



## Antigenic Epitope Prediction towards SARS Corona Virus using Bioinformatics Analysis

<sup>\*1</sup>Sunitha Panigrahi, <sup>2</sup>Nikkitha Umesh,  
Department of Biotechnology, St. Mary's College, Hyderabad  
*\* Email ID: [sritha17@yahoo.com](mailto:sritha17@yahoo.com)*

Received: 5<sup>th</sup> August 2014, Accepted: 25<sup>th</sup> September 2014, Published: 1<sup>st</sup> November 2014

### Abstract:

Severe Accute Respiratory Syndrome is one of the types of Pneumonia, wherein the Victim suffers from severe Breathing Problem. It is a viral infection where in the Organism responsible is SARS Co V. In worst condition the patient may even die. Due to its intense effect there is a huge demand to develop a treatment methodology to cure and save the lives of the people. The best approach to solve the problem is pretension rather than cure. To prevent this dreadful disease development of the Vaccine is the only approach possible. Here the research involves the use of Computer aided reverse vaccinology approach to identify and design the most potent antigenic peptide, the destruction of which can cure the disease. More than 100 proteins of the Virus have been screened from the Public database NCBI. The proteins that are confirmed to exist are further selected. The selected proteins are compared against the human proteome using BLAST analysis to eliminate the proteins that are common among the two species. The selected proteins are further analyzed to identify the antigenic regions upon it that can actually cause the effect of pathogen. The selected peptides are further compared to identify the most effective peptide that has maximum pathogenecity. The peptide is finally designed and optimized to calculate the energy using Argus lab software.

### Key Words:

SARS Co V, Antigenic Epitope, Reverse Vaccinology, BLAST

### Introduction:

A new and deadly clinical syndrome now called severe acute respiratory syndrome (SARS) was brought to the attention of the WHO by Dr. Carlo urbani and his colleagues in a Vietnamese hospital in February 2003<sup>[1]</sup>. WHO, the medical staffs in hospitals where the disease had appeared, and local and regional governments, together with a dozen cooperating laboratories across the globe, immediately responded. They provided a pro-visional case definition to identify the extent and geo-graphic distribution of the outbreak <sup>[2]</sup> the infectious agent,

and travel advisories and quarantines to limit the spread of the disease <sup>[3]</sup> (CoV), designated SARS-CoV was quickly identified as the etiological agent <sup>[4, 5]</sup> have not established, evidence suggests that it is an animal virus that was recently transmitted to humans <sup>[6]</sup> wildlife species consumed as delicacies in southern China, including Himalayan masked palm civets, Chinese ferret badgers, and raccoon dogs, possess antibodies consistent with natural infection with related CoVs, laboratory investigations to identify .Although the origin of the virus. A novel corona virus.

### Materials and Methods:

#### Collection of the Protein Sequences of SARS CoV From Public Data Base NCBI:

NCBI is one of the worlds premier websites which is a master database having linked to several other secondary data bases. It stores information related to the annotated sequences of Genes and proteins from more than 1000 organisms. It receives information from all round the globe. It is a redundant database accepting both individual and bulk submissions. In the current work NCBI was employed to collect more upto 100 protein sequences from SARS CoV.

#### Screening for the Similarity with Human Proteins:

To identify the protein of SARS that do not share any similarity with the human proteins BLAST tool from NCBI has been used. The BLAST P tool searches a specified database against the user entered sequences to identify the similarity if any between them. The tool is free and is available in the NCBI Data base. Those viral proteins that share similarity to the human proteins must be eliminated from the list of the selected proteins.

#### Protein Variability server for Antigenic Propensity Calculation:

To identify the antigenic regions and sites located within the protein sequences and also to compare the antigenic propensity of one peptide to another PVS Server has been used. The peaks obtained would guide to identify the highest antigenic regions.

**Calculation of Accessible Area on the Surface of the Peptide:**

After the selected of the peptides and calculation of their Propensity, the next parameter to screen the peptides is the accessible area on the surface of the peptide that can bind to the external ligands. The calculation can be done using the Formula

$$AA = \text{Valine} + \text{Leucine} + \text{Cysine} / \text{Peptide Length}$$

Those peptides having high Area can be preferred for the development of a vaccine.

**Peptide Construction in Argus Lab:**

From the results of PVS server the selected peptides are finalized and are constructed in Agrus lab for

energy calculation and optimization. Based upon the number of proteins selected after PVS and optimized in argus the peptides having lowest energy can be finalized for the development of a vaccine.

**Results and Discussion:**

Nearly 100 sequences from SARS CoV, were screened and out of which 50 were selected for the analysis. Among the proteins selected the screening was performed step by step to finalize and select a best peptide that can be used for the preparation of Vaccine.

**Table 1: The Proteins selected based on antigenic propensity along with their propensity values**

SNO	PROTEIN	PEPTIDE	ANTIGENIC PROPENSITY	ACCESSIBILITY AREA ON SURFACE
1	Nucleocapsid Protein Dimerization Domain, Chain B [gi 99032575]	FKDNVILLNKHIDAY	1.1	0.2
2	Nucleocapsid Protein Dimerization Domain, Chain A [gi 99032574]	FKDNVILLNKHIDAY	1.1	0.2
3	Oligomerization Domain Of Sars Coronavirus Nucleocapsid Protein, Chain H, [gi 146386432]	FKDNVILLNKHIDAY	1.125	0.2
4	Oligomerization Domain Of Sars Coronavirus Nucleocapsid Protein, Chain G, [gi 146386431]	FKDNVILLNKHIDAY	1.125	0.2
5	Oligomerization Domain Of Sars Coronavirus Nucleocapsid Protein, Chain F, [gi 146386430]	MHHHHH	1.125	----
6	Oligomerization Domain Of Sars Coronavirus Nucleocapsid Protein, Chain D, [gi 146386428]	FKDNVILLNKHIDAY	1.125	0.2
7	Oligomerization Domain Of Sars Coronavirus Nucleocapsid Protein, Chain C, [gi 146386427]	MHHHHH	1.125	---
8	Crystal Structure Of Oligomerization Domain Of Sars Coronavirus Nucleocapsid Protein, Chain A, [gi 146386425],	FKDNVILLNKHIDAY	1.125	0.2
9	Membrane protein, SARS coronavirus MA15, [gi 270309780]	RFLYIIKLVFLWLLWPVTLA CFVLAAYVRI	1.215	0.4
10	Membrane protein, SARS coronavirus MA15 ExoN1, [gi 270309749]	RFLYIIKLVFLWLLWPVTLA CFVLAAYVRI	1.215	0.4
11	Protein, SARS coronavirus MA15, [gi 270309733]	RFLYIIKLVFLWLLWPVTLA CFVLAAYVRI	1.215	0.4
12	Membrane protein, SARS coronavirus MA15, [gi 270309638]	RFLYIIKLVFLWLLWPVTLA CFVLAAYVRI	1.215	0.4
13	Membrane protein, SARS coronavirus MA15 ExoN1, [gi 270309623]	RFLYIIKLVFLWLLWPVTLA CFVLAAYVRI	1.215	0.4
14	membrane protein, SARS coronavirus HKU-39849, [gi 387913001]	RFLYIIKLVFLWLLWPVTLA CFVLAAYVRI	1.215	0.4
15	membrane protein [SARS coronavirus ExoN1, [gi 270309514]	RFLYIIKLVFLWLLWPVTLA CFVLAAYVRI	1.215	0.4

From the above table 1, it can be concluded that only 2 peptides namely **FKDNVILLNKHIDAY** and **RFLYIIKLVFLWLLWPVTLACFVLAAYVRI** can be selected for the development of a vaccine. Also it was found that the same two peptides are present in several proteins of virus. Thus selection of such proteins would be a very good target for the development of vaccine.

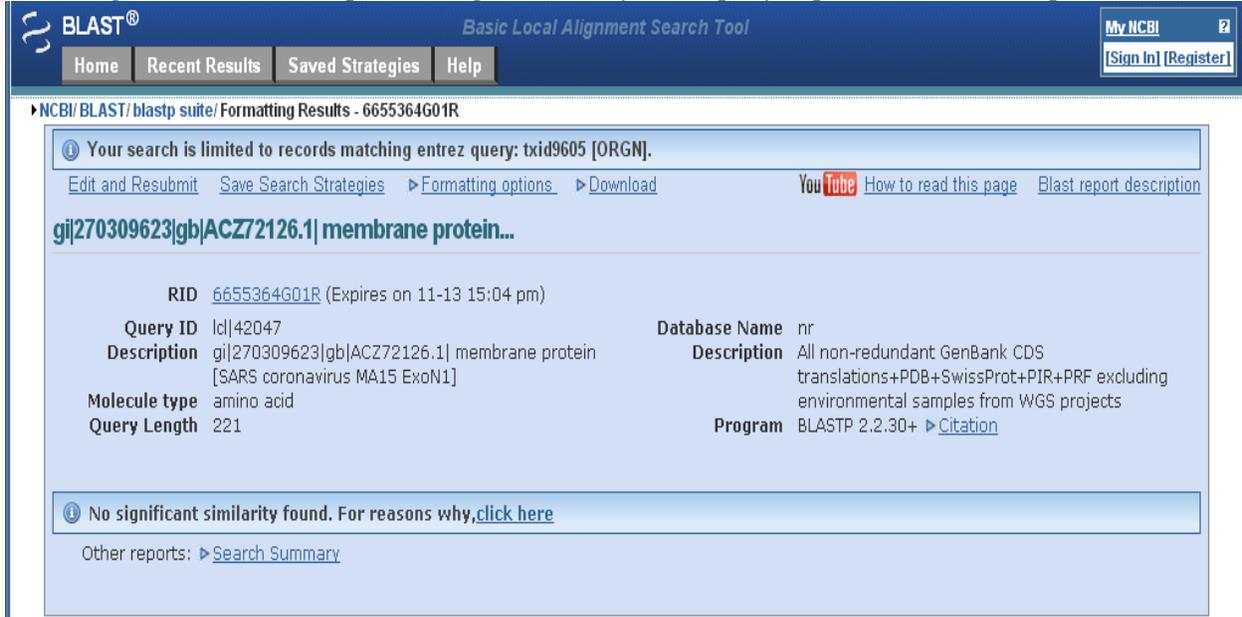
**Sequence selected in BLAST Analysis:**

Once all the sequences are collected from the database all were compared with the Homo sapiens proteome using the BLAST analysis tool. A total of 12 proteins were identified that did not share any similarity with the human proteome and were thus foreign to the human immune system. All the 13 sequences were further used for antigenic site prediction and propensity calculation.

**Sequence selected in BLAST Analysis:**

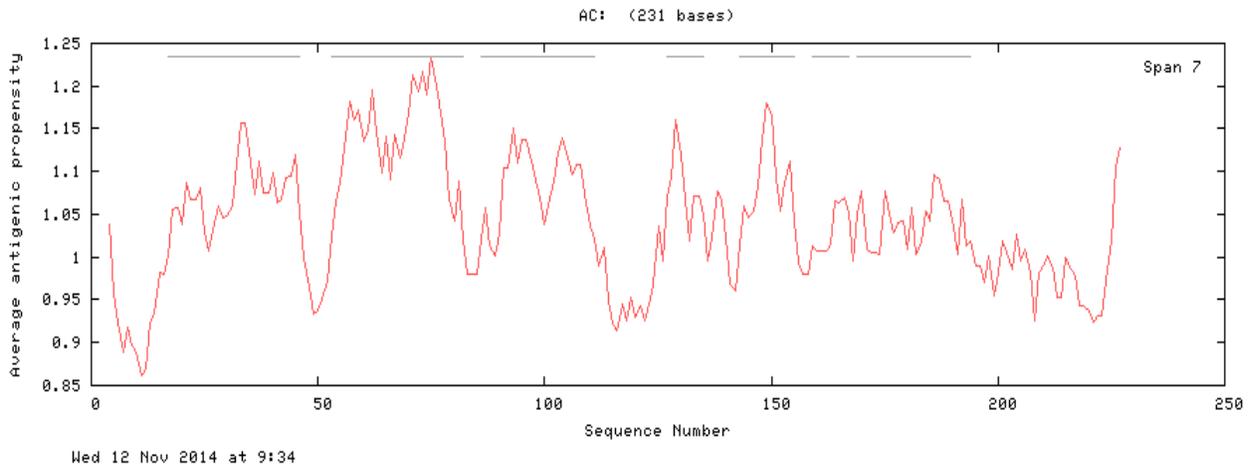
Once all the sequences are collected from the database all were compared with the Homo sapiens proteome using the BLAST analysis tool. A total of 12 proteins were identified that did not share any similarity with the human proteome and were thus foreign to the human immune system. All the 13 sequences were further used for antigenic site prediction and propensity calculation.

**Fig 1: The BLAST out put showing no similarity of the query sequence to the human proteome**



As from the above figure it can be concluded that the selected proteins shares no similarity to the human proteome. Thus, they can be antigenic to the human immune system. All the proteins that show similar results can be selected for further analysis.

**Table 2: Antigenic Propensity Cure Obtained for one of the Membrane protein**

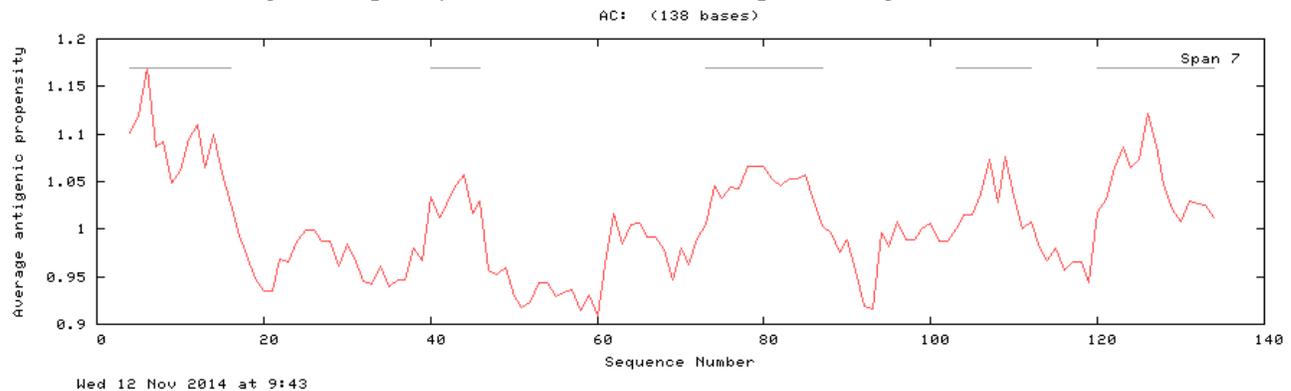


There are 7 antigenic determinants in your sequence:

n	Start Position	Sequence	End Position
1	17	ITVEKCLKQLLEQWNLVIGFLFLAWIMLLQF	46
2	53	RFLYIIKLVFLWLLWPVTLACFVLAAYRI	82
3	86	TGGIAIAMACIVGLMWLSYFVASFRL	111
4	127	ILLNVPLRG	135
5	143	MESELVIGAVIIR	155
6	159	RMAGHPLGR	167
7	169	DIKDLPKEITVATSRTLSTYYKLGASQ	194

The above Curve and the table are related to the peptide **RFLYIIKLVFLWLLWPVTLACFVLAAYRI** of the finalized peptides. In the graph only one of the peaks touches the threshold and that lies in the region 53 TO 82. Thus it was selected for further analysis. The same is the case with the second peptide as shown below.

Table 3: Antigenic Propensity Curve Obtained for the 2<sup>nd</sup> protein Oligomerization Domain



There are 5 antigenic determinants in your sequence:

n	Start Position	Sequence	End Position
1	4	CACSDSCMHHHHH	16
2	40	KQYNVTQ	46
3	73	KHWPQIAQFAPSASA	87
4	103	TWLTYHGAIK	112
5	120	FKDNVILLNKHIDAY	134

The above Curve and the table are related to the peptide **FKDNVILLNKHIDAY** of the finalized peptides. In this graph also only one of the peaks touches the threshold and that lies in the region 120 to 134. Thus it was selected for further analysis.

From among the two peptides finalized only one can be used for the vaccine development. The lengthy peptide no second i.e **RFLYIIKLVFLWLLWPVTLACFVLAAYRI**

Cannot be processed further due to its length as only short peptides can be used as Vaccine molecules. Thus only one peptide **FKDNVILLNKHIDAY** can be used for vaccine development.

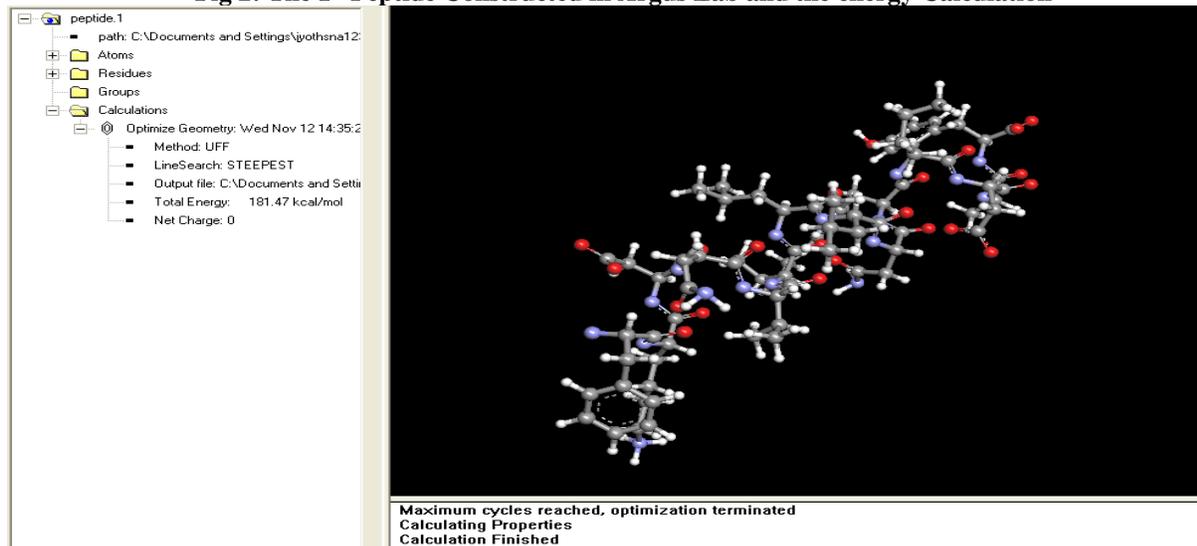
### Construction of Peptide in the ARGUS LAB:

As per the above analysis two peptides of the length 15 and 30 were selected after the final screening, however only one can be proceed for vaccine

development. Thus the second peptide with the Accessibility area 0.2 is used for model construction and energy evaluation.

### Model Construction of the Peptide FKDNVILLNKHIDAY Lab followed by Energy Calculation in ARGUS LAB

Fig 2: The 1<sup>st</sup> Peptide Constructed in Argus Lab and the energy Calculation



From the above Fig it can be observed that the peptide **FKDNVILLNKHIDAY** was constructed and the energy was found to be 181.47Kcal/mol.

### Conclusion:

The current work aimed to develop or identify a potential vaccine candidate (protein) against the SARS Corona Virus causing SARS Disease. The proteins of the organism are collected from the Public database NCBI followed by their comparison to human proteome using BLAST for identification of antigenic proteins. Later the antigenic proteins were further analyzed for the identification of Antigenic Propensity and antigenic region on the proteins. Based on the results one of the peptide is finalized and selected. The selected peptide was FKDNVILLNKHIDAY which was commonly present in several nucleocapsid proteins of the virus especially present in the dimerization region. The peptide was constructed in Argus lab and the energy was calculated to be 181.47Kcal/mol. The peptide can be artificially synthesized and attenuated so as to make it a potential vaccine to prevent SARS. This work lays a foundation for then development of the selected peptide vaccine.

### Reference:

- 1) 2003. World Health Organization. Severe acute respiratory syndrome (SARS): multi-

country outbreak.  
[http://www.who.int/csr/don/2003\\_03-16/en/](http://www.who.int/csr/don/2003_03-16/en/).

- 2) 2003. Preliminary clinical description of severe acute respiratory syndrome. MMWR Morb. Mortal. Wkly. Rep.52:255-256.
- 3) 2003. WHO recommended measures for persons undertaking international travel from areas affected by severe acute respiratory syndrome (SARS). Wkly. Epidemiol. Rec. 78:97-120.
- 4) Cyranoski D, Abbott A. Virus detectives seek source of SARS in China's wild animals. Nature. 2003;423:467
- 5) Holmes KV. Coronaviruses. In: Knipe DM, Howley PM, editors. Fields virology. Philadelphia: Lippincott-Raven; 2001. p. 1187-203.
- 6) van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJM, Wolthers KC, et al. Identification of a new human coronavirus Nat Med. 2004;10:368-73.