Journal of Advanced Bioinformatics Applications and Research.
Vol 1, Issue 1, June 2010, pp 50-58

IN SILICO PREDICTION OF INTERSPECIES RELATIONSHIP AND PROPERTY TO BE A VACCINE CANDIDATE OF SEGMENT 1 PROTEIN OF INFLUENZA A VIRUS FROM INDIA

Vishnu J. Gaikwad, Shraddha. S. Bhosale, Sandip S. Patel.
Dept. of Biotechnology Engineering,
Tatyasaheb Kore Institute of Engineering and Technology, Warananagar,
Dist- Kolhapur, MS, INDIA-416113.

ABSTRACT:

The ancestral information related to h1n1 swine flu segment 1 from India has the similarity with the h1n1 segment 1 from Chile.
The ambition of this research work includes analyzing the gene sequence of segment 1 from H1N1 (INDIA), to find out interspecific relationship between them by using advanced tools and techniques, for prediction of vaccine candidate we have predicted hydrophilic, property to be a vaccine candidate for future invention.
The nucleic acid sequence of H1N1 segment 1 has 2282 base-pairs. By using Bio-Edit we have predicted the sequence similarity interspecifically in between 20 species for evaluating the relationship between them, also calculated the entropy of the segment1.
These predictions are very useful, from which in future we can design a drug or vaccine against it.

Keywords: h1n1 swine flu segment 1, interspecies relationship, vaccine candidate.

Introduction

Swine influenza virus (**SIV**) is any strain of the influenza family of viruses that is endemic in pigs [1]. As of 2009, the known SIV strains include influenza C and the subtypes of influenza A known as H1N1, H1N2, H3N1, H3N2, and H2N3.
Influenza viruses belong to the family of Orthomyxoviridae; viruses with segmented RNA genomes that are negative sense and single-stranded [2].
Influenza remains an important cause of morbidity and death in Man, principally because of the ability of the surface haemagglutinin to undergo extensive antigenic variation [3].

A comparative sequence analysis of series of antigenically drifted strains of the H3 haemagglutinin subtype [4] clearly shows that variation occurs by the gradual accumulation of point mutations causing single amino acid substitutions. It appears that these cluster in 4 antigenic sites on the surface of the globular head of the haemagglutinin [5, 6] but amino acid substitutions are by no means confined solely to these relatively discrete sites.

Influenza A

Swine influenza is known to be caused by influenza A subtypes H1N1 [7], H1N2 [7], H2N3 [8], H3N1 [9], and H3N2 [7]. In pigs, three influenza A virus subtypes (H1N1, H1N2, and H3N2) are the most common strains worldwide [10]. In the United States, the H1N1 subtype was exclusively prevalent among swine populations before 1998; however, since late August 1998, H3N2 subtypes have been isolated from pigs. As of 2004, H3N2 virus isolates in US swine and turkey stocks were triple reassortants, containing

genes from human (HA, NA, and PB1), swine (NS, NP, and M), and avian (PB2 and PA) lineages [11].

Transmission

Transmission between pigs

Influenza is quite common in pigs, with about half of breeding pigs having been exposed to the virus in the US [12]. Antibodies to the virus are also common in pigs in other countries [12].

Transmission to humans

People who work with poultry and swine, especially people with intense exposures, are at increased risk of zoonotic infection with influenza virus endemic in these animals, and constitute a population of human hosts in which zoonosis and reassortment can co-occur [13]. Vaccination of these workers against influenza and surveillance for new influenza strains among this population may therefore be an important public health measure [14]. Transmission of influenza from swine to humans who work with swine was documented in a small surveillance study performed in 2004 at the University of Iowa [15]. This study among others forms the basis of a recommendation that people whose jobs involve handling poultry and swine be the focus of increased public health surveillance [13]. Other professions at particular risk of infection are veterinarians and meat processing workers, although the risk of infection for both of these groups is lower than that of farm workers [16].

Treatment

In humans

The important relationship between influenza in pig and human populations is due to the relative ease of transmission of viruses between the two species and is rejected in the prevalent subtypes. Influenza A viruses of the H1N1 and H3N2 subtypes have co-circulated and caused outbreaks of disease among pigs in Europe since the mid 1970s. The initial H3N2 viruses were antigenically related to contemporary human H3N2 viruses, such as Port Chalmers1}73 [17].

The ambition of this project includes analyzing the gene sequence of segment 1 from H1N1 (INDIA), to find out interspecific relationship between them by using advanced tools and techniques, for prediction of vaccine candidate we have predicted hydrophilic, antigenic peptides and plots.

Materials and Methods:

Data Collection and Analysis:

The sequence of Influenza A virus-segment 1 originated from India in 2009 (A/India/6263/1980(H1N1), gene index number is 133754167 by which we accessed it from National Centre for Biotechnology Information. This gene information we had applied for various software through which derives desired results.

We had translated this sequence from nucleic acid to amino acid and this sequence applied for modeling and antigenic fragments prediction as vaccine and or drug candidate.

Sequence analysis:

The programme used in sequence analysis was BioEdit VERSION 7.0.0, Tom Hall, Ibis Therapeutics, a division of Isis Pharmaceuticals, Inc. BioEdit is a biological sequence editor that runs in Windows 95/98/NT/2000 and is intended to provide basic functions for protein and nucleic sequence editing, alignment, manipulation and analysis.

We have also predicted Entropy (Hx) plot from the BioEdit programme.

Prediction of Protein Secondary Structure

The present version, GOR IV, uses all possible pair frequencies within a window of 17 amino acid residues and is reported by J. Garnier. J.F. Gibrat and B.Robson [18]. After cross validation on a data base of 267 proteins, the version IV of GOR has a mean accuracy of 64.4% for a three state prediction (Q3). The program gives two outputs, one eye-friendly giving the sequence and the predicted secondary structure in rows, H=helix, E=extended or beta strand and C=coil ;the second output gives the probability values for each secondary structure at each amino acid position. The predicted secondary structure is the one of highest probability compatible with a predicted helix segment of at least four residues and a predicted extended segment of at least two residues. This program was written by Jean-Francois Gibrat and was modified by Stephen Pheiffer in order to increase it's efficiency.
We have got the secondary structure information about alpha-helices, turns and coils by using GOR-IV.

Phylogenetic Analyssis:

Evolutionary relationship predictions are the most commonly used to derive ancestral relations amongst the species. We have predicted the relationship among 19 species with our H1N1 segment 1 from India.

Prediction of hydrophobicity:

Hydrophobic interactions are the most important non-covalent forces that are responsible for different phenomena such as structure stabilization of proteins binding of enzymes to substrates and folding of proteins [19]. This kind of interaction appears when non-polar compounds are put into water, and it is an entropy-driven process. The protein separation technique based on hydrophobic interactions, *Hydrophobic Interaction Chromatography* (HIC) is an important technique which exploits the reversible interaction between the hydrophobic patches on a protein's surface and the hydrophobic ligands of a chromatographic medium at moderately high concentrations of an antichaotropic salt.

These predicted surface sites or, in other words, the hydrophilic, accessible, or mobile regions were then correlated to the known antigenic sites from immunological studies and accessible sites determined by X-ray crystallographic data for several proteins [20]. A semi-empirical method which makes use of physicochemical properties of amino acid residues and their frequencies of occurrence in experimentally known segmental epitopes was developed to predict antigenic determinants on proteins [21]

Prediction of Antigenic Fragments-Peptides:

Identification of epitopes on proteins would be useful for diagnostic purposes and also in the development of peptide vaccines [22]. To aid experimental workers, Hopp and Woods have developed a method for prediction of antigenic determinants [23].
The approach of Hopp and Woods has been modified to take into account the fact that antigenic sites are on the surface of the protein and most surface residues are antigenic.

Results and Discussion:

The nucleic acid sequence of H1N1 segment 1 has 2282 base-pairs. By using Bio-Edit we have predicted the sequence similarity interspecifically in between 20 species for evaluating the relationship between them, also calculated the entropy of the segment1. (Fig.1)

Alignment of 20 species by using BioEdit.

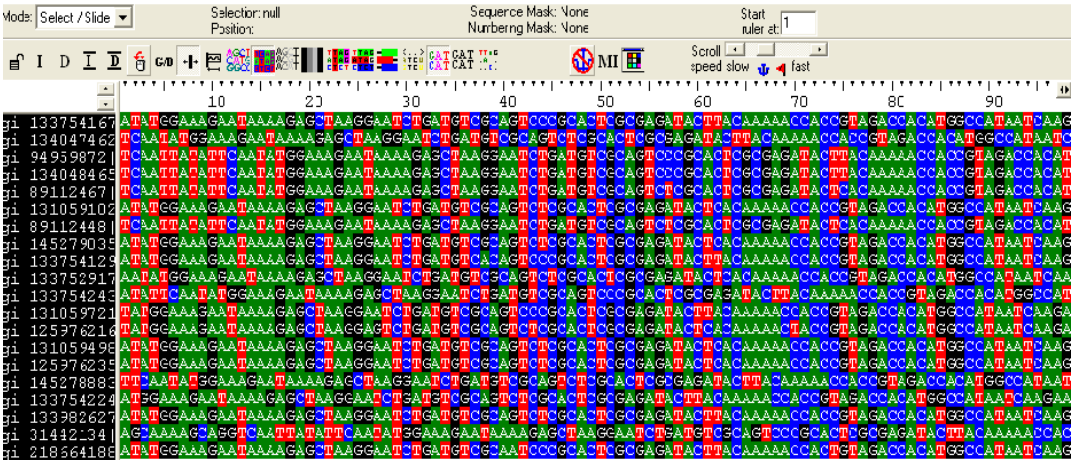


Fig.1: The figure shows the interspecies sequence alignment of 19 species with H1N1 segment 1 from INDIA.

Phylogenetic Analysis:

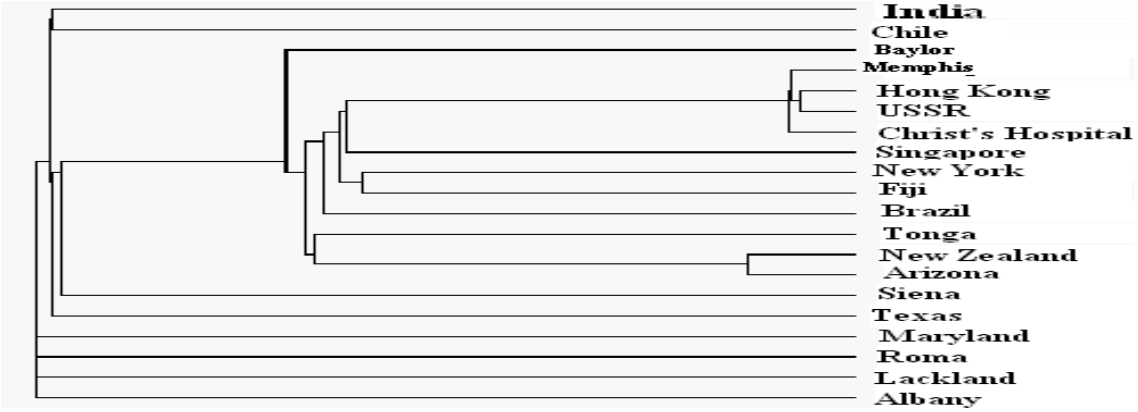


Fig.4: Phylogenetic Analysis Indian H1N1 segment 1 with 19 species around the world.

Above phylogenetic results shows that, the segment 1 from H1N1 India having homology in common ancestor with H1N1 of Chile rather than other continentals.

We had predicted some hypothesis that, the drug molecules and or vaccine candidates present to prevent the H1N1 virus in Chile should have same results about H1N1 virus in India

Hydrophobicity

Since the main feature of most globular protein domains is the hydrophobic core, the distributions of hydrophobic amino acids along the sequences of remote homologues and analogues ought to be similar. But is this pattern recognition already achieved by

standard alignment techniques using log-odds matrices derived from observed substitutions in sequence alignments[24]Most of these matrices cluster polar and nonpolar amino acids separately.

Analyses of conservation in multiple sequence alignments [25, 26] have also identified the amino acids which commonly substitute with each other and rationalise them in terms of structure and/or amino acid properties. Han and Bakerhan:sequence found that the most common clusters of single column amino acid profiles were predominantly hydrophobic or polar in nature. Consecutive segments of up to 13 residues were also often best explained in terms of hydrophobicity. Fiser *et al.*fiser:conservation looked at the conservation of amino acids with respect to structure and concluded that hydrophobicity was well conserved in the core. Ladunga and Smithladunga:substitution employed a robust all-or-nothing binary profile analysis on a number of large sequence databases and also reported common substitutions between hydrophobic or polar residues.

Antigenic Fragments Prediction of H1N1: Our sequence is 759 residues long

Average antigenic propensity for this protein is 1.0151

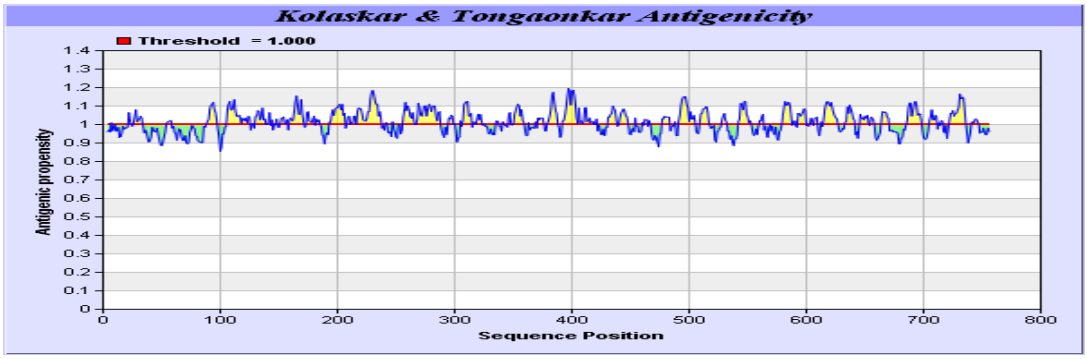
Antigenic plot for sequence

Sr. No.	Start Position	Sequence	End Position
1.	21	VDHMAIIK	32
2.	88	RVMISPLAV	96
3.	105	VTSTVHYPKIYKTYFEKVER	124
4.	129	TFGPVHFRNQVKI	141
5.	158	EAQDVIMEVVFPN	170
6.	173	GARILTS	179
7.	192	ELQNCKISPLMVAY	205
8.	207	LERELVRK	214
9.	216	RFLPVAGGTSSVYIEVLHLTQ	236
10.	255	VDQSLIIAAR	264
11.	267	RFLPVAGGTSSVYIEVLHLTQ	286
12.	306	QAVDICKA	313
13.	349	LQTLKLKVHE	358
14.	378	TRRLIQLIVS	387
15.	391	EQSIVEAIVVAMVFSQEDCMIKAVRGDLNF	420
16.	430	PMHQLLRH	437
17.	439	QKDAKVLFQ	447
18.	476	RGVRVSK	482
19.	490	NAERVVVSIDRFLRV	504
20.	508	RGNVLLSPE	516
21.	541	RGNVLLSPE	552
22.	579	PFQSLVPK	586
23.	592	YSGFVRT	598
24.	614	QIIKLLPFAAAPP	626
25.	632	QFSSLTVNV	640
26.	645	MRILVRGN	652
27.	663	KRLTVLG	669
28.	685	GVESAVLRGFLIL	697
29.	704	YGPALSI	710
30.	721	KANVLIGQGDVVLVM	735
31.	740	DSSILTD	746

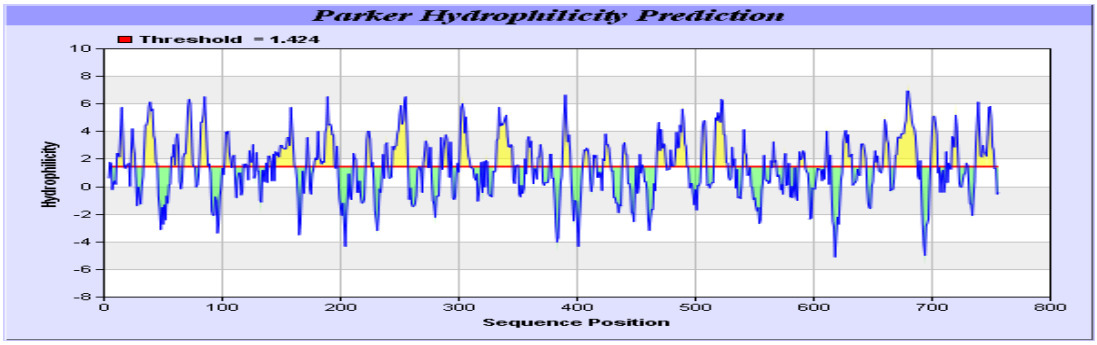
ALLELE: DRB1_0101: Threshold for 3 % with score: 0.14, Highest Score achievable by any peptide: 6

Rank	Sequence	At Position	Score	% of Highest Score
1.	FQSLVPKAI	535	2.6000	43.33
2.	WNRNGPVTS	54	1.7400	29.00
3.	WMMAMKYPI	4	1.7000	28.33
4.	LRISSSFSF	272	1.3000	21.67
5.	WEINGPESV	492	1.1400	19.00
6.	IIKLLPFAA	570	1.1000	18.33
7.	VRGNSPVFN	604	0.7400	12.33
8.	LEMCHSTQI	236	0.6900	11.50
9.	FKRTSGSSV	285	0.6000	10.00
10.	MRILVRGNS	600	0.4000	6.67
11.	IVRRAAVSA	221	0.4000	6.67
12.	FRNQVKIRR	90	0.4000	6.67
13.	VRDQRGNVL	459	0.3000	5.00
14.	VMGMIGILP	412	0.3000	5.00
15.	VKIQWSQNP	515	0.2000	3.33
16.	LIIAARNIV	214	0.2000	3.33
17.	LRKATRRLI	329	0.2000	3.17
18.	LVRGNSPVF	603	0.1700	2.83
19.	MRGVRVSKM	430	0.1300	2.17
20.	IRGQYSGFV	543	0.1000	1.67
21.	IRNWETVKI	509	0.1000	1.67
22.	FVNRANQRL	375	0.1000	1.67

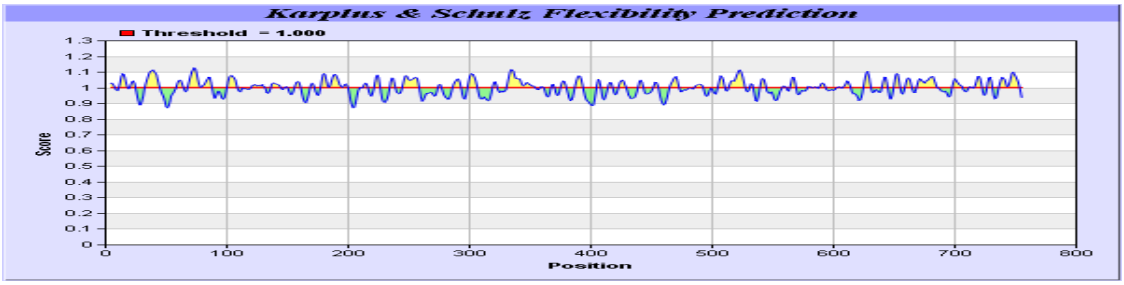
Kolaskar & Tongaonkar Antigenicity: Average: **1.016** Minimum: **0.856** Maximum: **1.195** Threshold: **1.000**



Parker Hydrophilicity Prediction: Average: **1.424** Minimum: **-5.114** Maximum: **6.943** Threshold:1.424



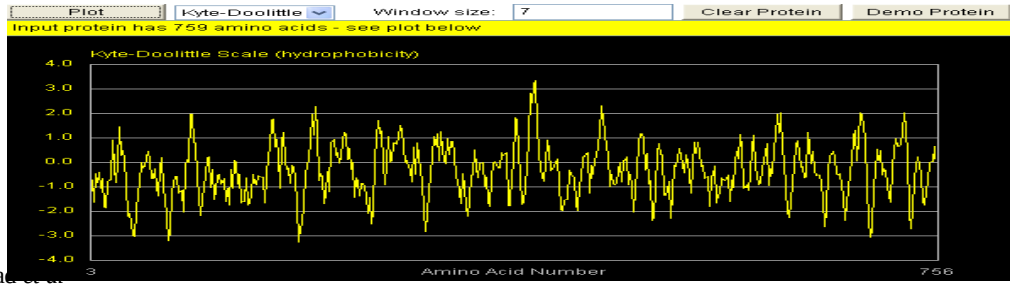
Karplus & Schulz Flexibility Prediction: Average: **1.002** Minimum: **0.872** Maximum: **1.126** Threshold: **1.000**



Bepipred linear epitope prediction: **Average:**-0.024 **Minimum:**-2.007 **Maximum:**2.367 **Threshold: 0.350**

Hydrophobicity plot:

1. Kyte-Doolittle Method



Conclusion and future scope:

We have predicted that the sequence of h1n1 of swine virus is similar to h1n1 segment 1 from Chile.

With prediction of antigenicity we got the sequences that have the antigenic sites by using that we can predict the vaccine and or drug formulations against that.

References:

1. Eccles, R (2005). "Understanding the symptoms of the common cold and influenza". *Lancet Infect Dis* **5** (11): 718–25. doi:10.1016/S1473-3099(05)70270-X. PMID 16253889
2. Baltimore, D. 1971. Expression of animal virus genome. Bacteriol. Rev. 35:235-241.
3. Stuart-Harris, C. (1979) Brit.Med.Bull., 35, 3-8.
4. Both, G.W., Sleigh, M.J., Cox, N.J. and Kendal, A.P. (1983) in press.
5. Wiley, D.C., Wilson, I.A. and Skehel, J.J. (1981) Nature 289, 373-378.
6. Webster, R.G., Laver, W.G., Air, G.M. and Schild, G.C. (1982) Nature 296, 115-121.
7. Villegas, P (1998). "Viral diseases of the respiratory system". *Poult Sci* **77** (8): 1143–5. PMID 9706079.
8. Horwood F, Macfarlane J (October 2002). "Pneumococcal and influenza vaccination: current situation and future prospects". *Thorax* 57 (Suppl 2): II24–II30. PMID 12364707. PMC: 1766003. http://thorax.bmj.com/cgi/pmidlookup?view=long&pmid=12364707.
9. Kawaoka Y (editor). (2006). *Influenza Virology: Current Topics*. Caister Academic Press. ISBN 978-1-904455-06-6. http://www.horizonpress.com/flu.
10. Vainionpää R, Hyypiä T (April 1994). "Biology of parainfluenza viruses". *Clin. Microbiol. Rev.* **7** (2): 265–75. PMID 8055470. PMC: 358320. http://cmr.asm.org/cgi/pmidlookup?view=long&pmid=8055470.
11. Hall CB (June 2001). "Respiratory syncytial virus and parainfluenza virus". *N. Engl. J. Med.* **344** (25): 1917–28. PMID 11419430. http://content.nejm.org/cgi/pmidlookup?view=short&pmid=11419430&promo=ONFLNS19.
12. Carrat F, Luong J, Lao H, Sallé A, Lajaunie C, Wackernagel H (2006). "A 'small-world-like' model for comparing interventions aimed at preventing and controlling influenza pandemics". *BMC Med* **4**: 26. doi:10.1186/1741-7015-4-26. PMID 17059593.
13. Hall CB (August 2007). "The spread of influenza and other respiratory viruses: complexities and conjectures". *Clin. Infect. Dis.* **45** (3): 353–9. doi:10.1086/519433. PMID 17599315. http://www.journals.uchicago.edu/doi/full/10.1086/519433.
14. Tellier R (November 2006). "Review of aerosol transmission of influenza A virus". *Emerging Infect. Dis.* 12 (11): 1657–62. PMID 17283614. http://www.cdc.gov/ncidod/EID/vol12no11/06-0426.htm.
15. Brankston G, Gitterman L, Hirji Z, Lemieux C, Gardam M (April 2007). "Transmission of influenza A in human beings". *Lancet Infect Dis* **7** (4): 257–65. doi:10.1016/S1473-3099(07)70029-4. PMID 17376383.
16. Cole E, Cook C (1998). "Characterization of infectious aerosols in health care facilities: an aid to effective engineering controls and preventive strategies". *Am J Infect Control* **26** (4): 453–64. doi:10.1016/S0196-6553(98)70046-X. PMID 9721404.
17. Tumova, B., Veznikova, D., Mensik, J. & Stumpa, A. (1980). Surveillance of in¯uenza in pig herds in Czechoslovakia in 1974±1979. I. Introduction of in¯uenza epidemic A (H3N2) viruses in pig herds. *Zentralblatt fuXr Veterinaermedizin Beiheft}Reihe B* **27**, 517±523.
18. Garnier J, Gibrat J-F, Robson B (1996). GOR secondary structure prediction method version IV. Meth Enzymol, 266:540-553
19. Dill KA. Dominant forces in protein folding. Biochemistry. 1990;29:7133–55. [PubMed]
20. Parker JM, Guo D, Hodges RS. 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 25:5425-5432. PMID: 2430611
21. Kolaskar AS, Tongaonkar PC. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*276:172-174. PMID: 1702393

22. Schmidt, A.M. (1989) Biotect. Adv. 7, 187-213.
23. Hopp, T.P. and Woods, K.R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences USA*, 78: 3824-3828.
24. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C., (1978). A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), volume 5, Suppl. 3, pp. 345-352. National Biomedical Research Foundation, Silver Spring, Washington D.C.
25. Han, K. F. & Baker, D. (1995). Recurring local sequence motifs in proteins. *J. Mol. Biol.* **251**, 176-187.
26. Ladunga, I. & Smith, R. F. (1997). Amino acid substitutions preserve protein folding by conserving steric and hydrophobicity properties. *Protein Eng.* 10, 187-196.