



In silico Vaccine designing for Lyme disease

*S. Deva Surya

Bharathidasan University-Trichy

*Email: devasurya27@gmail.com

Received: 23rd May 2016, Accepted: 30th May 2016, Published: 1st May 2016

Abstract

The vaccine designing project for Lyme disease involves various Bio-informatics tools and techniques for Insilico designing of a vaccine. Lyme disease is caused by the bacterium *Borrelia Burgdorferi* also called as Lyme Borreliosis or Lyme Arthritis. The bacterium uses 'Ticks' as vectors and infects Humans through tick bites. The symptoms can start from simple skin rashes to arthritis or paralysis or even memory loss if untreated. *Borrelia Burgdorferi* is a Gram Negative bacterium of *Spirochaetaceae* family. In the vaccine designing procedure all the protein sequences available for the bacterium are retrieved from NCBI and performed a primary screening to eliminate hypothetical or putative or unnamed sequences from the list and secondary screening to eliminate sequences that shows more than 30% of similarity with Human genome. As a vaccine is needed to be given to a human host it's very important to check for the maximum foreignness to humans, bio-informatics tools like BLAST and TFASTy were used for this purpose to eliminate similar and identical sequences. Using the TFASTy tool, (which compares our protein query against translated genome sequences of humans) the sequences that shows zero similarity are chosen for further steps. Surface Accessibility Area is another aspect to be considered and the sequence with highest SAA value is taken and it's antigenic sites were found using protein variability server and the sequence with highest SAA value among them was chosen as the potential vaccine candidate and further modelling was done using the software 'ARGUS LAB'.

Introduction

Lyme disease is transmitted to humans by the bite of infected ticks of the Ixodes genus. Usually, the tick must be attached for 36 to 48 hours before the bacteria can spread. In North America, the only bacterium involved is *Borrelia burgdorferi sensu stricto*, while in Europe and Asia, the bacteria *Borrelia afzelii* and *Borrelia garinii* are also causes of the disease. The disease does not appear to be transmissible between people, by other animals, or through food. Diagnosis is based upon a combination of symptoms, history of tick exposure,

and possibly testing for specific antibodies in the blood. Blood tests are often negative in the early stages of the disease. Testing of individual ticks is not typically useful.

Lyme disease was clinically described as an infectious illness by Dr. Alan Steere and colleagues in 1977 and is currently the leading vector-borne disease in the United States. Steere et al. suggested that the epidemiology of Lyme disease indicated transmission by an arthropod vector due to the geographic clustering of patients in rural areas and the seasonal occurrence of the symptoms. Subsequently, Dr. Willy Burgdorfer and co-workers observed spirochetes in the mid-gut tissues from ticks collected in a Lyme disease endemic area. These spirochetes produced a skin rash resembling erythema migrans when injected into rabbits, and sera from Lyme disease patients reacted with the bacteria in indirect immune-fluorescence assays. In recognition of this discovery, the bacterium was named *Borrelia Burgdorferi*.

The bacteria, known as spirochetes for their unique corkscrew shape and twisting propulsion, are slow to replicate and one of few bacteria that have learned to survive without iron, instead relying on a supply of manganese for their survival (lending to potential treatment strategies intending to starve the bacteria of manganese). Unlike most disease-causing bacteria, *Borrelia burgdorferi* do not appear to emit a toxin. Instead, the bacterium seems to have a direct interaction with the cell tissues it infects. Also interesting is the bacteria's slow replication rate, meaning that the number of bacteria found in a host remains quite small even after active infection.

B. burgdorferi has a very unique genome. It consists of one large linear chromosome that is 910,725 base pairs long. This chromosome contains approximately 853 genes coding for basic functions such as DNA replication, transcription, translation as well as transport and metabolic systems. However, in addition to this chromosome, the bacterium also contains 21 other linear and circular plasmids that add an additional 533,000 base pairs of DNA. The genomic organization of *B. burgdorferi* is unique due to the high number of plasmids. Research has shown that some bacteria lacking a complete set of plasmids are unable to successfully infect their host, leading researchers to

believe that the plasmids may encode virulent DNA (Fraser). One plasmid named lp25 has been found to be necessary for *Borrelia* infection. An lp25 gene named BBE22 has been discovered to encode for a nicotinamidase which by itself can maintain the ability for the bacteria to infect the host even when the plasmid is removed (Purser). However, in general, the sequenced genome does not contain any obvious genes coding for pathogenesis, and therefore, the mechanisms of *B. burgdorferi* infections are still a hot topic of research.

Materials and Methods:

Protein Sequences retrieval

The whole proteome of the bacterium are required to be screened for vaccine designing so, the protein sequences available till date are to be retrieved from NCBI protein database

Primary screening

This screening step is done to eliminate the sequences that are hypothetical, unnamed, predicted or putative.

BLAST (Basic Local Alignment Search Tool)

Using protein Blast the retrieved sequences were compared against *Homo sapiens* proteome and the sequences having more than 30% similarity are eliminated.

Secondary screening

To ensure maximum foreignness to humans, a secondary screening is done using the tool TFASTy, which compares our protein query against translated human genomes. From the result the sequences with 0% similarity to *Homo sapiens* are selected.

Surface Accessibility Area

From the remaining sequences the one with highest surface accessibility area is chosen, which is calculated by the number of Valine, Leucine and Cysteine residues divided by the length of the sequence.

Finding Antigenic sites

The antigenic sites present in that sequence must be found. These are the sites in the protein sequence where the antibodies from our body will bind. These sites from our sequence are found using Protein Variability Server available at <http://imed.med.ucm.es/PVS/>. This tool will find the possible antigenic sites in our sequence.

From all antigenic sites found, one sequence which has the highest SAA value and the lowest energy among them all is chosen as the vaccine candidate.

Results and discussions

Protein sequence retrieval:

The NCBI protein query for *Borrelia Burgdorferi* came out with more than fifty thousand sequences and they are retrieved in some standard format like FASTA.

Screening

By primary screening the hypothetical, unnamed, putative and predicted sequences are eliminated and got reduced to 26857 sequences, then following the BLAST and secondary screening further more sequences are reduced and sequences showing any similarity with human beings are eliminated

TFASTy:

The TFASTy tool is available at <http://workbench.sdsc.edu/> is used for the confirmation of maximum foreignness to *homo sapiens* genome and some best sequences with no similarity with humans are selected for further analysis.

Surface Accessibility Area

From the sequences obtained from screening procedures the SAA values are calculated and tabulated for each, which is done by adding the number of valine, leucine and cysteine residues and dividing it by the length of the sequence. From which the protein 'Undecaprenyl diphosphate' had the highest SAA value.

Protein Variability Server

To find the antigenic peptide sites present in the protein sequence, the Protein variability server available at <http://imed.med.ucm.es/PVS/> is used and the results showed that the protein sequence contains 10 antigenic peptide sequences.

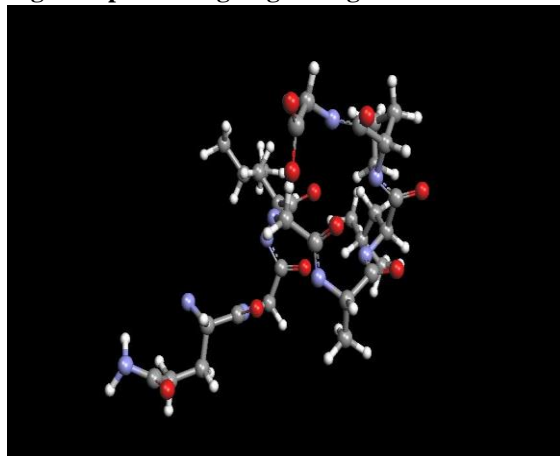
Fig 1: Antigenic sites found in PVS

Sequence
ILSAIILGIIQGITFLPISSSGHLLFRHFINKLSIIFDIYHLATVLIIVYRKRILELFLTFIRF
TNLKLLILIIITVTGVGTFS
FTLSFVLNFIITGILIMLEFNFKV
GNILLAGI
QGLGALPG
GITIFSASVIG
AFEISFLIPIVFGAILKH
EFYDIFMVLN
EINLGALVAFVVGIFSIN
KLYYFSIYLFALSIVCY

Argus Lab

The antigenic peptide sequences obtained from PVS are designed in Argus Lab using the amino acid builder tool and followed by energy minimisation step. The surface accessibility values for each residue are also found and the sequence with highest SAA value and lowest energy is chosen as the potential vaccine candidate.

Fig 2: Peptide designing at Argus Lab



The antigenic sequence “QGLGALPG” had the lowest Energy of 56.19kcal/mol and SAA value of 0.25, which makes it the best sequence to be selected.

Conclusion

From the Insilico project it's been concluded that the protein undecaprenyl diphosphate has the antigenic sequence ‘QGLGALPG’ which is a potential vaccine candidate made using immuno-informatics approach and could be taken for further synthesis in lab and tests on clinical trials. Even though the sequence retrieval gave 50 thousand and more sequences, there were only few sequences that are completely foreign to human genome. These Insilico works can save much of time and cost in designing a vaccine molecule also the Bio-informatics tools are greatly helpful in screening or designing processes. Through the Insilico designing all possible risks are eliminated and also unwanted sequences are easily removed and our time will be saved in finding a potential vaccine candidate. As our approaches are sequence and structure based the molecular designing works are also made easier with designing software like Argus lab. Still there are lots of diseases in need of drugs and vaccines to be created; hopefully these Insilico works will make the task easier for researchers to design vaccines or drugs in future.

References

- [1] Burgdorfer W, Barbour AG, Hayes SF, Benach JL, Grunwaldt E, Davis JP. **Lyme disease-a tick-borne spirochetosis?** PMID: 7043737
- [2] Claire M. Fraser¹, Sherwood Casjens², Wai Mun Huang², Granger G. Sutton¹, Rebecca Clayton¹, Raju Lathigra³, Owen White¹, Karen A. Ketchum¹, Robert Dodson¹, Erin K. Hickey¹, Michelle Gwinn¹, Brian Dougherty¹, Jean-Francois Tomb¹, Robert D. Fleischmann¹, Delwood Richardson¹, Jeremy Peterson¹, Anthony R. Kerlavage¹, John Quackenbush¹, Steven Salzberg¹, Mark Hanson³, Rene van Vugt², Nanette Palmer², Mark D. Adams¹, Jeannine Gocayne¹, Janice Weidman¹, Teresa Utterback¹, Larry Watthey¹, Lisa McDonald¹, Patricia Artiach¹, Cheryl Bowman¹, Stacey Garland¹, Claire Fujii¹, Matthew D. Cotton¹, Kurt Horst¹, Kevin Roberts¹, Bonnie Hatch¹, Hamilton O. Smith¹ & J. Craig Venter¹. **Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*** *Nature* **390**, 580-586 (11 December 1997) | doi:10.1038/37551; Received 3 November 1997; Accepted 18 November 1997
- [3] Allen C. Steere,¹ Jenifer Coburn,² and Lisa Glickstein¹. **The Emergence of Lyme disease** *J Clin Invest.* 2004 Apr 15; 113(8):10931101. doi:10.1172/JCI200421681
- [4] Steere AC. **Lyme disease: a growing threat to urban populations.** *Proc Natl Acad Sci U S A.* 1994 Mar 29;91(7):2378-83.
- [5] Allen C. Steere, M.D., Robert L. Grodzicki, M.S., Arnold N. Kornblatt, V.M.D., Joseph E. Craft, M.D., Alan G. Barbour, M.D., Willy Burgdorfer, Ph.D., George P. Schmid, M.D., Elizabeth Johnson, M.S., and Stephen E. Malawista, M.D. **The Spirochetal Etiology of Lyme Disease** DOI:10.1056/NEJM198303313081301
- [6] A. G. Barbour **Isolation and cultivation of Lyme disease spirochetes.** *Yale J Biol Med.* 1984 Jul-Aug; 57(4): 521–525. PMCID: PMC2589996.
- [7] Monica E. Embers^{1*} and Sukanya Narasimhan **Vaccination against Lyme disease: past, present, and future** *Front. Cell. Infect. Microbiol.*, 12 February 2013 |http://dx.doi.org/10.3389/fcimb.2013.00006

- [8] Centre for disease control and prevention Lyme Disease <http://www.cdc.gov/>
- [9] Steere AC, Grodzicki RL, Craft JE, Shrestha M, Kornblatt AN, Malawista SE. **Recovery of Lyme disease spirochetes from patients.** Yale J Biol Med. 1984 Jul-Aug;57(4):557-60. PMID: 6393606
- [10] Tilly K, Rosa PA, Stewart PE (2008).. InfectiousDiseaseClinicsofNorthAmerica22(2): 21734,v. **Biology of infection with Borrelia burgdorferi** doi:10.1016/j.idc.2007.12.013
- [11] Kurtenbach K, Hanincová K, Tsao JI, Margos G, Fish D, Ogden NH (2006).. Nature Reviews. Microbiology 4 (9): 660–9. **Fundamental processes in the evolutionary ecology of Lyme borreliosis.** doi:10.1038/nrmicro1475.
- [12] Allen C. Steere, MD; Elise Taylor; Gail L. McHugh, MS; Eric L. Logigian, MD. **The Over diagnosis of Lyme Disease** doi:10.1001/jama.1993.03500140064037
- [13] D. W. Rahn and S. E. Malawista **Lyme Disease** West J Med. 1991 Jun; 154(6): 706–714.PMCID: PMC1002871
- [14] Canadian Lyme Disease foundation <http://canlyme.com/>
- [15] American Lyme disease foundation <http://www.aldf.com/>
- [16] **Lyme Disease** <https://www.nlm.nih.gov/medlineplus/>
- [17] El Ghachi M1, Bouhss A, Blanot D, Mengin-Lecreulx D. **The bacA gene of Escherichia coli encodes an undecaprenyl pyrophosphate phosphatase activity.** (PMID:15138271)