OPEN ACCESS

Available online at www.ijasbt.org

International Journal of Applied Sciences and Biotechnology

A Rapid Publishing Journal

APPLIED SCIENCES		BIOTECHNOLOGY
Biochemistry	Immunobiology	Microbial biotechnology
Molecular biology	Bioinformatics	Medical biotechnology
Microbiology	Novel drug delivery system	Industrial biotechnology
Cell biology	Pharmacology	Environmental biotechnology
Cytology	Neurobiology	Nanotechnology
Genetics	Bio-physics	
Pathology	Botany	
Medicinal chemistry	Zoology	
Polymer sciences	Allied science	
Analytical chemistry	Earth science	
Natural chemistry		

If any queries or feedback, then don't hesitate to mail us at:
editor.ijasbt@gmail.com

SEM-Biotech Publishing

IC Value: **4.37**

D.S. Mundaganore (2013) Int J Appl Sci Biotechnol, Vol 1(4): 272-278
DOI: 10.3126/.ijasbt.v1i4. 9184

Research Article

International Journal of Applied Sciences and Biotechnology

ISSN: 2091-2609

IN SILICO VALIDATION OF MIDDLE EAST RESPIRATORY SYNDROME (MERS) VIRUS PROTEINS FOR BETTER DRUG DEVELOPMENT

D.S. Mundaganore[1], Y.D. Mundagnore[2]and K. V. Ashokan[3]

[1]Department of Zoology, Willingdon College, Sangli, Maharashtra, India-416416
[2]Department of Zoology, Miraj Mahavidyalya, Miraj, Sangli, Maharashtra, India-416410
[3]Department of zoology, P.V.P College, Kavathe Mahankal, Sangli, Maharashtra, India-416405

Corresponding author email: akvsangli@gmail.com

Abstract

MERS Co-virus protein sequence of N, M, E and S-protein was validated by bioinformatics servers and tools. The study revealed that S-protein shows highest percentage of amino acid and the least in M and E-protein. The bit score also shows the same trend but the E-value is maximum in E and M protein. The amino acid composition revealed that N-protein is rich in glycine and M-protein rich in leucin. Pyrrolysine, Selenocysteine and cystein (N-Protein) are absent in all the protein studied an indication of low thermo stability. The physico-chemical study showed that M, N, and E proteins are positively charged due to specific amino acids (Arginin and lysine) and S-protein is negatively charged due to aspartic acid and glutamic acid. EC is maximum in S-protein. Instability Index (II) shows that E and S-proteins are more stable in test tube than other proteins. Further all proteins are hydrophobic with GRAVY below '0'. To get clearer picture of the physical and chemical attribute of the protein we generated 3D model and the model was validated and observed that the 3D structure falls in the accepted limits. The practical implication of the study is that the result will assists the pharmacologist and other drug developers and Government bodies to get better knowledge on these proteins and develop possible new vaccine against MERS Co-virus.

Key words:MERS Co-virus**,** M, N, E, N and S-protein**.** Physico chemical character, SWISS MODEL, Model validation.

Introduction

Middle East Respiratory Syndrome (MERS) (De Groot *et al*, 2013) is an illness caused by virus. It was first reported in Saudi Arabia on 24 September on ProMED-mail 2012. The virus is specifically known as corona virus i.e. MERS-CoV. The closest phylogenetic neighbors to MERS-CoV are putative beta coronaviruses in China (Zaki *et al*, 2012), in Netherlands (van Boheemen *et al*, 2012, and a recently discovered isolate from South Africa (Ithete *et al*, 2012). Confirmed cases of MERS-CoV infection developed sever acute respiratory illness. They have fever, cough, and shortness of breath and half of the reported cases are dead. All the infected cases are linked to four countries in or near the Arabian Peninsula. It spread through close contact. But, the virus has not shown to spread in a sustained way in communities. The risk factor of this virus, spreading methods and how to prevent the infection is not clear yet. CDC (Center for Disease Control and Prevention) is working to understand these factors better. The more censes related to infection and death respectively is France (2, 1), Italy (3, 0), Jordan (2, 2), Qatar (5, 2), Tunisia (3, 1), United Kingdom (UK) (3, 2), United Arab Emeritus (UAE) (6, 2) with in all (114, 54) (http://www.cdc.gov/coronavirus/mers/). Molecular analysis including gentical and phylogenetic studies revealed that MERS-CoV falls under C-lineage of batracovirus genus and more related to Tylonycteris beta corona virus HKU4 and HKU5 Pipistrellus beta corona virus (Zaki *et al*, 2012, van Boheemen *et al*, 2012, Woo *et al*, 2012). Study indicate that spike (S) protein interact with and helps to enter into the host (Raj *et al*, 2013, Jiang *et al*, 2013). The available literature indicates the molecular organization of the virus is in juvenile stage. A lot of progress is needed to reach the target specific drugs and vaccine. The various proteins like M, N, E and S is not fully studied both concerned with its structure and function. To understand the structure of protein and its molecular organization is one of the requirement for developing effective vaccine and possible drugs against the dreaded disease MERS. Her we make an attempt to characterize different protein on MERS-Coronavirus with various bioinformatics tools and servers.

Materials and methods

Extraction of protein sequences

The four protein sequences of MERS Coronavirus were extracted from NCBI (National Center for Biotechnology). It is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health. It has

a series of database collection relevant to biotechnology and bioinformatics. The sequences retrieved from NCBI and its various specifications are given in Table 1.

Identification of amino acid percentage composition

The percentage composition of various amino acid and various physico-chemical properties of the retrieved sequence was done with the help of ProtParam bioinformatics server. The parameters computed by ProtParam include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). The amino acid and atomic compositions are self-explanatory. All the other parameters will be explained below.

Extinction coefficients

The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. It is useful to have an estimation of this coefficient for analyzing a protein with a spectrophotometer when purifying it. It has been shown (Gill and Hippel, 1989) that it is possible to estimate the molar extinction coefficient of a protein from knowledge of its amino acid composition. From the molar extinction coefficient of tyrosine, tryptophan and cystine (cysteine does not absorb appreciably at wavelengths >260 nm, while cystine does) at a given wavelength, the extinction coefficient of the native protein in water can be computed using the following equation:

E (Prot) = Numb(Tyr)*Ext(Tyr) + Numb(Trp)*Ext(Trp) +Numb(Cystine)*Ext(Cystine)

Where (for proteins in water measured at 280 nm): Ext(Tyr) = 1490, Ext(Trp) = 5500, Ext(Cystine) = 125;

The absorbance (optical density) can be calculated using the following formula:

Absorb(Prot) = E(Prot) / Molecular weight

Instability index (II)

The instability index provides an estimate of the stability of a protein in a test tube. Statistical analysis of 12 unstable and 32 stable proteins has revealed (Guruprasad, 1990) that there are certain dipeptides, the occurence of which is significantly different in the unstable proteins compared with those in the stable ones. The authors of this method have assigned a weight value of instability to each of the 400 different dipeptides (DIWV). Using these weight values it is possible to compute an instability index (II) which is defined as:

$$II = \frac{10}{L} * \sum_{i=1}^{i=L-1} DIWV(x[i]x[i+1])$$

Where: L is the length of sequence

DIWV(x[i]x[i+1]) is the instability weight value for the dipeptide starting in position i.

A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable.

Aliphatic index

The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins. The aliphatic index of a protein is calculated according to the following formula (Ikai, 1980):

Aliphatic index = X(Ala) + a * X(Val) + b * (X(Ile) + X(Leu))

Where X(Ala), X(Val), X(Ile), and X(Leu) are mole percent (100 X mole fraction)

of alanine, valine, isoleucine, and leucine.

The coefficients a and b are the relative volume of valine side chain (a = 2.9) and of Leu/Ile side chains (b = 3.9) to the side chain of alanine.

Table 1: Retrieved protein sequence and its specificity form MERS Coronavirus.

Identifier	Protein	Score in bits	Expect	Identities	Sequence length in number of amino acids
K0BVN3	**Nucleoprotein [N]**	696	0.0	84%	413
K9N7A1	**Membrane protein (M protein)**	441	e-121	100%	219
K9N5R3	**Envelope small membrane protein**	203	1e-50	100%	219
K0BRG7	**S protein**	2745	0.0	99%	1353

Table 2: Amino acid composition and physic-chemical properties of four proteins of MERS CoV.

Amino acids	N-Protein %	M-Protein %	E-Protein %	S-Protein %
Ala (A)	8	8.7	4.9	6.5
Arg (R)	6.3	4.8	3.7	3.3
Asn (N)	7.7	4.6	3.7	5.7
Asp (D)	4.8	2.7	2.4	4.9
Cys (C)	0	0.9	4.9	3.1
Gln (Q)	5.8	2.7	4.9	5.2
Glu (E)	3.3	2.3	2.4	3.4
Gly (G)	9.2	5	3.7	6.8
His (H)	1.5	1.4	0	1.6
Ile (I)	3.1	8.2	4.9	5.4
Leu (L)	6.3	9.6	13.4	8.9
Lys (K)	7	3.2	2.4	3.8
Met (M)	1.7	5	3.7	1.6
Phe (F)	3.4	4.6	9.8	5.2
Pro (P)	8.2	4.6	7.3	4.6
Ser (S)	8.5	9.1	2.4	9.9
Thr (T)	7.3	6.4	8.5	6.8
Trp (W)	1.5	3.2	1.2	0.7
Tyr (Y)	2.4	4.6	3.7	5.6
Val (V)	4.1	9.1	12.2	7.1
Pyl (O)	0	0	0	0
Sec (U)	0	0	0	0

Table 3: Physico-chemical properties of four proteins retrieved from NCBI

Types of protein	Negative charged residue	Positive charged residue	EC	II	AI	GRAVY
	Asp+Glu	Arg+Lys				
N-Protein	33	55	47900	48.62	56.76	-0.865
M-Protein	11	16	53525	43.67	104.61	0.436
E-protein	4	5	10220	33	111.59	0.795
S-Protein	112	96	170865	36.51	82.71	-0.074

3D Model generation

The 3D Model of the protein was generated by using SWISSMODEL package. SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modeling accessible to all biochemists and molecular biologists worldwide (Arnold *et al*, 2006; Kiefer *et al*, 2009; Peitsch, 1995). The three models was created with the help of YASARA software by using the pdb template generated from SWISSMODEL.

Validation of 3D model

The 3D model was predicted by various parameters given by SWISS model server as sequence identity percentage, E-value and QMEAN Z Square.

GRAVY (Grand Average of Hydropathy)

The GRAVY value for a peptide or protein is calculated as the sum of hydropathy values (Kyte and Doolite, 1982) of all the amino acids, divided by the number of residues in the sequence.

Prediction of phosphorylation sites

Phosphorylatiuon sites were predicted by using NetPhos. The NetPhos 2.0 server produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins (Blom *et al*, 1999).

Transmembrane sequence analysis

Transmembranase domains were predicted by using SOSUI server (Hirokawa *et al*, 1998; Mitaku *et al*, 1999; Mitaku *et al*, 2002). SOSUI which distinguishes between membrane and soluble proteins from amino acid sequences, and predicts the transmembrane helices for the former.

Prediction of hydrophobic residues

The hydrophobic residues were predicted by using Pepwheel program. pepwheel draws a helical wheel diagram for a protein sequence. This displays the sequence in a helical representation as if looking down the axis of the helix. It is useful for highlighting amphipathicity and other properties of residues around a helix. By default, aliphatic residues are marked with squares, hydrophilic residues are marked with diamonds, and positively charged residues with octagons, although this can be changed (Ramchandran *et al*, 1966).

Peroxysomal targeting signal prediction

The peroxysomal targeting signals were predicted by using PTS1 predictor. PTS1 is the most abundant target signal, and is located in the C-terminal. The PTS1 is highly conserved in evolution, with the following tripeptide consensus S/A-K/R-L/M. Moreover, beyond this tripeptide motif, the PTS1 motif has been recently enlarged up to 12 C-terminal residues (Nieberger *et al*, 2003) comprising: the tripeptide, 5 residues upstream the tripeptide and 5 residues with polar properties, more upstream.

Results

The sequence retrieved from NCBI (Table 1) shows that S-protein contain highest percentage of amino acid (1353 AA), the least in M-protein and E-protein (219 AA) and N-protein have 413AA. The score in bits also shows the same trend but identity shows maximum in M and E protein (100%) followed by S protein (99%) and the least in N-Protein (84%), but E-value shows maximum in E and M protein and least in N-and S-Protein. The highest E-value prove that E and M proteins are membranous in structure and N and S- protein with least, prove that it is intracellular in location.

The amino acid composition and various physical and chemical properties (Table 2) shows that N-protein is rich in glycine amino acid (9.2%) and least cystine, pyrrolysine and Selenocysteine, in the case of M-protein the richest one is leucine (9.6%) and the least one is pyrrolysine and Selenocysteine, E-protein also shown the same trend as M-protein but have more leucine (13.4%), but S-protein showed serine as the most abundant (9.9%) one and the least one is same as M-protein and E-protein. The least or zero occurrence of pyrrolysine and Selenocysteine in MERS Covirus indicating that it is thermo liable as these amino acids are more predominant in archae *Methanosarcina barkeri* (Srinivasan *et al*, 2005; Bing Hao *et al*, 2002). The lack of cysteine in N-protein also indicates the low stability of the protein due to lack of disulphide bonds. Leucine is most abundant in M-protein. It is observed that in HIV Nef protein containing leucine motif down-regulated CD4 from the cell surface and enhanced viral replication, the same mode of action may be involved in MERS Co Virus replication; hence it is more in M-protein. The abundant amino acid in s-protein is serine; it may be due to the fact that the virus multiplication by lytic cycle is predominant as in Epstein Barr virus (Amy *et al*, 1999).

The phisico-chemical properties showed that proteins M, N and E are positively charged and S protein is negatively charged one, this correlates the abundance of corresponding amino acids (Table 3). EC is observed maximum in S-protein and least in E-protein, M and N-protein shows more or less equal EC. Instability index shows that E-protein and S- proteins are more stable in test-tube as its II is less than 40 (Table 3) but M and N-protein less stable as its II is more than 40 (Guruprasad *et al*, 1990).The aliphatic index of a protein is a measure of the relative volume occupied by aliphatic side chain of the amino acids: alanine, valine, leucine and isoleucine. An increase in the aliphatic index increases the thermostability of globular proteins. The present study shows that M and E protein have more AI and hence more heat stable and N and S protein have low AI and hence less stable (Table 3). Grand average hydrophobicity shows that all proteins are hydrophobic with GRAVY below '0' (Table 3). To make a clear picture about the protein of MERS Co-virus we genarted 3D model by using SWISSMODEL (Fig. 1&2), it revealed that the N-protein has mass of 17168.094g/mol and S-protein has mass of 49937.865g/mol (Table4). The E-value and sequence identity is in agreement with good 3D model (Table 4). Comparison with non-redundant set of PDB structure is also in agreement with good model (Fig.3&4). The transmembarne sequence predicted by SOSUI (Table5 & Fig9n10) shows that hydrophobic residues are predominant than polar and charged residues.

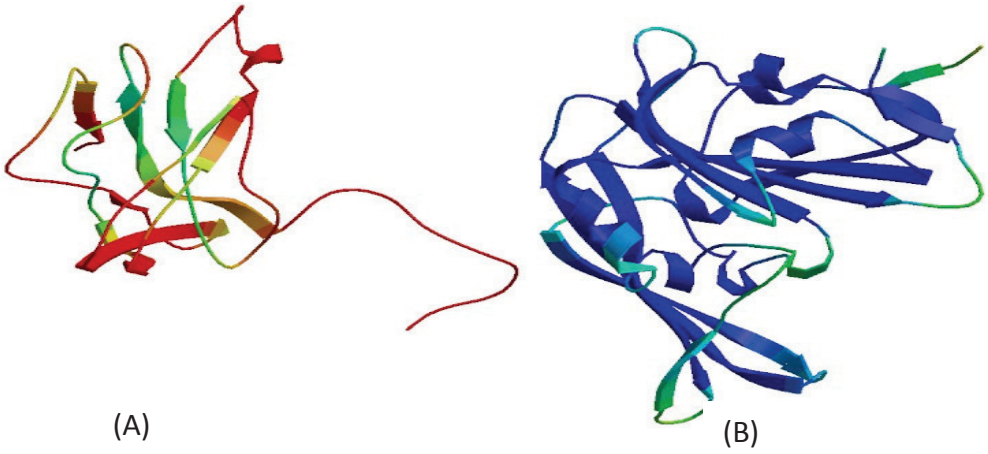

Fig 1: Predicted 3D Model (A): N-Protein (1ssk); (B): S-Protein (413nB)

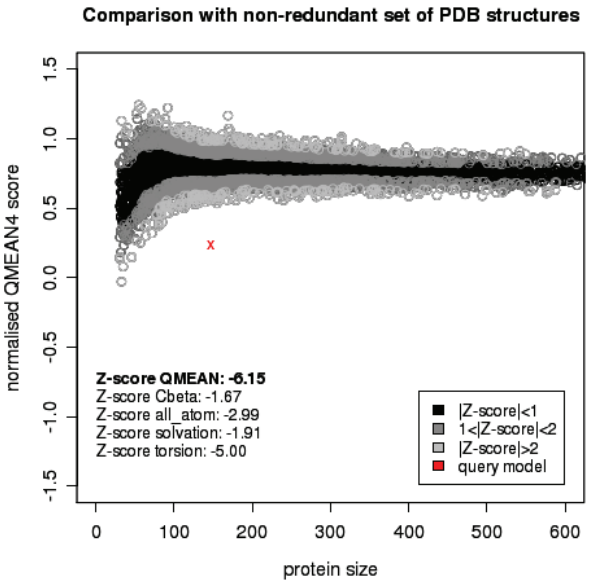

Fig. 3: Estimated absolute Model quality of 1ssk template

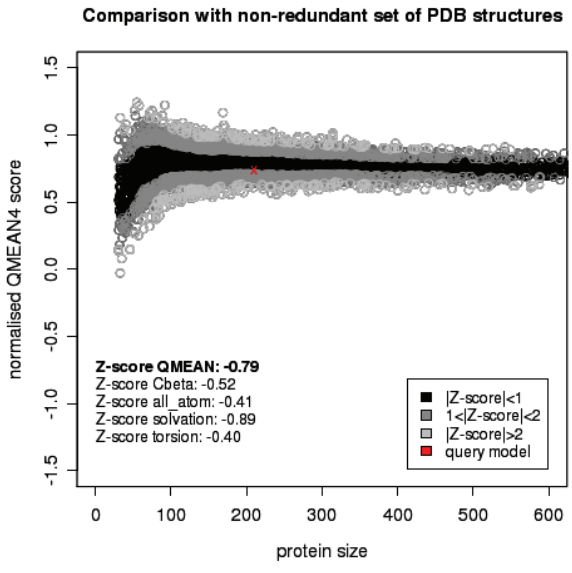

Fig. 4: Estimated absolute Model quality of4l3B template

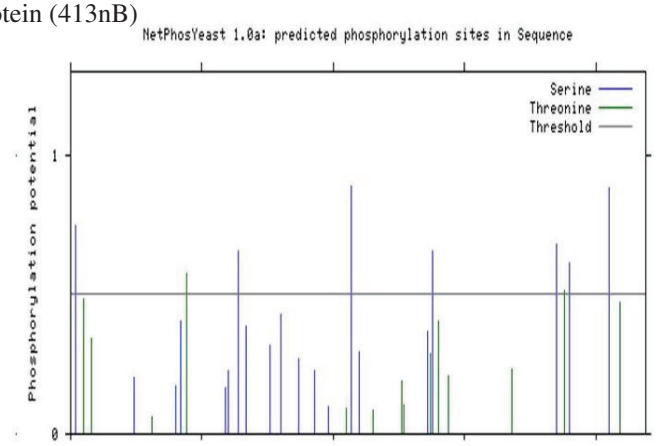

Fig 5: Predicted Phosphorylation site in N-Protein of MER

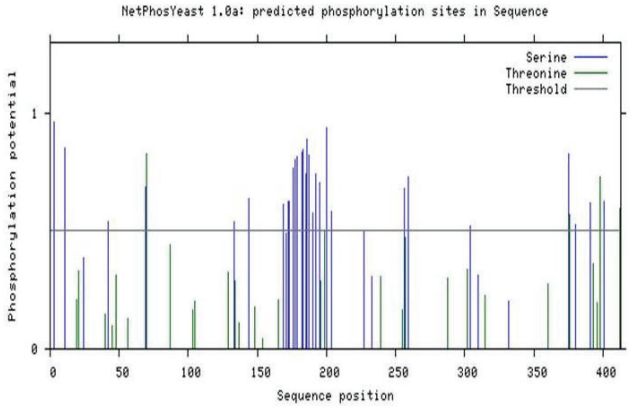

Fig. 6: Predicted Phosphorylation sites of M-Protein of MERS

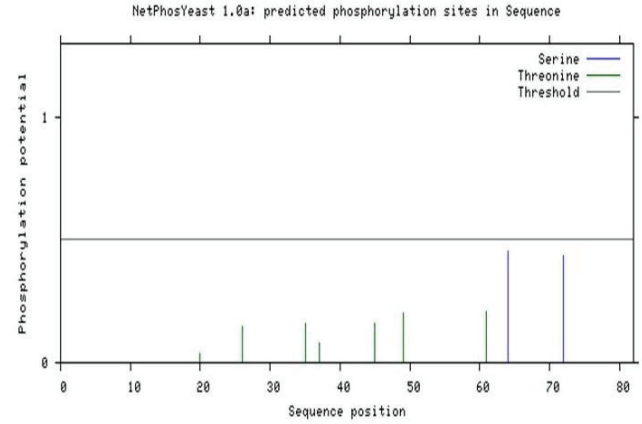

Fig7: Predicted Phosphorylation sites in E-Protein of MERS

:

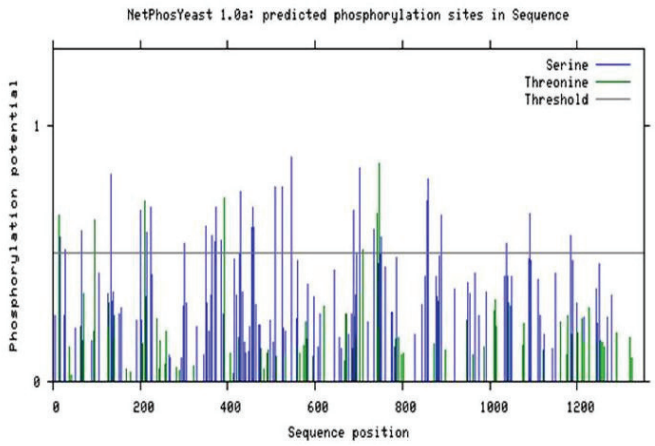

Fig 8: Predicted Phosphorylation sites in S-Protein of MERS

Table 5: Transmmebrane sequence in analysis by SOSUI

No.	N terminal	Transmembrane region	C terminal	type	length
E-protein	10	AQIIAIIKDWNFAWSLIFLLITI	32	PRIMARY	23
S-protein	60	SMALSIFSAVYPIDLASQIISGI	82	PRIMARY	23

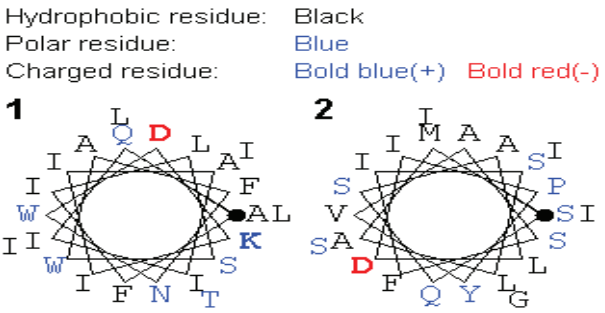

Fig. 9: Helical wheel predicetd by Pepwheel for E-Protein from MERS Co-virus

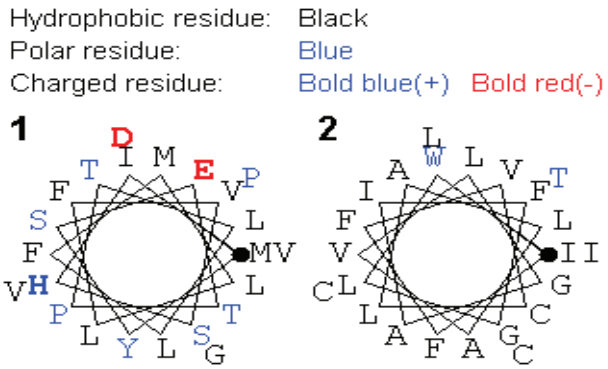

Fig. 10: Helical wheel predicetd by Pepwheel for S-Protein from MERS Co-virus

The phosphorylation site prediction is shown (Fig. 5-8). Phosphorylation site indicate the reaction pattern of the molecule during host interaction of the virus. Transmembrane predictions indicate that E protein and S-protein are primary in nature.

References

Amy F, Tobias R, Lyn G, Lee, Ayman E.G, Yoshimi E and George M. (1999). Amino Acid Substitutions Reveal Distinct Functions of Serine 186 of the ZEBRA Protein in Activation of Early Lytic Cycle Genes and Synergy with the Epstein - Barr virus R Transactivator, *J. Virol.* **73**(6): 4543-4551

Arnold K, Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*. **22**:195-201.

Bing Hao, Weimin Gong, Tsuneo K. Ferguson, Carey M. James, Joseph A. Krzycki, Michael K. Chan (2002). *A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase.* **296**(5572). Science. pp. 1462–1466.

Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology*. **294**(5): 1351-1362

De Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Ron.A.M, Galiano M, Gorballenya A.E, Memish Z.A, Erlman S.P, Leo L.M, (2013). "Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Announcement of the Coronavirus Study Group". *Journal of Virology*. **87**(14): 7790.

Gill SC and von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem*. **182**:319-326(1989).

Guruprasad K, Reddy BVB and Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng*. **4**:155-161.

Ikai AJ (1980) Thermostability and aliphatic index of globular proteins. *J. Biochem*. **88:** 1895-1898.

Ithete NL, Stoffberg S, Corman VM, Cottontail VM, Richards LR, *et al*. (2013) Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa [letter]. Emerg Infect Dis Available: http://wwwnc.cdc.gov/eid/article/19/10/13-0946_article.htm#suggestedcitation.

Jiang S, Lu L, Du L, Debnath AK. A predicted receptor-binding and critical neutralizing domainin S protein of the novel human coronavirus HCoV-EMC. *J Infect* 2013; **66**:464-466.

Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*. **37:** D387-D392.

Kyte, J and Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157** 105-132.

Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F. (2003). Motif refinement of the peroxisomal targeting signal 1 and evaluation of taxon-specific differences. J Mol Biol. 2003 May 2; **328**(3):567-79.

Peitsch, M. C. (1995) Protein modeling by E-mail Bio/Technology **13**: 658-660.

Raj VS, Mou H, Smits SL, *et al*. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC.*Nature* 2013; **495**:251-254.

Ramachandran, G.N.; Sasiskharan, V. (1968). "Conformation of polypeptides and proteins". *Advances in Protein Chemistry*. Advances in Protein Chemistry **23**: 283–437.

Srinivasan G, James CM, Krzycki JA. (2002). *Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA* **296** (5572). Science. pp. 1459–1462.

van Boheemen S, de Graaf M, Lauber C, *et al*. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio* 2012; **3**:e00473-e00412.

Woo PCY, Lau SKP, Li KSM, Tsang AKL, Yuen K-Y. Genetic relatedness of the novel human group C betacoronavirus to Tylonycteris bat coronavirus HKU4 and Pipistrellus bat coro- navirus HKU5. *Emerg Microbes Infect* 2012; **1**:e35.

Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 2012; **367**:1814-1820.