Analyses and Prediction of Antiviral proteins using Support Vector Machine

Akanksha Rajput and Manoj Kumar*

Bioinformatics Centre, CSIR-Institute of Microbial Technology, Sector 39A, Chandigarh-160036, INDIA
Email: akanksha@imtech.res.in ; manojk@imtech.res.in

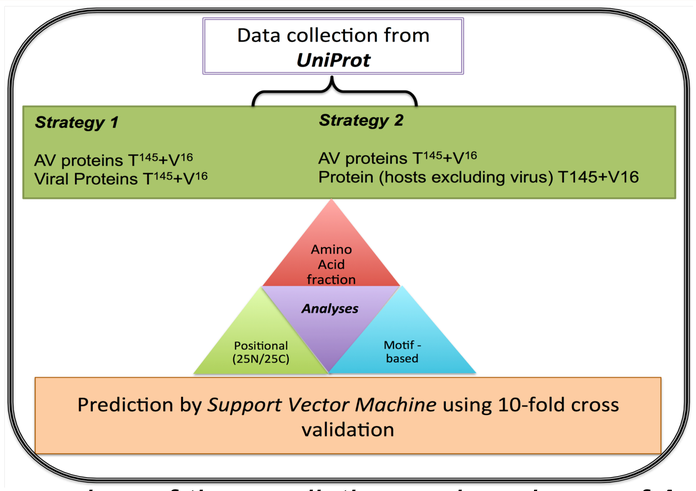
Introduction

- Antiviral (AV) proteins have therapeutic potential to interfere viral replication in hosts.
- These proteins are induced in host as an immune response against viral infection *via* interferon that further mediates them to inhibit virus replication.
- Despite having therapeutic potentials, these proteins are computationally unexplored.

Objectives

- Prediction of Anti-viral (AV) proteins using Support Vector Machine (SVM).
- Analyses of AV proteins

Methodology



Brief overview of the prediction and analyses of Anti-viral protein

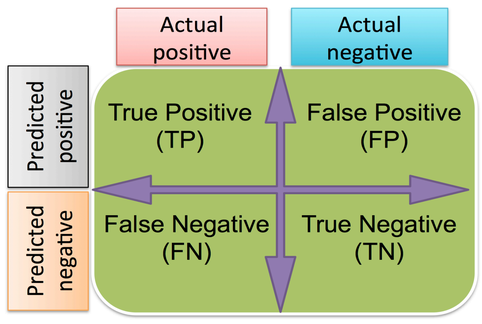
Feature selection algorithm

- Minimum redundancy maximum relevance (mRMR) feature selection method is employed to find most contributing features.

Classifier used

- Support Vector Machine are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

Performance measures



$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Results

I. *Performance of SVM models during 10-fold cross validation on training/ testing and validation data set*

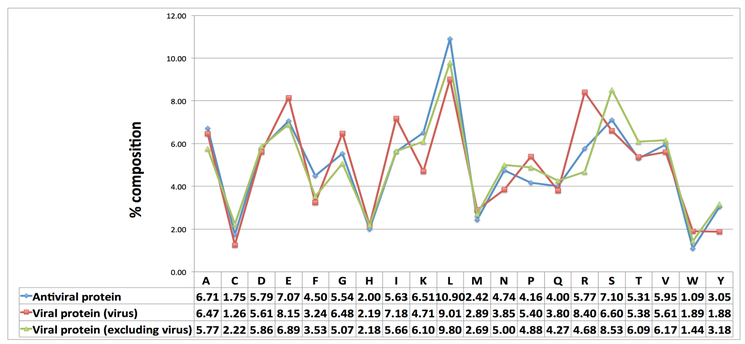
	Training/testing data set ($T^{145p+145n}$)					Validation data set ($V^{16p+16n}$)				
	Sen	Spec	Acc	MCC	AUC	Sen	Spec	Acc	MCC	AUC
AAC										
Strategy 1	99.31	97.93	98.62	0.97	1.00	100.00	93.75	96.88	0.94	1.00
Strategy 2	97.93	88.28	93.10	0.87	0.98	87.50	81.25	84.38	0.69	0.90
DPC										
Strategy 1	97.93	96.55	97.24	0.94	0.99	100.00	93.75	96.88	0.94	0.99
Strategy 2	97.24	90.34	93.79	0.88	0.98	87.50	87.50	87.50	0.75	0.93
PHY										
Strategy 1	97.93	97.93	97.93	0.96	0.99	100.00	93.75	96.88	0.94	1.00
Strategy 2	95.17	92.41	93.79	0.88	0.97	87.50	93.75	90.62	0.81	0.92
AAC+DPC										
Strategy 1	97.93	97.93	97.93	0.96	0.98	100.00	93.75	96.88	0.94	0.99
Strategy 2	95.17	90.34	92.76	0.86	0.98	87.50	87.50	87.50	0.75	0.91
AAC+DPC+PHY										
Strategy 1	97.93	97.24	97.59	0.95	0.98	100.00	93.75	96.88	0.94	1.00
Strategy 2	99.31	83.45	91.38	0.84	0.98	87.50	81.25	84.38	0.69	0.89

Sen, Sensitivity; Spec, Specificity; Acc, Accuracy; AUC, Area Under the Curve; AAC, Amino Acid Composition, DPC, Dipeptide Composition; PHY, Top 10 - physicochemical properties;

****Strategy 1 , 145 AV protein and 145 viral proteins;***
****Strategy 2 , 145 AV protein and 145 other host proteins (excluding viral)***

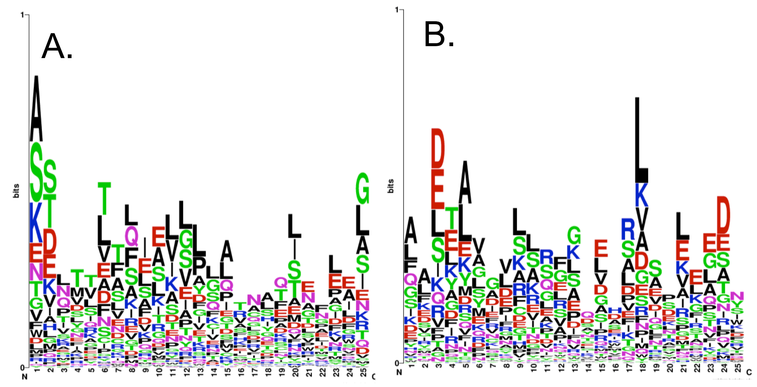
Results

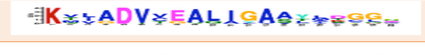
II. *Comparison of Amino Acid Composition of AVPs, Viral proteins and host proteins other than viruses*



	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Antiviral protein	6.71	1.75	5.79	7.07	4.50	5.54	2.00	5.63	6.51	10.90	2.42	4.74	4.16	4.00	5.77	7.10	5.31	5.95	1.09	3.05
Viral protein (virus)	6.47	1.26	5.61	8.15	3.24	6.48	2.19	7.18	4.71	9.01	2.89	3.85	5.40	3.80	8.40	6.60	5.38	5.61	1.89	1.88
Viral protein (excluding virus)	5.77	2.22	5.86	6.89	3.53	5.07	2.18	5.66	6.10	9.80	2.69	5.00	4.88	4.27	4.68	8.53	6.09	6.17	1.44	3.18

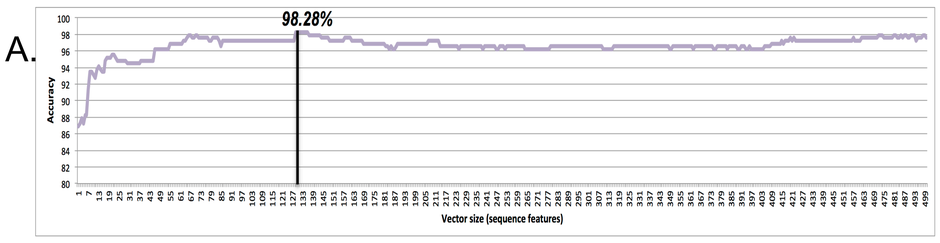
III. *Positional preference of Amino Acid* ***A)*** *25N and* ***B)*** *25C terminus of AVPs*



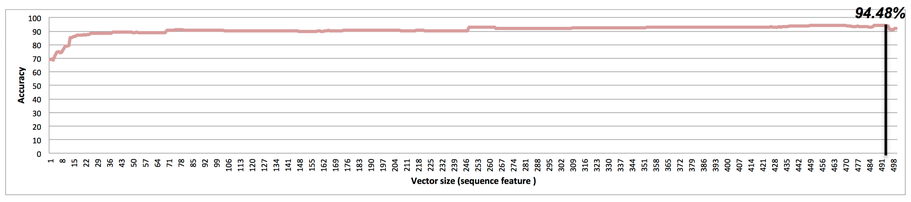
S.No	Motif logo	Motif Width	Sequence coverage
Motif 1		20	32
Motif 2		20	29
Motif 3		20	26
Motif 4		20	28
Motif 5		20	21
Motif 6		15	20
Motif 7		20	25
Motif 8		20	21
Motif 9		20	21
Motif 10		20	21

IV. *Motif extraction of AVPs using MEME/MAST suite*

A.



B.



V. *Performance of SVM hybrid (AAC+ DPC+PHY) model during mRMR feature selection technique on two data sets* ***A)*** *AVPs & Viral proteins* ***B)*** *AVPs & host proteins other than virus*

Conclusion

- Our study would help the researchers in exploring antiviral potential of protein(s) that further helps in determining their therapeutic potential against various viruses.

Acknowledgements

- Council of Scientific and Industrial Research (CSIR), India
- Department of Biotechnology (DBT), Govt. of India

References

- Samuel CE. Antiviral actions of interferons. ***Clin Microbiol Rev***. 2001 Oct;14(4):778-809.
- Cortes, C. & Vapnik, V. Support-vector networks. ***Mach Learn***. 1995 Feb; 20(273).
- Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy***. IEEE Trans Pattern Anal Mach Intell.*** 2005 Aug;27(8): 1226-38.
- Rajput A, Gupta AK, Kumar M. Prediction and analysis of quorum sensing peptides based on sequence features. ***PLoS One.*** 2015 Mar17;10(3):e0120066.
- Thakur N, Qureshi A, Kumar M. AVPpred: collection and prediction of highly effective antiviral peptides. ***Nucleic Acids Res***. 2012 Jul;40 (Web Server issue):W199-204.