International Journal of Advanced Biotechnology and Research
ISSN 0976-2612, Online ISSN 2278–599X,
Vol 4, Issue 4, 2013, pp 578-585
http://www.bipublication.com

IN-SILICO VACCINE DESIGNING FOR *Mycobacterium tuberculosis* USING REVERSE VACCINOLOGY

Aasma S Jamkhandi*, Mubashshira A Pathan and Anil Kumar S. Katti

Dept. of P.G Studies in Bioinformatics, Walchand Centre for Biotechnology,
Walchand college of Arts and Science, Solapur, Maharashtra, India.
*correspondence:E-mail ID: alicebioinfo@gmail.com

[Received-11/11/2013, Accepted-27/12/2013]

ABSTRACT

Tuberculosis is a major public health problem, causing death of over 1.8 million people every year. Tuberculosis is caused by the pathogenic bacterium *Mycobacterium tuberculosis*. The conventional approach has failed to provide a best solution because of increasing spread of multidrug resistant tuberculosis. In this work westudy *in-silico* peptide vaccine using reverse vaccinology. This method allows predicting the antigenic determinants from the bacterial proteins which shows the least identity with that of human proteins. The antigenicity prediction was carried out using VaxiJen v2; among 892 proteins 585 proteins were predicted as antigenic. These antigenic proteins were subjected to SDSC biological workbench to find least identity with human proteins using FASTA program. The GlmU protein (NCBI ID NP_335483.1) was found to have least identity of 20.75%. The antigenic determinants were predicted using Emboss antigenic. The selected peptide was subjected to MAPPP for identifying the MHC molecule to which it binds. The Hex server was use for docking studies.

Key words: -*Mycobacterium tuberculosis*, GlmU protein, reverse vaccinology, peptide vaccine, docking

[1] INTRODUCTION

Mycobacterium tuberculosis (MTB) is a pathogenic bacterium in the genus *Mycobacterium* and is a causative agent of most cases of tuberculosis (TB) (Michael Glickman and William Jacob, 2001). The World Health Organization (WHO) reports 9.2 million new cases of TB in 2006 of whom 7.7% were HIV-infected. Tuberculosis is the most common opportunistic infection in HIV-infected patients as well as the leading cause of death. Further, there has been an increase in rates of drug resistant tuberculosis, including multi-drug (MDRTB) and extensively drug resistant TB (XDRTB), which are difficult to treat and contribute to increased mortality (Swaminathan and Narendran, 2008).The conventional approach to vaccine development is based on dissection of the pathogen using biochemical, immunological and microbiological methods. Although successful in several cases, this approach has failed to provide a solution to prevent several major bacterial infections. The availability of

complete genome sequences in combination with novel advanced technologies, such as bioinformatics, microarrays and proteomics, have revolutionized the approach to vaccine development and provided a new impulse to microbial research. The genomic revolution allows the design of vaccines starting from the prediction of all antigens in-silico, independently of their abundance and without the need to grow the pathogen in vitro. This new genome-based approach, which is named "Reverse Vaccinology", has been successfully applied for Neisseria meningitides serogroup B for which conventional strategies has failed to provide an efficacious vaccine. The concept of "Reverse Vaccinology" can be easily applied to all the pathogens for which vaccines are not yet available and can be extended to parasites and viruses (Barbara Capecchi et al., 2004).

It is possible to predict the transmembrane proteins as well as the secretory proteins from the genome of organism using different tools like SOSUIDB server(Masahiro Gomi *et al.,* 2005).SOSUIDB does the predictive analysis of secretory proteins and membrane proteins coded by their respective genomes(Ryusuke Sawada *et al.,* 2011). The server VaxiJen allows categorization of antigenic proteins based on the physicochemical properties Irini Doytchinova and Darren Flower (2007), these antigenic proteins can be subjected to FASTA program to get the least identity protein(Lipmann and Pearson, 1999). The antigenic peptides can be predicted using method of Kolaskar and Tongaonkar.MAPPP is a bioinformatics tool for the prediction of potential antigenic epitopes presented on the cell surface by major histocompatibility complex class I (MHC I) molecules to CD8 positive T lymphocytes. It combines existing predictions for proteasomal cleavage with peptide anchoring to MHC I molecules (Hakenberg *et al.,* 2003). The epitope and MHC molecules can be docked using HEX 6.3 software and binding energy can be calculated.

[II] METHODS

2.1 Retrieval of proteome set of M*ycobacterium tuberculosis*

The proteome set of *Mycobacterium tuberculosis* CDC1551 (Oshkosh) was retrieved in fasta format from EBI Integr8 database.Integr8 (http://www.ebi. ac.uk/integr8/) has been developed to provide an integration layer for the exploitation of genomic and proteomic data (Manuela Pruess et al., 2004).

2.2 Prediction of transmembrane proteins

To predict the transmembrane proteins SOSUI server was used. The 4189 proteins from *Mycobacterium tuberculosis* were subjected to SOSUI server. This server has membrane protein prediction system SOSUI and signal peptide prediction system SOSUIsignal (Masahiro Gomi et al., 2005). By combination of those systems, number of transmembrane helices can be predicted based on physicochemical parameters. Therefore, it is possible to investigate the number distribution of transmembrane regions in membrane proteins comprehensively by using SOSUI server (Ryusuke Sawada et al., 2011).

2.3Retrieval of secretory proteins

The secretory proteins of *Mycobacterium tuberculosis* CDC1551 (Oshkosh) were retrieved from SOSUIDB server (http://bp.nuap.nagoyau .ac.jp/ sosui/sosui signal/SOSUIsignalDB/). Secretory proteins function as agents for numerous cell-cell interactions and determine the survival strategies adopted by organisms. Using the SOSUI system for membrane proteins and SOSUI signal for signal peptides, SOSUIDB does the predictive analysis of secretory proteins coded by their respective genomes (Masahiro Gomi et al., 2005).

2.4Prediction of Antigenicity

The server VaxiJen (http://www.jenner.ac.uk/VaxiJen) was used to

predict the antigenicity of the query proteins. Total 892 secretory and transmembrane proteins were predicted for antigenicity. VaxiJen was developed to allow antigen classification based on the physicochemical properties of proteins according to the method of Irini Doytchinova and Darren Flower (2007). Protein sequences can be submitted as single protein or as a multiple sequence file in fasta format.

2.5 Screening of least identity proteins

The 585 *Mycobacterium* antigenic proteins having least identity with that of human proteins were identified by using FASTA program provided by SDSC biological workbench (http://workbench.sdsc.edu). The *Mycobacterium* antigenic proteins were searched against Swiss-Prot Human database. BLOSUM50 scoring matrix was used for protein sequences to find the identity and other parameters were used as default values (Lipmann and Pearson, 1999).

2.6 Epitope prediction

Antigenic peptides were predicted using Predicting Antigenic Peptides and EMBOSS antigenic were used. Predicting Antigenic Peptides tool is provided by the Immunomedicine Group (http: //imed.med.ucm.es/Tools/antigenic.pl). These tools determine antigenic peptides using the method of Kolaskar and Tongaonkar (Kolaskar and Tongaonkar, 1990).

2.7 MHC-I binding prediction

MHC class I antigenic peptide processing prediction was done using MAPPP server (http://www.mpiib-berlin.mpg.de/MAPPP/expertquery.html).

MAPPP is a bioinformatics tool for the prediction of potential antigenic epitopes presented on the cell surface by major histocompatibility complex class I (MHC I) molecules to CD8 positive T lymphocytes. It combines existing predictions for proteasomal cleavage with peptide anchoring to MHC I molecules (Hakenberg et al., 2003).

2.8 Peptide designing

The peptide molecule was designed using software Accelrys Discovery Studio. Discovery Studio is built on the SciTegic Pipeline PilotScitegic Enterprise Server platform, open Operating platform, allowing faultless integration of protein modeling, pharmacophore analysis, structure based design, and peptide designing as well as visualization application(Discovery Studio 2.5 Guide Accelrys Inc 2009). The energy minimization of peptide molecule was done using Argus Lab (Mark A. Thompson).

2.9 Molecular Docking

Molecular docking of LADLIATHRAVSAAVTVLTT peptide and MHC-I molecule (PDB ID: 3UTQ) was carried out using Hex6.3. These docking algorithms employ a range of efficient search and energy-based scoring strategies, including fast Fourier transform (FFT) correlations, geometric hashing, and Monte Carlo (MC) techniques. Protein docking aims to predict how proteins interact based on analyses of known protein structures (Ritchie 2008). The complex obtained after docking was visualized under RasMol 2.6. It is a graphic molecular visualization tool the program is aimed at display, teaching and generation of publication quality images (Roger Sayle, 1996).

[III] RESULTS

3.1 Retrieval of proteome set of *Mycobacterium tuberculosis*

The *Mycobacterium tuberculosis* strain CDC1551 proteome set was retrieved in FASTA from the EBI Integr8. *M. tuberculosis* proteome contains 4189 proteins and further screened for transmembrane proteins and secretory proteins.

3.2 Prediction of transmembrane proteins

The proteome set containing 4189 proteins were screened to predict transmembrane proteins using SOSUI Batch server. Among 4189 proteins, total 495 proteins were predicted to be transmembrane proteins based on average hydrophobicity of a

polypeptide segment, which is considered to be the most important factor in the formation of transmembrane helices.

3.3 Retrieval of secretory proteins

The set of secretory proteins were obtained from SOSUIDB secretory database. *M. tuberculosi*s strain CDC1551 proteins contain 4189 open reading frame and 737 secretory proteins with their NCBI accession numbers. The SOSUIDB database uses genome parameters like G+C content and number of open reading frames to predict the secretory proteins.

3.4 Prediction of Antigenicity

The *Mycobacterium tuberculosis* transmembrane proteins and secretory proteins were used to predict the antigenicity using VaxiJen v2.0 as shown in Fig. 1. It is an alignment-independent prediction method which uses the physicochemical properties of proteins to predict the antigenicity. Among 892 proteins 585 proteins were predicted to be antigenic. These antigenic proteins were further analysed for least identity with human proteins. According to Prabhavathy et al the VaxiJen antigenicity score obtained for OMPLA and LsrC was 0.5326 and 0.4732 showing that the proteins were antigenic (Prabhavathy et al., 2011).

3.5 Screening of least identity proteins

The antigenic proteins were subjected to SDSC (San Digo Supercomputer center) biological workbench to identify the least identity with that of human proteins. Total 585 proteins were screened in which the secretory protein having NCBI accession number NP_335483.1 (GlmU) was found to have least identity of 20.755% as shown in Fig. 2. Similarly Ramani et al. (2012) found two proteins with Gene ID IF2_VIBC3 and SECA_VIBC3 of Vibrio cholera having least identity of 20.62% and 21.89% with human proteins respectively (Ramani et al., 2012).

3.6 Epitope prediction

The NP_335483.1 GlmU protein sequence having least identity and least E value was used for finding antigenic determinants using Emboss Antigenic tool and compared with Immunomed Predicting Antigenic Peptides. There were 28 antigenic determinants as shown in Table. 1 and graphical representation as shown in Fig. 3. Then the LCV values of the determinants were calculated. Based on the greater LCV value the LADLIATHRAVSAAVTVLTT antigenic determinant was selected as a best vaccine candidate and used for further study as shown in Fig. 4. In similar with Ramani and Sriranjini antigenic determinant NPLPCLIIIL of protein DAPE_RICRO found to have the grater LCV value of 40% (Ramani and Sriranjini, 2012).

3.7 MHC-I binding prediction

The MAPPP server is primarily used to identify the type of MHC molecule to which epitope molecule binds. MAPPP predicted the MHC binding probability and MHC cleavage score as shown in Fig.5. The type of MHC molecule predicted was HLA_A_0201, having pdb 3UTQ which was used for docking with the epitope molecule. In accordance with Pallavi and Chandan also got the H2_Kd MHC I molecule for RISSLVLAGIIGLSSTVAVKA peptide of antigenic protein (Pallavi and Chandan, 2012).

3.8 Peptide designing

Discovery studio 3.1 was used to design the peptide molecule. The peptide molecule LADLIATHRAVSAAVTVLTT was designed as shown in Fig. 6 and saved in .pdb format. Minimization of peptide molecule was done using Argus Lab. Energy minimization of molecule was found to be -808.017582. This molecule was further used in docking with MHC class I molecule. Similarly according to Dharmaiah et al three peptides were designed using Discovery studio (Dharmaiah et al., 2012).

3.9 Molecular Docking

Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Hex 6.3 software is used to dock the epitope molecule LADLIATHRAVSAAVTVLTT and the MHC molecule (3UTQ) as shown in Fig.7.

The binding energy was -461.86. According to Satya et al. (2011) human Brain-derived neurotrophic factor (BNDF) protein was succefully docked with ligands Astaxanthin and Beta carotene and binding energy was found to be -225.39 and 220.68 respectively (Satya et al., 2011).

[IV] CONCLUSION

Tuberculosis is life threatening infectious disease caused by *Mycobacterium tuberculosis.* It is found that it infects Human and some of the animals as well. Efforts during the 21st century vaccinology will witness more successes of application of vaccine informatics in vaccine research. This approach has potential to make dramatic advances and to improve human health. We retrieved the complete proteome of *M. tuberculosis* and screened by using SDSC Workbench. We got the sequence having the least identity as 20.75%. Antigenic determinant LADLIATHRAVSAAVTVLTT was selected based on least identity. This epitope was docked with MHC-1 molecule, the docking energy was found to be -461.86. Hence we conclude that this is the best Vaccine Candidate. The resultant vaccine can be sent to clinical trials and used for further research.

We suggest that *in silico* analyses be matched with *in vitro* assays (binding studies, MHC-tetramers and ELISpot assays). Furthermore, these discoveries should be validated in vivo, using HLA transgenic mice.

[V] REFERENCES

1. Barbara C, Davide S, Jeannette, Rino R, Mariagrazia P: The Genome Revolution in Vaccine Research. Curr. Issues *Mol. Biol*. 2004, 6:17-28.
2. David Ritchie: Recent Progress and Future Directions in Protein-Protein Docking. *Current Protein and Peptide Science* 2008 (9): 1-15.
3. Dharmaiah D, Rao A and Sridhar G: Modeling the Interactions between the BDNF Peptides and TrkB Receptor. *International Journal of Computer Applications* 2012 (55)9.
4. Discovery Studio 2.5 Guide Accelrys Inc., San Diego 2009.
5. Hakenberg J, Nussbaum A, Schild H, Rammensee, Kuttler H, Holzhütter H, Kloetzel P, Kaufmann S and Mollenkopf H: MAPPP - MHC-I Antigenic Peptide Processing Prediction. *Applied Bioinformatics* 2003 (2)3: 155-158.
6. Kolaskar A and Tongaonkar P: A semi-empirical method for prediction of antigenic determinants on protein antigens. *Federation of European Biochemical Societies* 1990 276(1.2): 172-174.
7. Manuela P, Paul K and Rolf A: Integrating Genomic and Proteomic Data the Integr8 Project. *Journal of Integrative Bioinformatics* 2004.
8. Masahiro G, Ryusuke S, Masashi S and Shigek M: Comparative Proteomics of the Prokaryota Using Secretory Proteins. *Chem-Bio Informatics Journal* 2005 5(3): 56-64.
9. Michael SG, William R J: Microbial pathogenesis of *Mycobacterium tuberculosis* Dawn of a discipline. *Cell* 2001, 104:477-485.
10. Pallavi K, Chandan D: In-silico design and development of vaccine by reverse vaccinology approach for anthrax. *Journal of Advanced Bioinformatics Applications and Research* 2012 (3)1: 262-266.
11. *Pearson W:* Flexible similarity searching with the FASTA3 program package. *Bioinformatics Methods and Protocols 1999 185-219.*
12. Prabhavathy k, Perumal P and Baalaji N: In sillico identification of B cell and T cell epitope on OMPLA and LsrC for *Salmonella typhi* for peptide based vaccine designing. *Indian Journal Of Biotechnology* 2011 (10): 440-451.
13. Ramani I and Sriranjini S:*In-silico* designing and development of vaccine for *rickettsia rickettssi* in Rocky Mountain spotted fever.
14. *Journal of Advanced Bioinformatics Applications and Research* 2012 (3): 339-344.
15. Ramani I, Sriranjini S and Jasmine: *In-silico* designing and development of vaccine for *v.cholerae* 0139 incholera disease. *International Journal of Scientific & Engineering Research* 2012 (3)7.
16. RasMol 2.6 Roger Sayle, ftp://ftp.dcs.ed.ac.uk/pub/rasmol 1996.

17. Ryusuke S and Shigeki M: Number Distribution of Transmembrane Helices in Prokaryote Genomes. *Computational Biology and Applied Bioinformatics* 2011.
18. Satya G, Rao A, Sridhar G and Chakravarthy M: Protein ligand interaction analysis an insilico potential drug target identification in diabetes mellitus and nephropathy. *Journal of Bioinformatics and Sequence Analysis* 2011 2(5): 95-99.
19. Swaminathan S, Narendran G: HIV and tuberculosis in India. *J. Biosci*. 2008, 33:527–537

List Of figures

Fig. 1. Antigenicity prediction of GlmU protein by VaxiJen

Alignments for Best Scores:

Back to table

>>SWISSPROT_HUMAN:ISPD_HUMAN (A4D126) Isoprenoid synthas (451 aa)

```
 initn:  60 init1:  60 opt: 111
Smith-Waterman score: 111;  20.755% identity (23.077% ungapped) in 159 aa overlap (22-169:44-197)

                      10        20        30        40        50
gi_158          MVQIVQVRVLATELQGGLMTFPGDTAVLVLAAGPGTRMRSDTPKVLHTLAG
                                  :  .:... :.: : ::   ::: .  .
SWISSP PGPCLSGQRGADHTASASLQSVAGTEPGRHPQAVAAVLPAGGCGERMGVPTPKQFCPILE
            20        30        40        50        60        70

               60        70        80        90       100
gi_158 RSMLSHVLHAIAKLAPQRLIVV--LGHDHQRIAPLVGELADTLGRTIDVALQDRPLGTGH
       : ..:..:.:. ..   . :::   :.. . .  .. .          ..: .  .   .
SWISSP RPLISYTLQALERVCWIKDIVVAVTGENMEVMKSIIQKYQHK-----RISLVEAGVTRHR
            80        90       100       110             120

     110       120               130       140          150       160
gi_158 AVLCGLSALPDDYAGN-------VVVTSGDTPLLDADTLADLI--ATHRAVSAAVTVLTT
       ... ::.:: .:  ..       :.. ..  :...  .:  ..  : ......:.  :..
SWISSP SIFNGLKALAEDQINSKLSKPEVVIIHDAVRPFVEEGVLLKVVTAAKEHGAAGAIRPLVS
      130       140       150       160       170       180
```

Fig.2. Screening of least identity proteins using SDSC

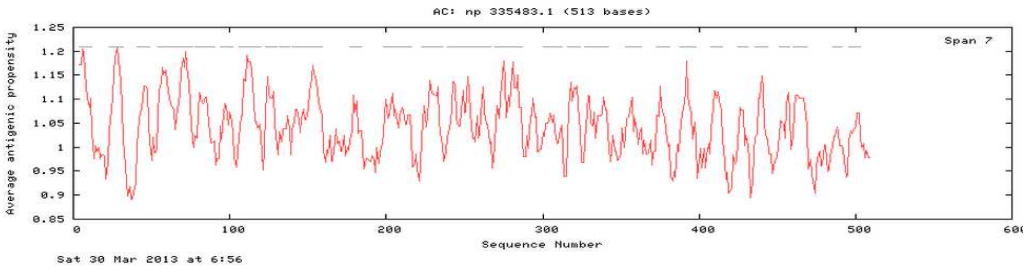

Fig. 3 Graphical representation of antigenic determinants

Score 1.171 length 20 at residues 141->160
Sequence: LADLIATHRAVSAAVTVLTT
| |
141 160
Max_score_pos: 154
LCV score: 35%

Fig. 4. Peptide with greater LCV value.

Query results								
Protein position	**Length**	**Sequence**						
Epitope	**Position**	**MHC type**	**n-mer**	**Overall score**	**Cleavage Probability**	**MHC binding score**	**Group**	
0..512	513	MVQIVQVRVLATELQGGLMT..RASEMACQQPTQPPDADQTP						
APQRLIVVL	65	H2_Ld	9	0.8754	0.9444	0.8065	same length	
APQRLIVVL	65	HLA_B_0702	9	0.8571	1.0000	0.7143	c-term. trimmed	
APQRLIVVL	65	H2_Ld	9	0.9032	1.0000	0.8065	c-term. trimmed	
RIAPLVGEL	79	HLA_A_0201	9	0.9028	1.0000	0.8056	n-term. trimmed	
RPLGTGHAVL	102	HLA_B_0702	10	0.8906	1.0000	0.7812	same length	
RPLGTGHAVL	102	HLA_B_0702	10	0.8906	1.0000	0.7812	c-term. trimmed	
LLDADTLADL	134	HLA_A_0201	10	0.9118	1.0000	0.8235	n-term. trimmed	
ALRSALSRL	208	HLA_A_0201	9	0.8611	1.0000	0.7222	n-term. trimmed	
ALRSALSRL	208	HLA_A_0201	9	0.8531	0.9839	0.7222	same length	
YLTDVIAIL	226	HLA_A_0201	9	0.8889	1.0000	0.7778	n-term. trimmed	
LYLTDVIAIL	225	H2_Kd	10	0.9107	1.0000	0.8214	same length	
YLTDVIAIL	226	HLA_A_0201	9	0.8889	1.0000	0.7778	trimmed twice	
LYLTDVIAIL	225	H2_Kd	10	0.9107	1.0000	0.8214	c-term. trimmed	
AGVNNRVQL	253	H2_Db	9	0.8857	1.0000	0.7714	c-term. trimmed	
QLAELASEL	260	HLA_A_0201	9	0.8720	0.9940	0.7500	n-term. trimmed	

Fig. 5. MAPPP result for MHC binding probability and MHC cleavage score

Fig. 6. Peptide designing using discovery studio

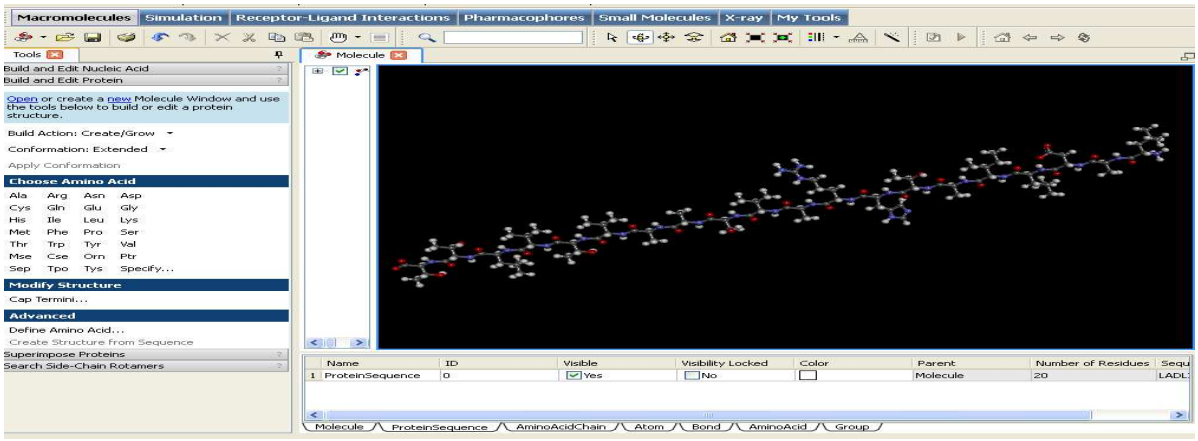

Fig. 7. Docking of peptide molecule with MHC-I molecule using Hex 6.3

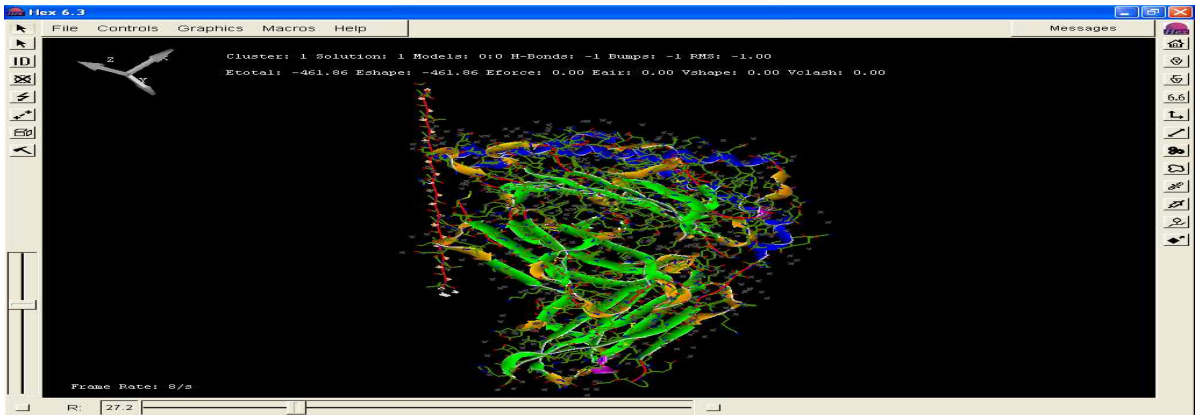

Table.1 Antigenic determinants of GlmU sequence.

There are 28 antigenic determinants in your sequence:

n	Start Position	Sequence	End Position
1	4	IVQVRVLAT	12
2	23	GDTAVLVLAA	32
3	41	DTPKVLHTL	49
4	54	MLSHVLHAIAKLAPQRLIVVLGH	76
5	78	HQRIAPLVGELAD	90
6	94	RTIDVALQD	102
7	106	GTGHAVLCGLSALPD	120
8	123	AGNVVVTS	130
9	132	DTPLLDA	138
10	140	TLADLIATHRAVSAAVTVLT	159
11	177	HEVMAIVE	184
12	198	VNAGVYAFDIAALRSALSR	216
13	223	QQELYLTDVIAILR	236
14	239	GQTVHASHVDDSALVAGVNNRVQLAELAS	267
15	269	LNRRVVAAHQLAGVTVVDP	287
16	300	RDTVIHPGTQLLG	312
17	316	IGGRCVVGP	324
18	327	TLTDVAVGDGASVVRT	342
19	353	AAVGPFTYLRP	363
20	373	LGAFVEVKN	381
21	388	TKVPHLTYVGD	398
22	408	GASSVFVN	415
23	425	TVGSHVR	431
24	436	TMFVAPVTI	444
25	451	GAGTVVRE	458
26	460	VPPGALAVSA	469
27	485	GSPAAQA	491
28	496	SEMACQQP	503