



Insilico Vaccine Design against Zaire Ebolavirus

*Ambika Iyer

Department of Biotechnology, B.V. Bhoomaraddi College of Engineering & Technology, Hubli, Karnataka

*Email Id: ambikaiyer@yahoo.com

Received: 6th Oct 2015, Accepted: 13th Oct 2015, Published: 1st Jan 2016

Abstract

The best approach for the development of a vaccine against a specific infection or a microbial disease is Insilico vaccine designing that is not only accurate but also requires minimum time, labour and economy. This approach can also be used to analyze the target structures for possible binding/active sites, generate candidate molecules, check for their drug likeness, dock these molecules with the target, rank them according to their binding affinities, further optimize the molecules to improve binding characteristics. During vaccine designing the disease to be selected should be a viral or a bacterial disease but not due to a protozoan, as they cannot be attenuated and injected. The main goal of the current work is to develop a vaccine against the disease Ebola. Computational tools offer the advantage of delivering new drug candidates more quickly and at a lower cost. This review provides succinct overview of several methods in drug discovery by using insilico approaches on identifying drug targets where there are genes or proteins associated with specific diseases.

Keywords

Ebola, protein-protein BLAST, PVS tool, insilico methods, *Zaire ebolavirus*

Introduction



Ebolavirus is a virological taxon included in the family *Filoviridae*, order *Mononegavirales*. It is a disease of humans and other primates caused by ebolaviruses. The disease has a high risk of death, killing between 25 and 90 percent of those infected, with an average of about 50 percent. [Kuhn J. H. et al, 2010] The Ebola virus (EBOV) protein VP24 inhibits type I and II interferon (IFN) signaling by binding to NPI-1 subfamily karyopherin α (KPNA) nuclear import proteins, preventing their interaction

with tyrosine-phosphorylated STAT1 (phospho-STAT1). This inhibits phospho-STAT1 nucleaimport. A biochemical screen now identifies heterogeneous nuclear ribonuclear protein complex C1/C2 (hnRNP C1/C2) nuclear import as an additional target of VP24. Co-immunoprecipitation studies demonstrate that hnRNP C1/C2 interacts with multiple KPNA family members, including KPNA1. Interaction with hnRNP C1/C2 occurs through the same KPNA1 C-terminal region (amino acids 424–457) that binds VP24 and phospho-STAT1. The ability of hnRNP C1/C2 to bind KPNA1 is diminished in the presence of VP24, and cells transiently expressing VP24 redistribute hnRNP C1/C2 from the nucleus to the cytoplasm. These data further define the mechanism of hnRNP C1/C2 nuclear import and demonstrate that the impact of EBOV VP24 on nuclear import extends beyond STAT1. [Isaacson M. et al, 1976] The five characterized species of the *Ebolavirus* genus are Zaire ebolavirus, Sudan ebolavirus, Reston ebolavirus, Tai Forest ebolavirus and Bundibugyo ebolavirus. [Feldmann H. et al, 2005] The species *Zaire ebolavirus* is a virological taxon. The species has a single virus member, Ebola virus. The species was introduced in 1998 as *Zaire Ebola virus*. In 2002, the name was changed to *Zaire ebolavirus*. It has a genomic sequence that differs from the type virus by less than 30%. [Mayo M. A. et al, 2002]

Materials & Methods

Collection of Gene Sequence:

The protein sequences were collected from National Center for Biotechnology Information NCBI database. URL: www.ncbi.nlm.nih.gov. The organism selected is *Zaire ebolavirus*. To receive the protein sequence one has to type the name of the organism in the search box of the NCBI homepage and set the database to proteins and search. Collect the FASTA sequences of all the proteins present in Ebola Virus. These proteins will serve as pool for vaccine screening. Exclude those proteins that are prefixed as Predicted, Hypothetical, Unnamed, etc. Then we have to check whether any of it matches with the human proteome.

Primary screening to Eliminate the Non foreign proteins:

The Basic Local Alignment Search Tool Protein (BLAST) is a standard protein – protein BLAST. URL:

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome

BLAST of each of these sequences with the Homo sapiens proteins has been performed to identify the similarity among the sequences. Input can be given in the form of accession number or gi number of the Protein sequence or Protein sequence in the FASTA format can be given in the search box present in the Enter query sequence column. All the parameters are default and can be changed as per user requirements. For executing Blast program one has to click on the "BLAST" button. In this step our main aim is to eliminate all those proteins of pathogen that share more than 30% identity with the human proteome. A vaccine is the protein of the pathogen that has no similarity or very less (less than 30%) similarity to the human proteome. [Altschul S.F. et al, 1990]

Protein Variability Server Tool:

Next step is to determine the antigenic regions on the protein using Protein Variability Server (PVS) Tool. URL: <http://imed.med.ucm.es/PVS/> In the home page shown above click on 'Tools' then click on 'Antigenic Peptide Prediction Tool' the click on 'www server'. In the page replace the existing sequence with the sequence of your interest and then click on 'Submit'. [Diez-Rivero C.M. et al, 2009]. The antigenic propensity of the sequences is calculated by the tool and the result is interpreted from the graph.

SAA Calculation.

All the peptides selected based on their PVS result are now subjected for SAA calculation. Here the total no of Valines, Cysteins and Leucines present in the peptide are counted and divided by the length of the peptide to obtain the Surface Accessibility area value. This is very important for docking to the specific binding site. Thus those peptides with highest SAA are only considered for the peptide development

Building Peptide in Argus Lab:

The next step is to build the 3D structure of the Peptide in Argus lab. Argus Lab is free software and can be downloaded from the URL: <http://www.arguslab.com/arguslab.com/ArgusLab.html>

This step is very important to calculate the energy of the peptide and select the best one based on the lowest energy possible. The peptide with lowest energy would have a high stability. The peptide building in the Argus lab can be done using the option Builder toolkit. The peptide builder is available in the package thus the 3D structure can be built. Once the structure is developed than the energy is calculated using clean geometry option.

Results & Discussion

The organism causing the disease Ebola is *Zaire ebolavirus*. The protein sequences of the organism are collected which will pool for vaccine screening. These proteins were collected from NCBI database. However any source can be selected to collect the proteins of the organism. Few among these proteins that are foreign to the human proteome may be responsible for the onset of the diseased condition. To select the protein several screening techniques are employed.

BLAST P is a tool to compare two or more protein sequences in a pairwise manner. The sequences that share either no similarity or less than 30% similarity with the human proteome are selected for the further screening. This pairwise sequence alignment tool would enable this comparison of all the bacterial protein to the complete proteome of the human.

Through BlastP has been used to eliminate the non-foreign proteins. The query sequence is represented by the broad red bar at the top. Database hits are shown aligned to the query as colored bars/lines. The color of the bar indicates the alignment score with reference to color key. The hits are arranged in descending order based on alignment score. By moving the mouse over the bar directs the user to the specific alignment of the corresponding hit and query sequence. In this step our main aim is to eliminate all those proteins of pathogen that share more than 30% identity with the human proteome. A vaccine is the protein of the pathogen that has no similarity or very less (less than 30%) similarity to the human proteome.[Gish W. et al, 1993] The proteins selected are,

Peptide 1:

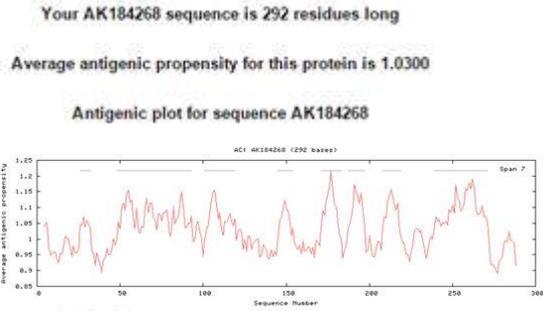
VP30 minor nucleoprotein [Zaire ebolavirus]

Accession number: AKI84268

>gi|824041183|gb|AKI84268.1| VP30 minor nucleoprotein [Zaire ebolavirus]

MEASYERGRPRAARQHSRDGHDHHVRRASS
SRENYRGEYRQRSASQVRVPTVFHKKRVEP
LTVPPAPKDICPTLKKGFCDSSFCCKDHQLE
SLTDRELLLLIARKTCGSVEQQLNITAPKDSR
LANPTADDFQQEEGPKITLLTLIKTAEHWARQ
DIRTIEDSKLRALLTLCAVMTRKFSKSQLSLL
CETHLRREGLGQDQAEPVLEVYQRLHSDKG
GSFEAALWQQWDRQSLIMFITAFLNIALQLPC
ESSAVVVSGLRTLVPQSDNEEASTNPGTCSW
SDEGTP

The final score obtained was 37 and the Energy value is 8.2 Through protein variability server tools we can determine the antigenic regions on the protein. [Garcia-Boronat M. et al, 2008]



The above picture shows the result of PVS server. It indicates the presence of only one highest peak touching the threshold value of propensity.

There are 8 antigenic determinants in your sequence:

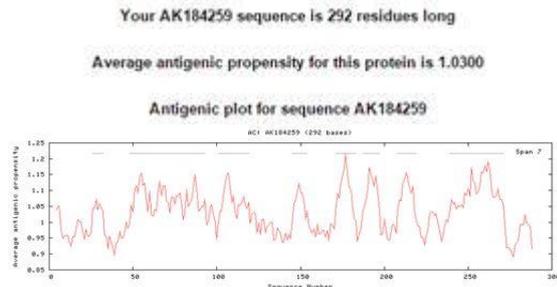
n	Start Position	Sequence	End Position
1	26	HDHHVRA	32
2	48	RSASQVRVPTVFHKKRVEPLTVPPAPKDIPTLKKGFLCDSFFCKK	93
3	101	TDRELLLIARKTCGSVEQ	119
4	145	PKITLLTIK	154
5	171	KLALLTLCAVMT	183
6	187	SKSQLLCET	197
7	208	QAEPLVEVYQRL	219
8	239	SLIMFITAFNLIALQLPCESSAVVVSGLRTLVP	271

The above tables shows that the antigenic propensity of the peptide that touches the threshold line is 187 to 197 which can be selected for further peptide building.

Peptide 2:
VP30 minor nucleoprotein [Zaire ebolavirus]

Accession number: AKI84259
 >gi|824041130|gb|AKI84259.1| VP30 minor nucleoprotein [Zaire ebolavirus]
 MEASYERGRPRAARQHSRDGHDHHRARSS
 SRENYRGEYRQSRASQVRVPTVFHKKRVEP
 LTVPPAPKDIPTLKKGFLCDSFFCKKDHQLE
 SLTDRELLLIARKTCGSVEQQLNITAPKDSR
 LANPTADDFQEEGPKITLLTIKTAEHWARQ
 DIRTIEDSKLRALLTLCAVMTRKFSKSQLSLL
 CETHLRREGLGQDQAEPLVEVYQRLHSDKG
 GSFEAALWQQWDRQSLIMFITAFNLIALQLPC
 ESSAVVVSGLRTLVPQSDNEEASTNPGTCSW
 SDEGTP

The final score obtained was 37 and the Energy value is 8.2 Through protein variability server tools we can determine the antigenic regions on the protein.



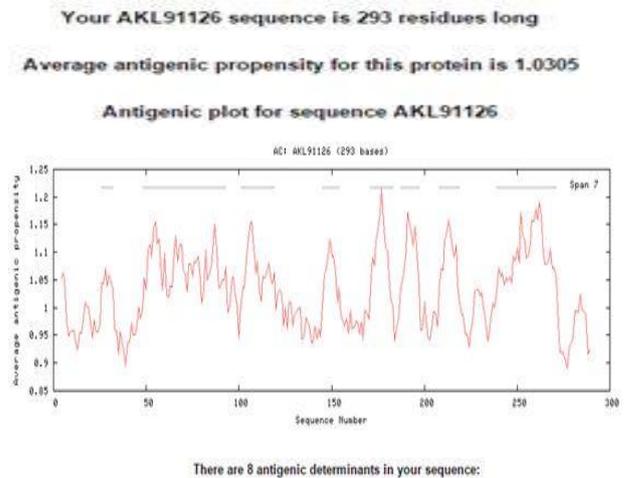
There are 8 antigenic determinants in your sequence:

n	Start Position	Sequence	End Position
1	26	HDHHVRA	32
2	48	RSASQVRVPTVFHKKRVEPLTVPPAPKDIPTLKKGFLCDSFFCKK	93
3	101	TDRELLLIARKTCGSVEQ	119
4	145	PKITLLTIK	154
5	171	KLALLTLCAVMT	183
6	187	SKSQLLCET	197
7	208	QAEPLVEVYQRL	219
8	239	SLIMFITAFNLIALQLPCESSAVVVSGLRTLVP	271

Peptide 3
VP30 [Zaire ebolavirus]

Accession number: AKL91126
 >gi|831182909|gb|AKL91126.1| VP30 [Zaire ebolavirus]
 MEASYERGRPRAARQHSRDGHDHHRARSS
 SRENYRGEYRQSRASQVRVPTVFHKKRVEP
 LTVPPAPKDIPTLKKGFLCDSFFCKKDHQLE
 SLTDRELLLIARKTCGSVEQQLNITAPKDSR
 LANPTADDFQEEGPKITLLTIKTAEHWARQ
 DIRTIEDSKLRALLTLCAVMTRKFSKSQLSLL
 CETHLRREGLGQDQAEPLVEVYQRLHSDKG
 GSFEAALWQQWDRQSLIMFITAFNLIALQLPC
 ESSAVVVSGLRTLVPQSDNEEASTNPGTCSW
 SDEGTP

The final score obtained was 37 and the Energy value is 8.2 Through protein variability server tools we can determine the antigenic regions on the protein.



n	Start Position	Sequence	End Position
1	26	HDHHVRA	32
2	48	RSASQVRVPTVFHKKRVEPLTVPPAPKDIPTLKKGFLCDSFFCKK	93
3	101	TDRELLLIARKTCGSVEQ	119
4	145	PKITLLTIK	154
5	171	KLALLTLCAVMT	183
6	187	SKSQLLCET	197
7	208	QAEPLVEVYQRL	219
8	239	SLIMFITAFNLIALQLPCESSAVVVSGLRTLVP	271

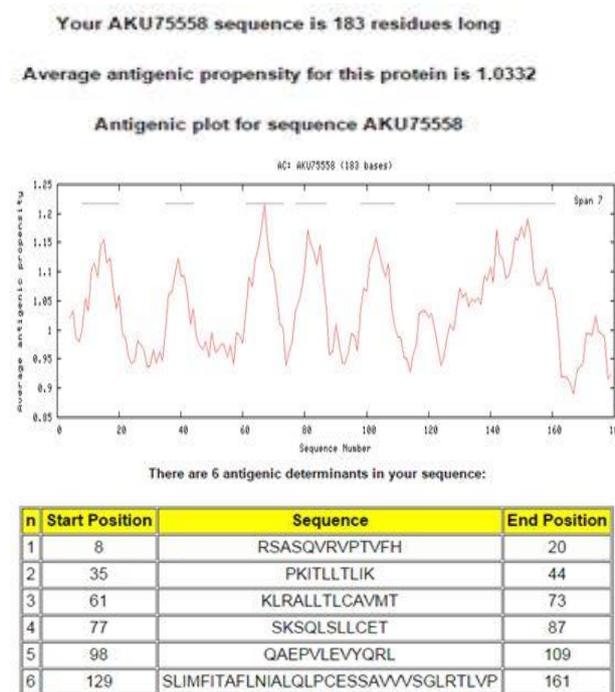
Peptide 4:

VP30, partial [Zaire ebolavirus]

Accession number: AKU75558

>gi|913075117|gb|AKU75558.1| VP30, partial [Zaire ebolavirus]

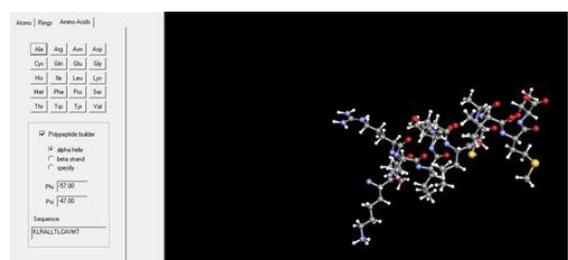
RQSRASQVRVPTVFHKXXXXXXXXXXXXXXXXXXXX
 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 XANPTADDFQEEGPKITLLTLIKTAEHVAR
 QDIRTIEDSKLRALLTLCVMTTRKFSKSQLSL
 LCETHLRREGLGQDQAEPLVEVYQRLHSDKG
 GSFEAALWQQWDRQSLIMFITAFNLIALQLPC
 ESSAVVSGLRRTLVPQSDNEEASTNPGTCSW
 SDEGTP



The peptides with the highest peak is selected as the antigenic peptides. These peptides are designed using the software Argus Lab. Highest peak = 171-183

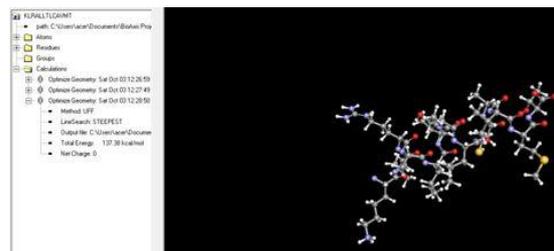
Construction of the peptide 3D structure in Argus Lab

The peptide KLRALLTLCVMT with the highest peak Highest peak = 171-183



Total Energy = 137.38kcal/mol

Antigenic Peptide: SLIMFITAFNLIALQLPCESSA VVVSGLRTLVP at the position 239-271



Total Energy = 514.06kcal/mol

Summarization of the result:

Length of Peptide	SAA	Total Energy of Peptide
13	177	137.38
33	255	514.06

Based upon the results of the Peptide SAA and total energy, the best peptides is selected which is having Maximum SAA and Minimum Total Energy. Then it is said to be the best vaccine candidate to be used for the vaccination against the disease Ebola.

Conclusion

In vaccine designing the user is working on the complete proteome of the selected pathogenic virus. We have selected the organism as *Zaire ebolavirus*. Once the proteome is downloaded the next step is to remove all the proteins that are Predicted, Hypothetical, Unnamed or Putative. Then the selected proteins must be screened for the Foreignness based on the tool BlastP. In BlastP all the foreign peptides can be selected which can be further processed to Antigenic Peptide Prediction Using Protein Verification Server tool. Based on the results of PVS, best peptides having maximum Antigenic Propensity can be selected. From all the peptides selected based on PVS, SAA calculation must be performed to detect the best peptide having both antigenicity and accessibility to the external ligand. The Highest peaks were found to be 171-183 and 239-271

Antigenic peptide = KLRALLTLCAVMT and SLIMFITAFNLIALQLPCESSAVVVSGLRTLVP. The selected peptides are designed in Argus lab and they are optimized followed by their energy Calculation. Finally the peptide that has maximum SAA, Foreignness and Minimum energy can be selected as the Best Vaccine candidate. The peptide with maximum SAA and minimum total energy is KLRALLTLCAVMT which can be further used to develop vaccine against the disease Ebola.

Acknowledgement

I would like to acknowledge BioAxis DNA Research Center Pvt. Ltd. for their endeavor support & logistics to carry out this project.

References

1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." *J. Mol. Biol.* 215:403-410.
2. Diez-Rivero CM & Reche PA. (2009). Discovery of conserved epitopes through sequence variability analysis. *Bioinformatics for immunomics; Series: Immunomics Reviews: Vol. 3.* Pg. 95-101.
3. Feldmann, H.; Geisbert, T. W.; Jahrling, P. B.; Klenk, H.-D.; Netesov, S. V.; Peters, C. J.; Sanchez, A.; Swanepoel, R.; et al. (2005). "Family Filoviridae". In Fauquet, C. M.; Mayo, M. A.; Maniloff, J.; Desselberger, U.; Ball, L. A. *Virus Taxonomy—Eighth Report of the International Committee on Taxonomy of Viruses*. San Diego, USA: Elsevier/Academic Press. pp. 645–653.
4. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA (2008) PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res.*, **36**:W35-41
5. Gish, W. & States, D.J. (1993) "Identification of protein coding regions by database similarity search." *Nature Genet.* 3:266-272.
6. Isaacson, M; Sureau, P; Courteille, G; Pattyn, SR;. "Clinical Aspects of Ebola Virus Disease at the Ngaliema Hospital, Kinshasa, Zaire, 1976"
7. Kuhn, J. H.; Becker, S.; Ebihara, H.; Geisbert, T. W.; Johnson, K. M.; Kawaoka, Y.; Lipkin, W. I.; Negredo, A. I.; Netesov, S. V.; Nichol, S. T.; Palacios, G.; Peters, C. J.; Tenorio, A.; Volchkov, V. E.; Jahrling, P. B. (2010). "Proposal for a revised taxonomy of the family Filoviridae: Classification, names of taxa and viruses, and virus abbreviations". *Archives of Virology* **155** (12): 2083–2103.
8. Mayo, M. A. (2002). "ICTV at the Paris ICV: results of the plenary session and the binomial ballot". *Archives of Virology* **147** (11): 2254–60.