

# Linear B cell epitope prediction for epitope vaccine design against meningococcal disease and their computational validations through physicochemical properties

Sharat Chandra · Tiratha Raj Singh

Received: 4 June 2012/Revised: 18 September 2012/Accepted: 20 September 2012/Published online: 12 October 2012  
© Springer-Verlag Wien 2012

**Abstract** *Neisseria meningitidis* serogroup B is predominantly known for its leading role in bacterial meningitis and septicemia worldwide. Although, polysaccharide conjugate vaccines have been developed and used successfully against many of the serogroups of *N. meningitidis*, such strategy has proved ineffective against group B meningococci. Here, we proposed to develop peptide epitope-based vaccine candidates from outer membrane (OM) protein contained in the outer membrane vesicles (OMV) based on our in silico analysis. In OMV, a total of 236 proteins were identified, only 15 (6.4 %) of which were predicted to be located in the outer membrane. For the preparation of specific monoclonal antibodies against pathogenic bacterial protein, identification and selection of B cell epitopes that act as a vaccine target are required. We selected 13 outer membrane proteins from OMV proteins while taking into consideration the removal of cross-reactivity. Epiptopia web server was used for the prediction of B cell epitopes. Epitopes are distinguished from non-epitopes by properties such as amino acid preference on the basis of amino acid composition, secondary structure composition, and evolutionary conservation. Predicted results were subject to verification with experimental data and we performed string-based search through IEDB. Our finding shows that epitopes have general preference for charged and polar

amino acids; epitopes are enriched with loop as a secondary structure element that renders them flexible and also exposes another view of antibody–antigen interaction.

**Keywords** B cell epitopes · Meningococcal · *Neisseria meningitidis* · Structural features · Vaccine candidates

## 1 Introduction

The interaction between an antibody and its antigen is at the heart of the humoral immune response. The detection of highly immunogenic regions within a given protein, specifically those that elicit a humoral immune response, i.e., B cell epitopes, is central to many immunodetection and immunotherapeutic applications (Irving et al. 2001). Antibodies bind to their corresponding antigens at discrete sites known as antigenic determinants of epitopes. The precise localization of an epitope can be essential in the development of biomedical applications such as designed vaccines, diagnostic kits, and immune-therapeutics (Westwood and Hay 2001). Predicting epitopes is fundamental to the understanding of the basis of immunological discrimination between self and non-self as well as for mechanisms of bio-recognition in general. Since proteins are one of the most abundant and diverse class of antigens, antigens of infectious agents, and allergens, much of the interest in antigen characteristics is focused on antigenic proteins (Rubinstein et al. 2008).

B cell epitopes are classified into two different groups. The first group consists of linear or continuous epitopes. A continuous epitope comprises a single, consecutive stretch of amino acids in the protein sequence, which is specifically recognized by an antibody raised against the intact protein. The second group is formed by conformational or

S. Chandra  
Molecular and Structural Biology Division, CSIR-Central Drug Research Institute, Council of Scientific and Industrial Research, Chattar Manzil Palace, PO Box 173, Lucknow 22600, India

T. R. Singh (✉)  
Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology (JUIT),  
Waknaghat 173234, India  
e-mail: tiratharaj@gmail.com

discontinuous epitopes. These are epitopes composed of residues separated in the protein sequence, but in spatial proximity because of the protein fold (Van Regenmortel et al. 1996). It has been suggested that sequence-level analysis at exon and peptide level can be a valuable profile (Liou, and Huang, 2012). A specific and isolated linear peptide, derived from the sequence of a given antigen, was able to elicit antibodies that not only bound the peptide, but also strongly cross-reacted with the native antigen. Peptides containing linear epitopes are considered to have a high potential for vaccines. In addition to the advantages of being used for vaccines; peptides are easily synthesized, purified, stored, and handled (Andersen et al. 2008).

*Neisseria meningitidis* is a major cause of childhood meningitis and septicemia worldwide. Meningococcal disease is associated with a high fatality rate (approximately 10 %) and survivors may develop permanent abnormalities such as deafness, seizures, amputation, and mental retardation (Offit and Peter 2003). A Gram-negative encapsulated bacterium *Neisseria meningitidis* is classified into groups according to the chemical composition and immunogenic properties of capsular polysaccharides. Serogroups A, B, C, W135, and Y account for >95 % of the infections. Capsular polysaccharide or capsular polysaccharide conjugate vaccines are available against serogroup A, C, Y, and W135 strains (Jóðar et al. 2002; Morley et al. 2001; Roupael and Stephens 2012). However, no capsule-based vaccine is available for *N. meningitidis* serogroup B. The immune system tolerates serogroup B capsular polysaccharide because of its similarity to human carbohydrate  $\alpha(2 \rightarrow 8)$ N-acetyl neuraminic acid or polysialic acid, both consisting of repeated units of two to eight linked sialic acid (Finne et al. 1983). Therefore, our study is focused on proteins of the outer membrane for development of effective epitope vaccine against meningococcal diseases. In our previous work, we predicted and characterized the T cell epitopes for epitope vaccine design from OMV proteins of *N. meningitidis* serogroup B (Chandra et al. 2010).

An unguided experimental search for such regions is clearly laborious and resource intensive. Thus, computational approaches that are able to perform this task are desired. The humoral immune response is based on antibody–antigen interaction which leads to a solution of numerous protein–protein interfaces. We compared these interfaces with respect to physico-chemical properties at the amino acid level from amino acid composition. Aside from the physico-chemical character of the interfaces, secondary structure content was also found to be an important property of protein–protein interfaces, and has been used as one of the analysis parameters. In the present study, the outer membrane proteins of outer membrane vesicles of *Neisseria meningitidis* serogroup B were used

for the prediction of linear B cell epitopes. Amino acid composition, physico-chemical properties, and secondary structure element preference were analyzed to identify putative patterns in the predicted epitopes.

## 2 Materials and methods

### 2.1 Collection of data

A set of 15 OM protein complements was selected from the OMV 236 non-redundant proteins and represented only 6.4 % of the total number of proteins detected (Williams et al. 2007). The complete genome and protein sequences of *N. meningitidis* serogroup B (MC58) were taken from Genbank (NCBI) and UniProtKB. The selected protein sequences were retrieved in FASTA format and used for further analyses.

### 2.2 Vaccine candidate characterization

Amino acid compositions were computed by using Expasy's ProtParam server (Gasteiger et al. 2005). The similarity of human protein to OM proteins of OMV were searched using BlastP. Protein secondary structure prediction was done by APSSP2: Advanced Protein Secondary Structure Prediction Server (Raghava 2002).

### 2.3 Epitope prediction

The Epitopia server (Rubinstein et al. 2009a, b) has been used to predict epitopes in the manually curated dataset. The Epitopia is a Web-based server that aims to predict immunogenic regions in either a protein three-dimensional structure or a linear sequence. Epitopia implements a machine-learning algorithm that has been trained to discern antigenic features within a given protein. The Epitopia algorithm (Rubinstein et al. 2009a, b) uses a naïve Bayes classifier to predict the immunogenic potential of protein regions. The classifier was trained to recognize immunogenic properties using a benchmark dataset of 66 non-redundant validated epitopes derived from antibody–antigen co-crystal structures and 194 non-redundant validated epitopes derived from antigen sequences.

A given antigen input is divided into overlapping surface patches (or stretches in the case of a linear sequence input), with the size of a typical epitope. Epitopia then computes for each patch (or stretch) the probability that it was drawn from the population of epitopes on which the classifier had been trained, with respect to each one of its physio-chemical and structural geometrical properties. The immunogenicity score is thus the sum of logs of these probabilities and is assigned to the central residue of the

patch or to the middle residue in the linear stretch (Rubinstein et al. 2009a, b).

#### 2.4 Validation of predicted epitopes

Predicted epitopes were searched in the IEDB database ([www.immuneepitope.org](http://www.immuneepitope.org)) (Vita et al. 2010), to find out the known experimental entry of epitopes. This search result shows the possibility of epitope immunogenic property on the basis of an already known experiment. All predicted epitopes were searched for linear epitope match as substring (Table 1).

### 3 Results and discussion

In OMV, out of 236 non-redundant proteins, only 15 (6.4 %) proteins were predicted to be located in the outer membrane. In our previous study, we found the similarity of two proteins with human proteins, when OM proteins were searched against translated human genome through BlastP. The fkpA (NMB1567) has limited similarity to the human FK506 binding proteins 2 (FKBP2) and AIP aryl hydrocarbon receptor interacting proteins, and omp85 (NMB0182) has limited similarity with human SAMM50 sorting and assembly machinery component 50 homolog, which might prove a plausible immunological cross-reactivity between vaccine and host cell proteins, and therefore could lead to autoimmunity. To overcome this technical limitation, we selected 13 out of 15 OM proteins for epitope prediction (Chandra et al. 2010). These 13 OM proteins did not show any significant similarity with human proteins and neglected the chances of autoimmunity response. The summary of the results of the prediction server EpiToPIA used for B cell epitope prediction in the present study is given in Table 1. The peptides are given according to their rank (from 1 to 5) of the most significant immunogenic stretches obtained by the clustering procedure (Rubinstein et al. 2009a, b). The predicted epitopes are searched in the IEDB database for their possible similarity to the linear epitope data present in the IEDB. We found that the predicted epitopes of porA and porB proteins had positive entry as linear epitope when searched as substring. Two predicted epitopes of the NMB1961 protein were also present in the IEDB database as linear epitopes. Some entries were also found in IEDB, while others were not found because they were not utilized as vaccine candidate in any previous study.

#### 3.1 Amino acid preference of epitopes

Feature selection is an important aspect of multifaceted analysis and has to be implemented properly for

multidimensional datasets in bioinformatics (Hulse et al. 2012). The peptides of 13 antigenic proteins were classified into epitope and non-epitope groups. The predicted peptides signify the epitopic region, while protein sequence of OM protein minus epitope sequence signifies the non-epitopic region. The dataset was constructed for both groups individually. The amino acid preference of epitopes differs from that of the remaining antigen surface (Jones and Thornton 1995; Lo Conte et al. 1999). Here, the amino acid preference is evaluated using the amino acid frequencies (Fig. 1). The overall amino acid composition is found to differ significantly between the epitope and non-epitope surface. The amino acid composition was computed and results were evaluated in three categories of amino acids: hydrophobic, charged and polar and other small amino acid. Our data show that charged and polar amino acids are found more frequently in epitopic group than in the non-epitopic group (Fig. 1). On the other hand, the non-epitopic group has more hydrophobic amino acid as compared to other amino acids (Fig. 1). These findings are in support of previous reports (Bogan and Thorn 1998; Jackson et al. 1999) claiming that charged residues are generally preferred in protein–protein interaction due to their capability to form a multitude of interactions.

#### 3.2 Epitopes' secondary structure

The important structural aspect of an epitope that may distinguish it from the remaining antigen protein part is its secondary structure composition. To test which kind of specific secondary structure elements the epitopes were enriched with, each amino acid was assigned to either of the following three secondary structure groups: (1) alpha-helix, 310 helix, and pi-helices were grouped as helix; (2) isolated beta bridges and extended beta strands were grouped as strands; and (3) turns, bends, and irregular structures were grouped as loops. This analysis revealed that epitopes were significantly enriched with loops and significantly depleted of helix and strands. The frequency of distribution in our analysis for loops, helices, and strands was 0.58, 0.25, and 0.17, respectively (Fig. 2). Since loops tend to be more flexible than other organized secondary structure elements (Jemmerson and Paterson 1985; Pellequer et al. 1991), these results suggest that epitopes are relatively flexible (Fig. 2). The abundance of flexible secondary structure elements in epitopes may facilitate the capacity demonstrated by epitopes to undergo conformational adjustment upon antibody binding.

#### 3.3 Evolutionary conservation of epitopes

Functional regions on protein surfaces tend to be evolutionarily conserved relative to other regions (Zhou

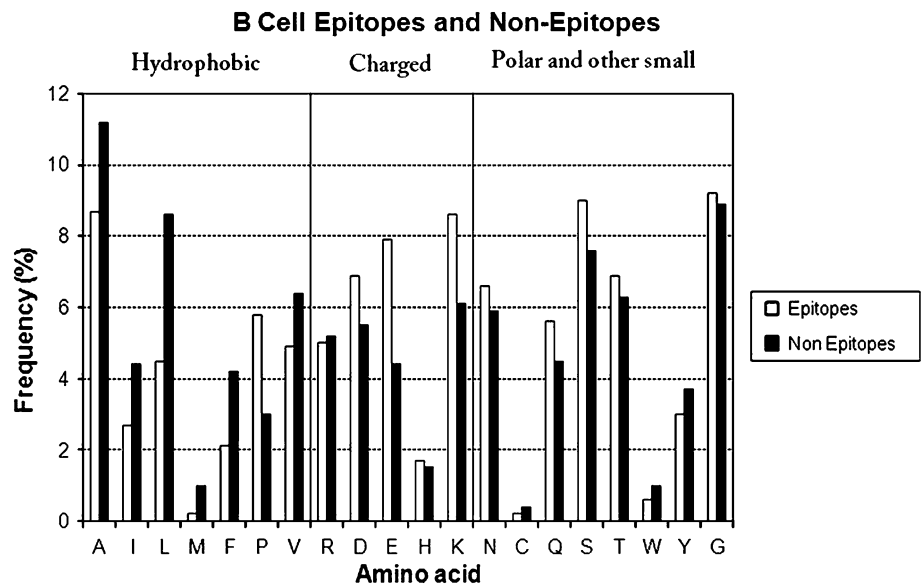
**Table 1** Epitope predictions with position and number of residues in each predicted instance. Epitopes were organized according to rank of antigenicity (top 5)

Protein name	Epitope (according to rank of antigenicity)	Position	Number of residues	IEDB epitope entry (substring search)
porA NMB1429	1. IQNSKSAYTPAYYTKNTNNNLT	186–207	22	Present
	2. DASQAIDPWDSNNDVASQLGIFKRHDDMP	136–164	29	Present
	3. GSDQAKGTDPLKNH	259–279	14	Absent
	4. RNTGIGNYTQIN	371–382	12	Present
	5. GRNYQLQLTEAQAANGGASGQVK	33–55	23	Present
porB NMB2039	1. TSRSVFHQNGQVTEVTTATG	33–52	20	Present
	2. QDAKLTDASNSHNSQ	234–248	15	Present
	3. RHHQVQEGLNIEKYQ	198–212	15	Present
	4. DADIGNEYD	278–286	9	Present
	5. RVGRLNSVLKDTGDINPWDSKSD	118–130	13	Present
Opc NMB1053	1. LETDENKLGKTKNVKLPTGVPENRID	85–110	26	Absent
	2. SISSTRAFLKEKHKAACH	37–54	18	Absent
	3. QRTYKESGEFSVTTKSGDVSL	234–253	20	Absent
	4. FESSKDSIKTTKHTLHSSRQS	135–155	21	Absent
	5. SRYKLNTG	183–193	8	Absent
pilQ NMB1812	1. PKKQNLESGGVNNAKPTFTGRK	326–347	22	Absent
	2. GGDDKWGAETKIN	561–573	13	Absent
	3. KDSPAQCASGNQ	676–688	13	Absent
	4. IANGGSSTNTELKKAVALGLT	638–652	15	Absent
	5. PAKQAAAAPSTK	146–157	12	Absent
rmpM NMB0382	1. EPEPEPEPAPA	71–81	11	Absent
	2. KLGAKVSKAKKRE	197–209	13	Absent
	3. QGRVECGD	58–65	8	Absent
	4. KDSLRAE	109–115	7	Present
	5. SNGVPSRISAVGLGESQAQ	167–186	20	Absent
NMB0345	1. DSREVKVPSFDE	246–257	12	Absent
	2. KTQPVSEQEVKAAAYDNISGFYKGTQE	129–154	26	Absent
	3. AENSRAEDTPQ	45–55	11	Absent
	4. RLKLDRAE	76–84	9	Absent
	5. DLEQGVPP	208–215	8	Absent
mtrE NMB1714	1. PIPEDLPAGLPLDKQFFVEKLPAGLS	249–274	26	Absent
	2. QYEQPKVEVAETFKNDTADS	26–45	20	Absent
	3. ANDSRQGSLSGGNVSSSYK	106–124	19	Absent
	4. LRYKAGV	202–208	7	Present
	5. GGLKRDTQTDK	457–467	11	Absent
NMB0088	1. KAELVFEKEKTVVKGKSDRRTTITPN	328–352	25	Absent
	2. DTSYRTAKASGNDVD	427–441	15	Absent
	3. SATDFTGLPVQGSKSGK	85–101	17	Absent
	4. WGIKSKAEILTAKPPKPNG	190–208	19	Absent
	5. IQADGHADVKGSD	215–227	13	Absent
Hap NMB1985	1. ETAREPVYHAAGGVNSYRPLNNGEN	369–394	26	Absent
	2. YSFNDDNNGTGKINAKHEH	330–358	29	Absent
	3. QREAETRPATTAFPARRARRDLPLQLPQ	1136–1164	29	Absent
	4. RNNYKAGTKGHPYGG	143–157	15	Absent
	5. DHWSQKEGIPRGE	656–668	13	Absent

**Table 1** continued

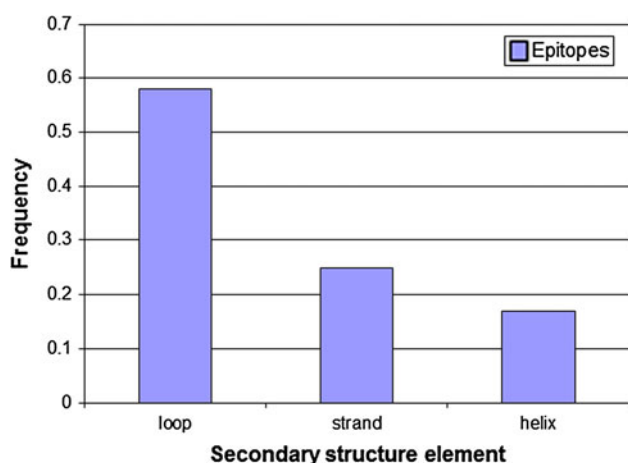
Protein name	Epitope (according to rank of antigenicity)	Position	Number of residues	IEDB epitope entry (substring search)
IptD NMB0280	1. NESGSPERTEAAVQGSGEASIPEDY	65–89	25	Absent
	2. EETDNPTAGESVRSVSSEPIQP	31–51	21	Absent
	3. RFGSQEARRVSR	488–499	12	Absent
	4. NEKIYLKSDGSYFYDK	679–694	16	Absent
	5. QQTGEAHNVRMEIEQGRRLLQSV	160–182	23	Absent
nspA NMB0663	1. PNVDLDAGYRYNYIGKVNTVKNVRSGE	139–165	27	Absent
	2. NYKAPSTDFK	70–79	10	Present
	3. ASSSLGSAKGFS	36–47	12	Absent
	4. NRASVDLGGSDSFSQTS	108–124	17	Absent
	5. AEGASGFY	19–26	8	Absent
NMB1961	1. SAETGAAETA VQEDSVSETQAEAAGEAET	241–269	29	Present
	2. APAFAETRPADPYEG	17–31	15	Absent
	3. RQTGATPAEGTEDNID	219–234	16	Present
	4. GGIPDNKNT	117–125	9	Absent
	5. TPVGRW	172–177	6	Absent
tbpB NMB0460	1. DEVKLDDESDWEATGLPDEPKELPKRQKSV	79–107	29	Absent
	2. TEAPRPAPKYQDVFSEKPAQKDQGG	35–60	26	Absent
	3. PSNHQNGNTGNGINQPKNQAKDYEN	127–151	25	Absent
	4. HAKREFNLKVEPKSAKNGDD	163–182	20	Absent
	5. DDGEEYSNKNKSTLTDGQEGY	240–260	21	Absent

**Fig. 1** Amino acid preference of epitope and non-epitopes. All 20 standard amino acids and their frequency of usage have been represented on the *x* and *y* axes, respectively. Amino acids have also been classified into three distinct classes of biochemical properties (hydropathy, charged, polar and other small), indicated on top of the image



and Shan 2001). Epitopes may overlap such functional regions due to shared constraints imposed by the nature of protein–protein interactions. If so, epitopes should be more evolutionarily conserved than the remaining antigen surfaces. But epitopes are enriched with unorganized secondary structures (loops). It is claimed that amino acid replacements in surface loops usually do not

perturb the three-dimensional structure of the protein, since surface loops are relatively flexible (Saunders and Baker 2002). Thus, the conservation variability of epitopes might be biased by the abundance of loops in epitopes. These results imply that epitopes do not tend to overlap functional regions, but rather cover separate regions.



**Fig. 2** Distributions of secondary structure elements in epitopes. Major secondary structure elements such as loops, helices, and sheets have been incorporated. These elements and their frequency of distribution have been represented on the *x* and *y* axes, respectively

This study also proposed some new directions for further analysis of this kind of dataset. One such approach could be the comparative analysis using various available tools/servers/databases under similar conditions and parameters to check the performance of these methods. Such kind of analyses also provide robust and repetitive data which could help experimental biologists to verify it experimentally (Gupta et al. 2011). This kind of approach could also be applied to other pathogens and diseases. The generated data could also be implicated in the development of novel prediction methods for linear and discontinuous epitope vaccine candidates by applying machine-learning techniques.

#### 4 Conclusion

The predicted epitopes from the genome/proteome sequences of the pathogens would greatly reduce the time as well as cost and be useful for experimental planning in the development of epitope vaccines. We have predicted numerous epitopes in OMV proteins, which would be useful for the earlier identification of meningitis and septicemia. The predicted epitopes may be used for safe vaccine development against meningococcal diseases. Epitopes are enriched with charged and polar residues and depleted of hydrophobic ones. Epitopes are enriched in loops and depleted in helix to beta strands. The lack of conservation can be partially explained by the enrichment of loops in epitopes, as they are relatively tolerant to amino acid replacements. An additional explanation for the lack of epitope conservation involves self-tolerance, as the conserved regions of antigen, which are expected to have much lower potential for eliciting an immune response, may also present in the host itself. These results have

important implications for the development and use of vaccines based on epitopes of OM proteins subsequent to *in vitro* validations. However, all the findings of this study will definitely aid in the process of vaccine design.

#### References

- Andersen P, Mkhailov D, Lund O (2008) Prediction and identification of B cell epitopes using protein sequence and structure information. Computational structural biology (Methods and Applications), World Scientific Press, Singapore
- Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol* 280:1–9
- Finne J, Leinonen M, Mäkelä PH (1983) Antigenic similarities between brain components and bacteria causing meningitis. Implications for vaccine development and pathogenesis. *Lancet* 2(8346):355–357
- Gasteiger E, Hoogland C, Gattiker A (2005) Protein identification and analysis tools on the ExPASy Server. In: John M. Walker (ed) *The Proteomics Protocols Handbook*, Humana Press, pp 571–607
- Gupta A, Chaukiker D, Singh TR (2011) Comparative analysis of epitope predictions: proposed library of putative vaccine candidates for HIV. *Bioinformatics* 5(9):386–389
- Hulse JV, Khoshgofaar TM, Napolitano A, Wald R (2012) Threshold-based feature selection techniques for high-dimensional bioinformatics data. *Netw Model Anal Health Inform Bioinforma* 1:47–61
- Irving MB, Pan O, Scott JK (2001) Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics. *Curr Opin Chem Biol* 5(3):314–324
- Jackson RM (1999) Comparison of protein–protein interactions is serine protease-inhibitor and antibody–antigen complexes: implications for the docking problem. *Protein Sci* 8:603–613
- Jemmerson R, Paterson Y (1985) Mobility and evolutionary variability factors in protein antigenicity. *Nature* 317:89–90
- Jódar L, Feavers IM, Salisbury D, Granoff DM (2002) Development of vaccines against meningococcal disease. *Lancet* 359(9316): 1499–1508
- Jones S, Thornton JM (1995) Protein–protein interactions: a review of protein dimer structures. *Prog Biophys Mol Biol* 63:31–65
- Liou S-W, Huang Y-F (2012) An exon/intron disparity framework based on the nucleotide profile of single sequence. *Netw Model Anal Health Inform Bioinforma* 1:19–25
- Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein–protein recognition sites. *J Mol Biol* 285:2177–2198
- Morley SL, Pollard AJ (2001) Vaccine prevention of meningococcal disease, coming soon? *Vaccine* 20(5–6):666–687
- Offit PA, Peter G (2003) The meningococcal vaccine: public policy and individual choices. *N Engl J Med* 349:2353–2356
- Pellequer JL, Westhof E, Van Regenmortel MH (1991) Predicting location of continuous epitopes in proteins from their structures. *Methods Enzymol* 322:891–901
- Raghava GPS (2002) APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning. *CASP5 A-132*
- Rouphael NG, Stephens DS (2012) *Neisseria meningitidis*: biology, microbiology, and epidemiology. *Methods Mol Biol* 799:1–20
- Rubinstein ND, Mayrose I, Halperin D, Yekutieli D, Gershoni JM, Pupko T (2008) Computational characterization of B-cell epitopes. *Mol Immunol* 45(12):3477–3489

- Rubinstein ND, Mayrose I, Martz E, Pupko T (2009a) Epitepia: a web server for predicting B-cell epitopes. *BMC Bioinfo* 10:287
- Rubinstein ND, Mayrose I, Pupko T (2009b) A machine-learning approach for predicting B-cell epitopes. *Mol Immunol* 46(5): 840–847
- Saunders CT, Baker D (2002) Evaluation of structural and evolutionary contributions to deleterious mutation prediction. *J Mol Biol* 322:891–901
- Van Regenmortel MHV (1996) Mapping epitope structure and activity: from one dimensional prediction to four dimensional description of antigenic specificity. *Methods* 9:465–472
- Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B (2010) The immune epitope database 2.0. *Nucleic Acids Res* 38(Database issue):D854–62 (Epub 11 Nov 2009)
- Westwood OMR, Hay FC (2001) *Epitope mapping: a practical approach*. Oxford University Press, London
- Williams JN, Skipp PJ, Humphries HE, Christodoulides M, O'Connor CD, Heckels JE (2007) Proteomic analysis of outer membranes and vesicles from wild-type serogroup B *Neisseria meningitidis* and a lipopolysaccharide-deficient mutant. *Infect Immun* 75(3): 1364–1372
- Chandra S, Singh D, Singh TR (2010) Prediction and characterization of T-cell epitopes for epitope vaccine design from outer membrane protein of *Neisseria meningitidis* serogroup B. *Bioinformation* 5(4):155–161
- Zhou HX, Shan Y (2001) Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins: Struct Funct Genet* 44: 336–343