Distribution of physiochemical properties of antiviral and viral proteins
The blue bars represent 20 antiviral protein composition and red bars represent 20 viral protein composition

In most of the protein molecules, the hydrophobic amino acids (Ala, Ile, Phe, Trp, Tyr) are accumulated in the core, where as the hydrophilic amino acids (Asn, Gln, Gly, Met, Pro, Ser, Thr, Val), positive charge amino acids (Lys,

Arg) and negative charge amino acids (Asp, Glu) are present in the surface of the protein. The hydrophilic amino acids can interact with solvents through hydrogen bonds either by donating or accepting protons. The salt bridges (side chain-side chain, side chain- main chain) are also formed by the charged amino acids in protein

Table 3. The correlation results of physiochemical properties

		Hydrophilic	Positive charge
Hydrophilic	Pearson correlation	1	0.97
	Sigma (2=tailed)		0.002
	N	20	20
Positive charge	Pearson correlation	0.97	1
	Sigma (2=tailed)	0.002	
	N	20	20

		Hydrophilic	Negative charge
Hydrophilic	Pearson correlation	1	0.915
	Sigma (2=tailed)		0.002
	N	20	20
Negative charge	Pearson correlation	0.915	1
	Sigma (2=tailed)	0.002	
	N	20	20

		Hydrophilic	AA composition (%)
Hydrophilic	Pearson correlation	1	0.963
	Sigma (2=tailed)		0.002
	N	20	20
AA composition (%)	Pearson correlation	0.963	1
	Sigma (2=tailed)	0.002	
	N	20	20

		Hydrophobic	Positive charge
Hydrophobic	Pearson correlation	1	0.946
	Sigma (2=tailed)		0.002
	N	20	20
Positive charge	Pearson correlation	0.946	1
	Sigma (2=tailed)	0.002	
	N	20	20

		Hydrophobic	Negative charge
Hydrophobic	Pearson correlation	1	0.878
	Sigma (2=tailed)		0.002
	N	20	20
Negative charge	Pearson correlation	0.878	1
	Sigma (2=tailed)	0.002	
	N	20	20

		Hydrophobic	AA composition (%)
Hydrophobic	Pearson correlation	1	0.97
	Sigma (2=tailed)		0.002
	N	20	20
AA composition (%)	Pearson correlation	0.97	1
	Sigma (2=tailed)	0.002	
	N	20	20

		Positive charge	Negative charge
Positive charge	Pearson correlation	1	0.925
	Sigma (2=tailed)		0.002
	N	20	20
Negative charge	Pearson correlation	0.925	1
	Sigma (2=tailed)	0.002	
	N	20	20

		Positive charge	AA composition (%)
Positive charge	Pearson correlation	1	0.923
	Sigma (2=tailed)		0.002
	N	20	20
AA composition (%)	Pearson correlation	0.923	1
	Sigma (2=tailed)	0.002	
	N	20	20

		Negative charge	AA composition (%)
Negative charge	Pearson correlation	1	0.832
	Sigma (2=tailed)		0.002
	N	20	20
AA composition (%)	Pearson correlation	0.832	1
	Sigma (2=tailed)	0.002	
	N	20	20

AA – Aminoacid

In our study, the twenty amino acids divided into six groups: hydrophilic, hydrophobic, positive, negative, aromatic and flexible residues. The hydrophilic and aromatic (Phe, Trp, Tyr) amino acid ratios of antiviral proteins are more

compare to viral proteins. In viral proteins, the hydrophobic, positive and negative charge and flexible (Gly) amino acid ratios are greater than antiviral proteins.

Statistical analysis:
The statistical analysis of the physicochemical properties of the antiviral protein was done. The following properties were examined, including amino acid composition, charge, hydrophobicity, hydrophilicity. The Pearson's r for the correlation between the two variables close to 1 and positive value that, hydrophilic and positive charge is 0.970, hydrophilic and negative charge is 0.915, hydrophilic and amino acid composition ratio is 0.963, hydrophobic and positive charge is 0.946, hydrophobic and negative charge is 0.878, hydrophobic and amino acid composition ratio is 0.970, positive charge and negative charge is 925, positive charge and amino acid composition ratio is 923, negative charge and amino acid composition ratio is 832 which shows strong relationship between two variables and positive correlation i.e., one variable increases in value and the second variable also increase in value (Table 3).

CONCLUSION

Antiviral proteins have inhibitory effect against viral infections. These proteins exhibit specific changes for different types of viral influence in various organisms. According to our current work, the antiviral proteins are found to be more hydrophilic or water soluble in nature. Viral proteins are hydrophobic in nature and having more charged residues. These hydrophilic and hydrophobic natures have a powerful impact on three dimensional structure of the protein. Our analysis demonstrated that, finding the characteristics of viral and antiviral protein providing gateway for antiviral drug discovery.

Acknowledgement
We are grateful to the management of VIT University for supporting this investigation.

REFERENCES

[1] MA Wainberg. *Antiviral Res.*, **2009**, 81(1), (1–5).
[2] E De Clercq. *Annu Rev Pharmacol Toxicol*., **2011**, 51, (1–24).
[3] HJ Field; E De Clercq. *Microbiol Today.*, **2004**, 31, 58–61.
[4] HJ Schaeffer; L Beauchamp; P de Miranda; GB Elion; DJ Bauer; P. Collins. *Nature*., **1978**, 272(5654), 583–585.
[5] H Mitsuya; KJ Weinhold; PA Furman; MH St Clair; SN Lehrman. *Proc Natl Acad Sci U S A.*, **1985**, 82(20), 7096– 7100.
[6] M Baba; R Pauwels; P Herdewijn; E De Clercq; J Desmyter. *Biochem Biophys Res Commun*., **1987**, 142(1), 128–134.
[7] Y Hamamoto; H Nakashima; T Matsui; A Matsuda; T Ueda. *Anti microb Agents Chemother.*, **1987**, 31(6), 907–910.
[8] TS Lin, RF Schinazi, WH Prusoff. *Biochem Pharmacol*., **1987**, 36(17), 2713–2718.
[9] SD Young; SF Britcher; LO Tran; LS Payne; WC Lumma. *Antimicrob Agents Chemother*., **1995**, 39(12), 2602–2605.
[10] W Lewis; MC Dalakas. *Nat Med*., **1995**, 1, 417–422.
[11] C Wild; T Greenwell; T Matthews. *AIDS Res Hum Retroviruses*., **1993**, 9(11), 1051–1053.
[12] N Thakur; A Qureshi; M Kumar. *Nucleic Acids Res.*, **2012**, 40, W199–204.
[13] A Qureshi; N Thakur; M Kumar. *PLoS One.*, **2013**, 8(1), e54908.
[14] DM Rausch; KM Hwang; M Padgett; AH Voltz; A Rivas. *Ann N Y Acad Sci*., **1990**, 616, 125–148.
[15] W Pang; SC Tam; YT Zheng. *Antivir Chem Chemother*., **2009**, 20, 1–18.
[16] JM Louis; F Dyda; NT Nashed; AR Kimmel; DR Davies. *Biochemistry.*, **1998,** 37, 2105–2110.
[17] I Oz Gleenberg; O Avidan; Y Goldgur; A Herschhorn; A Hizi. *J Biol Chem.*, **2005**, 280, 21987–21996.
[18] IO Gleenberg; A Herschhorn; A Hizi. *J Mol Biol*., **2007**, 369, 1230–1243.
[19] G Castel; M Chteoui; B Heyd; N Tordo. *Molecules*., **2011,** 16(5), 3499–3518.