Original Research Article

Identification of immunogenic T and B-cell epitope peptides of Rubella virus E1 glycoprotein towards development of highly specific immunoassays and vaccine

UNDER PEER REVIEW

Abstract

Introduction: The Rubella virus has worldwide occurrence and congenital Rubella syndrome are widely recognized as emerging infection in several parts of the world. Miscarriage, perinatal mortality, and stillbirth can develop in pregnancy during first trimester. The most frequent techniques for laboratory diagnosis of Rubella virus infection are IgM and IgG-based serological detection methods. Such emerging viral and bacterial pathogen emphasizes for the development of fast diagnostic devices, there is a need for enhanced and quicker methods. **Material and method:** Search for peptide vaccine with specific T and B-cell epitopes was identified through bioinformatics-based approaches. These were identified utilizing available Rubella virus E1 glycoprotein sequence databases. The outer-membrane glycoprotein, E1 is a target protein for the prediction of best antigens. **Results:** bioinformatics online software Bepipred 2 was used for prediction of potential B-cell epitope **PFCNTPHGQLEVQVPPDPGD** was identified and it has shown high conversation against E1 glycoprotein and few other bioinformatics online software NetMHCpan, IEDB and NetCTL was used to identify maximum surface-exposed residues. T-Cell epitope **RPVALPRAL** was identified and shown to be conserved to E1 Glycoprotein. Predicted epitopes were found to had promiscuous class-I major histocompatibility complex binding affinity to major histocompatibility complex supertypes , antigenicity scores and high proteasomal cleavage. The three-

dimensional modeled structures was created using I-TASSER online server for highlighting the predicted T- and B- cell epitopes.

Conclusion: The predicted T and B cell epitope could be used for development of immunoglobulin assay and vaccines.

KEY WORDS: Epitope, Rubella virus, peptide vaccine, B and T cell peptide.

UNDER PEER REVIEW

INTRODUCTION:

Rubella virus belongs to rubivirus genus of the Togaviridae family and it is a positive sense single-strand RNA. About 25% to 50% of infections are asymptomatic (1,2). In newborn children, CRS (Congenital Rubella Syndrome) generally cause mild illness, with visible manifestations. In infants a red rash is usually the first sign in those who do develop symptoms. The rash usually starts on the face and then spreads to the rest of the body, lasting around three days. (https://www.cdc.gov/rubella/about/symptoms.html) Worldwide more than 100000 infants are infected by CRS each year(3). Rubella is a highly contagious disease which is spread by respiratory droplets that often manifests in babies as a low-grade fever, arthralgia, maculopapular rash, and myalgia. (WHO 2011)(4). Infection in pregnant women in the first 12 weeks can cause foetal death or CRS, which manifests as cataract and glaucoma, neural deafness, heart disease and microcephaly in 80-90% of cases(5,6). Studies in India indicated a high seroprevalence of rubella infection(7). After vaccination and infection immune response developed in body with development of antibodies against the structural protein, which aimed towards E1, E2 and capsid protein (8,9). The antibodies produced against glycoprotein E1 glycoprotein continues to stay in the body for long time, but the antibodies produce against E2 and capsid protein stay for short time, it might differed by the accessibility of immune system(10,11)

The present vaccine is live attenuated and available as Measles-Mumps- Rubella (MMR) vaccine and is known to have some acute side effects(12,13). The seroconversion to the rubella virus is about <90% with one dose of MMR. Peptide-based vaccines with reduced adverse reactions could be developed. The present method of serological diagnosis is IgM (acute infection) or IgG (past exposure). These assays use recombinant antigens or viral antigens purified from lysate and could have low sensitivity ≤90% and a false positive rate of 5%. It is possible to improve the performance of these assays by peptide antigen(14). Rubella virus infections are common for males and females, however highly risk to immunosuppressed patients and pregnant women. Our study was aimed to select T and B cell peptides of E1 glycoprotein using bioinformatics tools and the predicted peptides can be used in the development of immunoassay and vaccine. This study is focused on rubella virus vaccine development and this vaccine candidate peptide can induce an immune response against rubella virus infection.

MATERIALS AND METHODS:

All available full-length coding sequence (n=181) of Glycoprotein E1 gene of rubella virus was downloaded from in the NCBI database. A consensus sequence was selected from the 181 Glycoprotein E1 sequences using the CLC sequence viewer 8.0 software. Linear B cell epitopes was predicted using two different programmes: BepiPred 2.0(15) and IEDB analysis resource (16). The Glycoprotein E1 gene consensus sequence was uploaded to each programme. Using the CLC Sequence viewer software, the epitope predicted by the two algorithms was evaluated for conservancy, and a conserved B-cell epitope candidate vaccine was chosen. The glycoprotein E1 gene coding sequence 3D protein structure stimulated by I-TASSER web server program.

RAMPAGE program(18) was used to examine the conformation of the modelled structure and the Ramachandran plot. For backbone dihedral angles ψ versus φ amino acid residues, the Ramachandran plot was used to visualize energetically allowed areas in the simulated protein structure. RAMPAGE shows the proportion of residues in different places, such as the preferred region, permitted region, and outlier region. The protein becomes more stable as the number of residues in the preferred area increases.

UNDER PEER REVIEW

To find discontinuous epitopes, the Discotope 2 program (19) was used using default prediction settings. The estimate was based on the 3D structure created by I-TASSER. The final scores are derived by adding the propensity scores of nearby residues and the surface accessibility (20). The ElliPro software assigns a score to each anticipated epitope, which was calculated by averaging the Protrusion Index (PI) over all epitope residues.

The IEDB analysis resource software was used to predict Class-I MHC binding T-cell epitopes. The human HLA allele reference set (27 alleles) with both 9-mer and 10-mer peptide length lengths was utilized for the prediction, which was done using the IEDB-recommended prediction technique. According to the percentile rank, the epitopes were predicted. Each MHC allele has a high affinity for it, as well as its length. is indicated by a tiny numerical percentile rank (1%). As a result, a available epitopes with percentile rankings below the threshold was select for further analysis.

MHC-I binding T-cell peptide lengths of 8 - 11mer were also predicted using NetMHC 4.0 server(21), Artificial Neural Networks (ANN) are used in this system. For prediction, the HLA representatives of supertypes and other defaults criteria (0.5 percent Rank) was chosen. The program's strongest binders was chosen for further investigation.

TAPred(22) was also used to predict TAP-binding peptides. Binders with high and moderate affinity was chosen using a cascade SVM prediction method. NetChop 3.1 sever(23) was used to predict protoplasmic cleavage.

The IEDB analysis resource was used to predict the immunogenicity of the chosen peptides in Class I. This bioinformatics tool predicts the immunogenicity of a peptide MHC complex based on amino acid characteristics and their location within the peptide. The Vaxijen online server software was used to forecast protective antigens without regard to alignment. The physicochemical characteristics of proteins are used to classify antigens.

RESULTS:

UNDER PEER REVIEW

B-CELL PEPTIDE PREDICTION

Separate epitopes was identified using two different algorithms, then listed and evaluated for conservation among the 181 sequences examined. The BepiPred 2 programme produced four peptide epitopes of different lengths. One epitope, however, was found in the highly conserved region (PFCNTPHGQLEVQVPPDPGD).

Among the four peptide epitopes produces by IEDB algorithm only one peptide (PFCNTPHGQLEVQVPPDPGD), fell within the appropriate size range. However, because the epitope was located in a very changeable area, it was deemed unsuitable.

In the I-TASSER-generated protein model, the chosen peptide PFCNTPHGQLEVQVPPDPGD was highlighted (Figure 1). The epitopes in the model was labelled using the Pymol software.

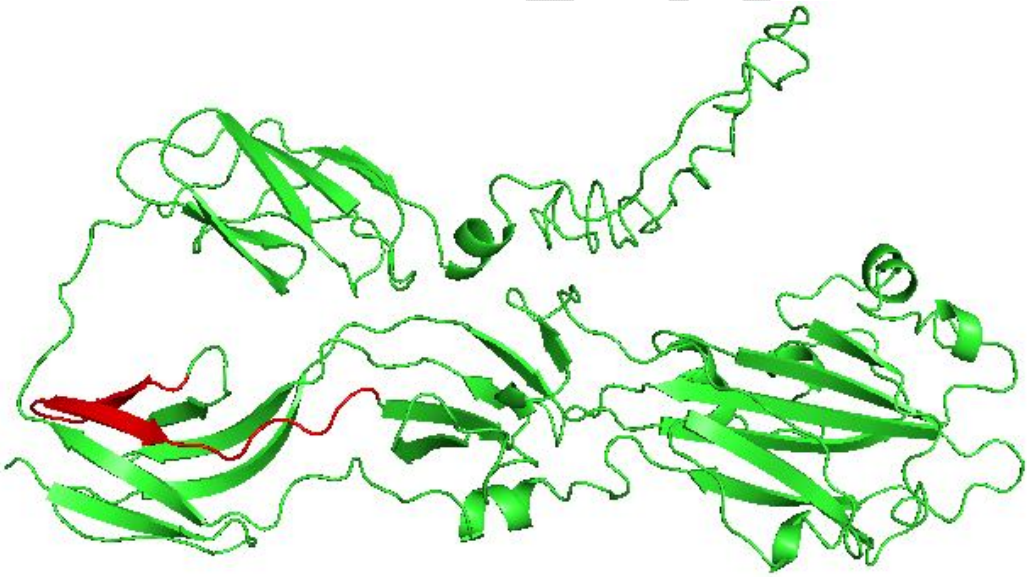


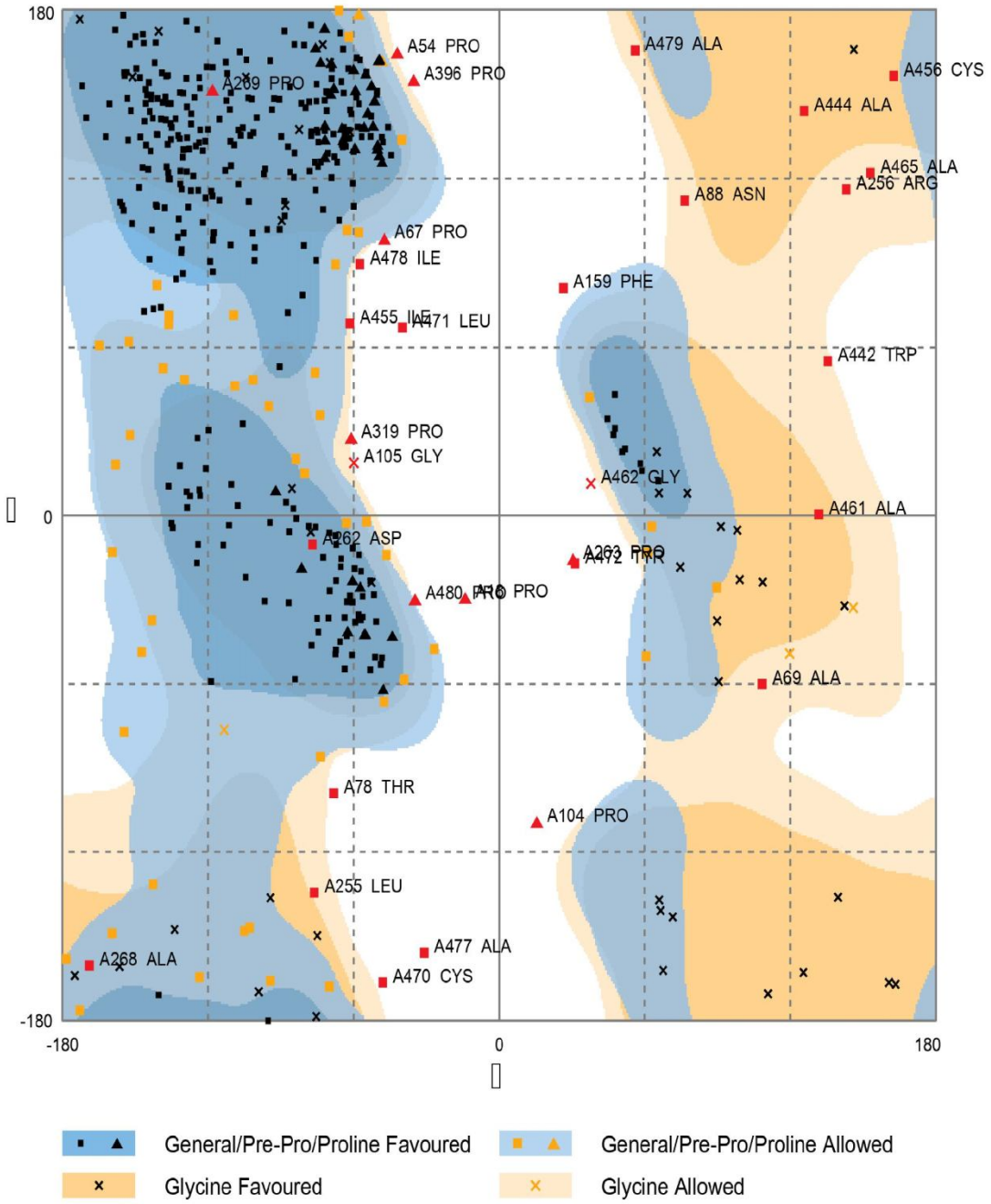
Figure 1: 3D protein structure of E1 of Rubella virus. The linear B-cell epitope predicted by Bepipred and IEDB is highlighted in red.

The estimated TM-score was 0.78+-0.10, and the estimated RMSD was 6.2+-3.8. The C-score was 0.49, the estimated TM-score was 0.78+-0.10, and the estimated RMSD was 6.2+-3.8. The C-score is a

confidence score used by I-TASSER to estimate the quality of projected models. It's computed using the importance of threading template alignments and the structure assembly simulations' convergence parameters. The C-score is usually in the range of [-5,2], with a higher C-score indicating a more confident model and vice versa. a TM-score of >0.5 suggests a model with correct topology, A random similarity is indicated by a TM-score of <0.17. The Ramachandran plot analysis found on the Rampage server revealed that the model created by I-TASSER was close to acceptable. 82.5 percent, 11.1 percent, and 6.5 percent of residues was found in the preferred, permitted, and outlier areas, respectively (Figure 2).

Figure 2: Molecular characterization of Rubella virus E1 glycoprotein using RAMPAGE program with residues scores

UNDER PEER REVIEW

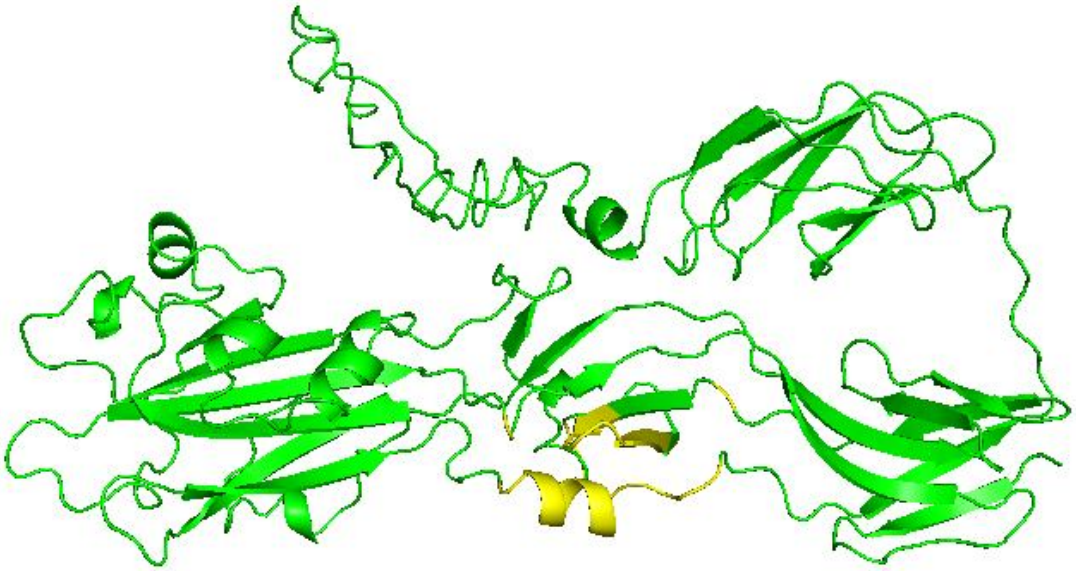


DISCONTINUOUS B-CELL PEPTIDE PREDICTION

Using the modelled 3D protein structure of the Glycoprotein E1 gene in PDB format, the Discotope 2 software predicted discontinuous epitopes, which are indicated in the modelled protein structure (Figure

3). The positions of the detected residues are 152; 201; 209-218; 220; 226-228; 302-134. Using the Pymol software, the epitopes was highlighted in the 3D structure.

Figure 3: Discontinous epitopes predicted by Discotope 2 bioinformatics tool. The predicted discontinuous epitopes are highlighted in yellow.



UNDER PEER

T-cell peptide prediction for MHC CLASS-I:

All three algorithms generated MHC-I predicted T cell peptide, which were given together with other prediction scores (Table1). One epitope, RPVALPRAL, TAPred score was high, and the protein survived proteasomal cleavage, as predicted by the NetChop 3.1 algorithm, as well as a high Class I Immunogenicity and Vaxijen score among the epitopes found by all three systems. HLA supertypes HLA-B*07:02 and HLA-B*39:01 exhibited a strong affinity for it. This was followed by another epitope, RPRLRLVDA, which showed strong MHC binding, TAP prediction, NetChop 3.1-predicted proteasomal cleavage, and good antigenicity ratings.

The MHC supertypes HLA-B*07:02 and HLA-B*08:01 were predicted to have a strong affinity for this epitope.

Table 1: T-cell peptides identified by three different programmes(IEDB, NetMHCPan and NETCTL)

List of epitopes	Survive Proteasomal cleavage	TAPred score	Vaxijen score	Immunogenicity score
AQSFTGVVY	Yes	3.841	0.1614 (Probable NON-ANTIGEN).	0.19276
RPVALPRAL	Yes	8.363	0.6515 (Probable ANTIGEN)	0.09807
GEVWVTPVI	Yes	8.573	-0.1840 (Probable NON-ANTIGEN)	0.32779
RPRLRLVDA	Yes	8.129	1.3364 (Probable ANTIGEN).	0.0934
CTFWAVNAY	Yes	3.483	0.3954 (Probable NON-ANTIGEN).	0.35525
MSVFALASY	Yes	4.219	0.3212 (Probable NON-ANTIGEN)	0.9522
DLVEYIMNY	Yes	3.974	-0.2477 (Probable NON-ANTIGEN)	0.08385
ETRQTWAEW	Yes	5.317	1.0473 (Probable ANTIGEN)	0.23807
FHTETRTVW	Yes	6.713	0.4227 (Probable NON-ANTIGEN)	0.25675
TPERPRLRL	Yes	3.880	-0.1644 (Probable NON-ANTIGEN)	0.14338
TETRTVWQL	Yes	3.841	0.6486 (Probable ANTIGEN)	0.14338

Proteasomal cleavage prediction by the NETchop 3.1program
TAPred score intermediate binding affinity
Vaxijen score with probable antigen, antigen and non-antigenic

UNDER PEER REVIEW

DISCUSSION:

The goal of this study was to find possible epitopes that might trigger cellular and humoral immune responses that could be used to development immunological assays and vaccines. We utilized bioinformatics tools to predict epitopes for rubella virus E1 glycoprotein. rubella virus has three structural proteins: E1 and E2, as well as a capsid protein. The glycoprotein is responsible for cellular response recognition[24].

In the prediction of B cell peptides, the Bepipred 2.0 server with an artificial neural network and the IEDB server was utilized. IEDB and NetMHC 4.0 servers based on artificial neuron networks were also used to predict T-cell epitopes. Consensus sequence was obtained using CLC sequence viewer 8.0 program. In addition, the physicochemical properties of B cell epitopes, including Discontinue epitopes using discotope programme, epitopes affinity prediction using TAPred, proteosomal cleavage analysis using NETCHOP, 3D structure was evaluated using I-TASSAR, and VAXIJEN online software was used for alignment-independent prediction. Results from the above analyses were scored to establish the reliability of predicted antigenic epitopes. Rubella virus E1 glycoprotein was subjected to Bepipred 2.0 cell epitope prediction tests. There were four peptides identified, with the 20 amino acid PFCNTPHGQLEVQVPPDPGD B cell epitope from 183 to 201 being the most acceptable. For the prediction of discontinuous B-cell epitopes, we applied Ellipro and Discotope. RPVALPRAL, the suggested T cell epitope, is highly conserved in rubella E1 glycoprotein. Three distinct bioinformatics programmes were used in the study. In E1 glycoprotein, the peptide between 330 and 339 was predicted.

The BepiPred2 algorithm predicted that the B-cell peptide epitope PFCNTPHGQLEVQVPPDPGD contained 12 exposed amino acid residues. The functional Bcell epitopes have a length of 15 - 25 residues and are surface accessible for effective antigen-antibody interactions. B-cell receptors detect surface accessible clusters of amino acids, which can activate a cellular immune response.

The antigenicity of the chosen epitope was predicted using the VaxiJen server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) with default values, which was designed to predict powerful antigen and subunit vaccines with an accuracy of 70 percent – 85 percent. Antigenic peptides with a

UNDER PEER REVIEW

higher antigenicity score (>1.0) did not survive the program's prediction of proteasomal cleavage. However, the chosen peptide, RPVALPRAL, got a score of 0.6515, suggesting that it was antigenic to a high degree. The C-score in the ITASSER model was 0.69, while the projected TM-score was 0.78+-0.10. The bioinformatics tools listed above was carefully considered in our study because different bioinformatics tools provide varied results. Several programmes were used for T and B cell peptide prediction, and only the harmonized consensus findings are given here.

CONCLUSION:

The best possible B-cell epitope was predicted as PFCNTPHGQLEVQVPPDPGD, while the best candidate T-cell epitope was identified as RPVALPRAL, which is conserved to rubella virus, based on the mean percent prediction probability score. These epitopes are ideal for the development of vaccines and immunoassays in the detection of antibodies to the rubella virus E1glycoprotein gene.

UNDER PEER REVIEW

REFERENCES:

1. Dominguez G, Wang CY, Frey TK Sequence of the genome RNA of rubella virus: evidence for genetic rearrangement during togavirus evolution. Virology (1990) 177: 225-238.

2. Battisti AJ, Yoder JD, Plevka P, Winkler DC, Prasad VM, et al. Cryo-electron tomography of rubella virus. J Virol (2012)86: 11078-11085.

3. George S, Viswanathan R, Sapkal GN. Molecular aspects of the teratogenesis of rubella virus. Biol Res. 2019 Aug 28;52(1):47.

4. WHO organization (2011) Weekly epidemiological record. Rubella vaccines: WHO position paper 86: 301-316.

5. Best JM (2007) Rubella. Semin Fetal Neonatal Med 12: 182-192.

6. Atkinson W, Wolfe S, Hamborsky J Epidemiology and Prevention of Vaccine-preventable Diseases, Centers for Disease Control and Prevention. 2011

7. Muliyil DE, Singh P, Jois SK, Otiv S, Suri V, Varma V, et al Sero-prevalence of rubella among pregnant women in India, 2017. Vaccine. 2018 Dec 18;36(52):7909-7912.

8. Mitchell LA, Tingle AJ, Décarie D, Shukin R Identification of rubella virus T-cell epitopes recognized in anamnestic response to RA27/3 vaccine: associations with boost in neutralizing antibody titer. Vaccine (1999) 17: 2356-2365

9. Petrova EK, Dmitrieva AA, Trifonova EA, Nikitin NA, Karpova OV The key role of rubella virus glycoproteins in the formation of immune response, and perspectives on their use in the development of new recombinant vaccines. Vaccine (2016) 34: 1006-1011

10. Gießauf A, Letschka T, Walder G, Dierich MP, Würzner R A synthetic peptide ELISA for the screening of rubella virus neutralizing antibodies in order to ascertain immunity. J Immunol Methods (2004) 287: 1-11.

11. Seto NO, Gillam S Expression and characterization of a soluble rubella virus E1 envelope protein. J Med Virol (1994) 44: 192-199.

12. Yavuz ST, Sahiner UM, Sekerel BE, Tuncer A, Kalayci O, et al. Anaphylactic reactions to measles–mumps–rubella vaccine in three children with allergies to hen's egg and cow's milk. Acta Paediatr (2011) 100: e94-e96.

13. Sukumaran L, McNeil MM, Moro PL, Lewis PW, Winiecki SK, et al. Adverse Events Following Measles, Mumps, and Rubella Vaccine in Adults Reported to the Vaccine Adverse Event Reporting System (VAERS), 2003- 2013. Clin Infect Dis (2015) 60: e58-65

14. Binnicker MJ, Jespersen DJ, Harring JA. Multiplex detection of IgM and IgG class antibodies to Toxoplasma gondii, rubella virus, and cytomegalovirus using a novel multiplex flow immunoassay. Clin Vaccine Immunol. 2010; 17(11):1734-1738

UNDER PEER REVIEW

15. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res 2017

16. Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006;2:2.

17. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: protein structure and function prediction. Nat Methods. 2015;12:7-8.

18. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins. 2003 Feb 15;50(3):437-50.

19. Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking Jens Vindahl Kringelum, Claus Lundegaard, Ole Lund, and Morten Nielsen Plos Computational Biology, 2012

20. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput Biol. 2012;8(12):e1002829.

21. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics. 2016 Feb 15;32(4):511-7.

22. Bhasin M, Raghava GP. Classification of nuclear receptors based on amino acid composition and dipeptide composition. J Biol Chem. 2004 May 28;279(22):23262-6.

23. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. *Genome Res*. 2005;15(11):1566 1575.

UNDER PEER REVIEW