



Structural Bioinformatics – cum- Immunoinformatic Approaches on the Major Glycoproteins to screen Potential Antigenic Determinants of Varicella Zoster virus in Vaccine design

Priya Roosvelt¹, Indu Purushothaman², Rajarajan S^{3*}

Bioinformatics Infrastructure Facility Centre (BIFC) of DBT,
P.G & Research Department of Microbiology and Biotechnology,
Presidency College (Aut), Chennai , India

Abstract: The Varivax vaccine presently used against chickenpox is live attenuated, and induces both cell-mediated and humoral immune response whose relative contributions and duration of protection is limited. The physio-chemical properties, prediction, validation of 3D structure, pairwise protein alignment and biological functions of five glycoproteins: gE, gB, gH, gI and gC, and their putative epitopes for potential T cell of HLA I and II have been predicted, using immunoinformatics and structural Bioinformatics tools to identify the effective peptide. gB protein was found to have highly similar structure based on its TM- score. The peptide showing the maximum binding with HLA class 1 molecule is VLLCLVIFL (gE) while LVIFLICTA (gE) and MVIVIVISV (gI) are peptides with HLA class 2 molecules. Hence considering this fact, even though gB had obtained the best TM-score it cannot be considered for proposing an effective subunit vaccine as it has very less allele predicted for both HLA class I and class II. The next best TM-score and hence the maximum alleles shown by gE for HLA class I and gI for HLA class II molecules could be considered in proposing an effective subunit vaccine.

Keywords: Varicella zoster virus, immunoinformatics, structural bioinformatics, protein structure, Major histocompatibility complex, class I, class II

I. Introduction

The viral genome Varicella zoster virus (VZV) consists of a long unique region (U_L) -10Kbp, short unique region (U_S) - 5.2Kbp with a pair of repeat elements in opposite directions (R_L - 88bp and R_S - 7.3bp) with terminal (T) and internal (I) copies of the repeats. The complete DNA molecule has the arrangement $TR_L-U_L-IR_L-IR_S-U_S-TR_S$. When DNA replication occurs, the U_S region inverts, resulting in the formation of two isomeric forms of VZV DNA. More than thirty structural proteins including six to seven glycoproteins are present in the VZV virions. The size of the messenger RNA (mRNA) ranges from 0.8 to 6.5 Kb [1]. The G+C contents of the U_L and U_S , TR_S and IR_S , TR_S and IR_L are 43- 44%, 59% and 68% respectively. The VZV is made of five important glycoprotein genes, namely, gpI, gpII, gpIII, gpIV and gpV.

The gpI (gE) is located in gene 68 in the U_S region of the genome and forms the major glycoprotein in the outer membrane of an infected cell with a relative molecular mass of 98,000D [2]. Two mammalian protein kinases namely casein kinases I and II are responsible for the phosphorylation of the polypeptide backbone on its serine and threonine residues [3]. Throughout the Alpha herpesviruses, it was found that the gE/gI heterodimer is conserved and is dispensable during replication [4]. It is reported that gE through its unique terminal extracellular domain binds to a cellular receptor, insulin-degrading enzyme (IDE) making it important for virus infectivity. Thus, the early stage of VZV infection was found to be related to an increase in virus internalization and cell-to-cell spread with the help of the recombinant soluble IDE [5].

gpII (gB) is the second most abundant and conserved VZV glycoprotein of molecular mass of 140,000D located in gene 31 present in the U_L region of the VZV genome [6]. In C57BL/6 mice, it is found to possess an immunodominant epitope that is mapped from position 498 to 505 amino acids, and predicted by common $CD8^+$ T cells after the infection of the target cells within 2 hours [7,8]. It plays an important role in the viral entry into the host cell by initial attachment, and the fusion of viral and cellular membranes. With the help of gK, it is found to be a part in virus-induced cell-to-cell fusion, which is widespread in VZV infection [9].

gpIII (gH) contains mainly complex N-linked oligosaccharide chains and is present in gene 37 located in the U_L region with a molecular mass of 118,000 D [10]. During virion morphogenesis, gH is found to be involved in the synthesis between the virion envelope and the outer nuclear membrane. The exact processing and cell surface expression of gH is due to the interaction that takes place between gH and gI [11]. The specific roles played by each domain of gH in VZV replication, was studied by homology modeling. For skin tropism, entry and fusion, the distal tip of the domain I (DI) is vital,

while DII helices and a conserved disulfide bond are necessary for the gH structure and VZV replication, and for DIII structural stability and membrane fusion, ⁷²⁴CXXC⁷²⁷ is important [12].

VZV gI (gpIV) is required for the formation of an envelope of viral particle in Trans-Golgi network [13]. It is located in the immediate upstream from the gpI in gene 67 in the Us region of the viral genome. It encloses mainly complex N-linked oligosaccharide chains with a molecular mass of 118,000D [10]. gI is found to be involved in cell-cell spread, secondary envelopment of virions and posttranslational modification [14-18]. For VZV infection of human skin and T cells, it is established that gI is necessary [19,20] in SCHIDhu mouse studies.

Finally, the gpV (gC) codes for gene 14 posses a molecular mass ranging from 95, 00 to 105,000D [21]. The gpV is found to have 34% amino acid homology to HSV-1 gC and is necessary for the early attachment to heparan sulfate moieties of the host cell surface proteoglycans [22], while it is not essential for the replication of VZV, as its synthesis is variable, when the virus is grown in tissue culture [23,24].

In this paper, the development of an effective vaccine for T cell immunity was identified by analyzing putative epitopes of the Varicella zoster virus Dumas strain. Among 47 strains/isolates sequences available in NCBI, the Dumas strain was selected in this study, as major works related to Varicella zoster virus have been carried out with the help of this strain. The Varivax vaccine presently used against chickenpox is live attenuated and induces both cell-mediated and humoral immune response, whose relative contributions and duration of protection is limited, and the manufacturer's claim is unknown. The vaccine used for protection against zoster is also a live attenuated vaccine of chickenpox origin. The physiochemical properties, such as molecular weight, isoelectric point, aliphatic index, antigenicity value and location of the glycoproteins, gE, gB, gH, gI and gC (homologous to HSV-1 and HSV-2 [25]), were predicted, using different latest bioinformatic tools and nonameric peptides, of important glycoproteins of Varicella zoster virus binding to Human leukocyte antigen (HLA) class I molecules by CTLPred and class II alleles by ProPred.

II. Materials and Methods

Retrieval of protein sequences and prediction of its physiochemical properties

The glycoprotein gE, gB, gH, gI and gC are coding for genes 68, 31, 37, 67 and 14 of the Varicella zoster virus have been used for immunoinformatic analysis. The sequence of the genes 68,31,37,67 and 14 were retrieved from NCBI protein database (<http://www.ncbi.nlm.nih.gov/protein>) for the Varicella zoster virus Dumas strain. The immunogenicity rates were found by VaxiJen v2. 0, at a threshold of 0.5% (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). The physiochemical properties of the selected glycoprotein such as molecular weight, isoelectric point, aliphatic index was predicted using the CLC BIO Main workbench version 6.8.1 (www.elcbio.com) and the location of the protein was predicted by Virus- PLoc (www.csbio.sjtu.edu.cn/bioinf/virus/).

Prediction of HLA class I and class II epitopes

All five glycoproteins were analyzed for potential T cell epitope, using immunoinformatic tools. All HLA class I and class II alleles were analyzed, using CTLPred (www.imtech.res.in/raghava/ctlpred) and ProPred (<http://www.imtech.res.in/raghava/PROPRED/>) with a threshold of 5%. The antigenic determinant peptides in all the five glycoproteins, binding to several HLA molecules with good binding affinity were identified using this tool.

The epitopes predicted for class I using CTLPred were cross verified, by using the existing online, namely, Epitope prediction (<http://atom.research.microsoft.com/bio/epipred.aspx>), NetHLA 3.2 version (<http://www.cbs.dtu.dk/services/NetHLA/>), IEDB (www.iedb.org) and SYFPEITHI version 1.0. (www.syfpeithi.de).

Molecular modeling and validation of predicted proteins

Homology modeling was performed using Phyre 2, Raptor X, CPH model 3.2 server and PS2-V2. Automated homology modeling was carried out by Phyre 2 to predict the 3D structure of protein. Raptor X predicts the structure of protein that are not similar to the ones available in protein data bank. CPH model server 3.2 uses profile-profile alignment along with secondary structure for prediction of 3D structure of protein.

Errat predict the overall quality factor of the predicted protein and it is performed by Structural analysis and verification server (SAVES version 4). Based on the structural features of the proteins, its quality is predicted by Pro Q (<http://www.sbc.su.se/~bjornw/ProQ/ProQ.html>).

What IF

It's a molecular modeling package (<http://swift.cmbi.ru.nl/whatif/>) which was used to complete the structure of predicted protein and remove bumps in the completed protein structure if any.

Structural Bioinformatics

TM-Align Based on structural similarity, residue-to-residue alignment is generated by TM-Align (<http://zhanglab.ccmb.med.umich.edu/TM-align/>) Cofactor. The biological function of the protein is predicted based on the structure of protein.

III. Result and Discussion

The glycoproteins which had the highest and lowest molecular weight were found to be gB (98.066kDa) and gI (39.364kDa) respectively. The stability of the protein, in particular, the isoelectric points are indicated by the pI value that ranged from 5.59 to 8.89. The relative volume occupied by the aliphatic side chains, namely, alanine, Valine, Isoleucine and

leucine is termed as Aliphatic index. The value obtained from the index can be regarded as the positive factor responsible for increasing the thermostability of proteins. Glycoprotein (gI) and gE had the highest (100.452) and lowest aliphatic index 74.591 (Table 1). Glycoprotein gI was the only glycoprotein which was found to be non-antigenic in nature, when compared with the other glycoproteins by alignment based prediction methods. The Virus-PLoc server was used for predicting the subcellular localization of viral proteins within hosts and virus-infected cells. While gC, gB, gH and gI were predicted to be found in the plasma membrane, gE was predicted to interact with the DNA as identified by the Virus-PLoc tool. The potent immunodominant epitopes for vaccine design were selected as peptides showing the coverage of the maximum number of HLA alleles and highest binding score. Using CTLpred and PROPRED, the binding specificity of the peptides to the HLA class I and class II molecules, was evaluated at 5% threshold value respectively (Table 2 and Table 3). The potent antigenic determinants presented by the HLA class I and II super types were found to be a total of 23 and 37 peptides.

The quality of proteins predicted by Phyre 2, Raptor X and CPH model 3.2 server was checked by Errat (Table 4). It was found that gB glycoprotein (78.60%) predicted by CPH model 3.2 server based on the Errat score was the best protein structure among the other three structure prediction tools used in this study (Figure 1).

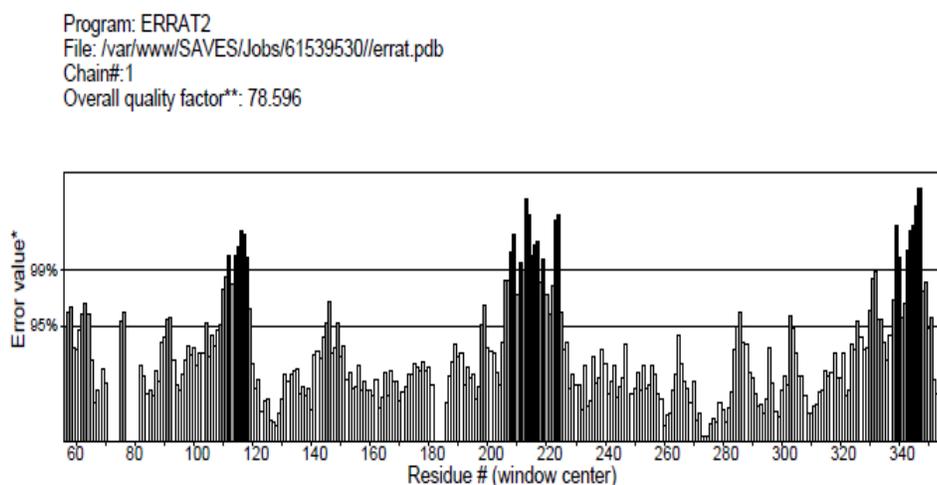


Figure 1: Errat score of gB predicted by CPH model 3.2 server

The local structural alignment of protein was produced based on continuous global optimization [26] and the finest alignment of protein was obtained from the source of information of predicted secondary structure and profile of a PSI-BLAST sequence [27]. These two techniques were exploited to identify the most potent target for subunit vaccine.

Structural similarity search for the glycoprotein protein and its corresponding template was performed using TM-align in this study. gE had a TM score of 0.98 and 0.95, gB had a TM score of 0.99 and 0.99, gH had a TM score of 0.95 and 0.94, gI had a TM score of 0.41 and 0.74 & gC had a TM score of 0.76 and 0.32 for chain 1 and chain 2 respectively. The structure of glycoprotein B (gB) was found to be highly similar to its template based on its TM-score (Table 5).

The Biological function of the protein was annotated using the Cofactor in order to identify functional site homologous to following glycoproteins. It was observed that gE had a structural similarity with protein 2giyA (HSV-1gE) and enzyme 1zvyA (Hydrolase), while gB had a structural similarity with protein 2gumb3 (HSV-1gB) and enzyme 2jfdA (Transferase), gH had a structural similarity with protein 3mlcA (regulator gH-gL) and enzyme 1cb8A (chondroitin AC lyase), gI had a structural similarity with protein 2giyA (HSV-1 gE ectodomain) and enzyme 2e1qA (human xanthine oxidoreductase mutant), gC had a structural similarity with protein 3f8uD (Tapasin/ ERp57 heterodimer) and enzyme 2bk8A (M1 domain from Titin) (Table 6).

Studies have been carried out earlier on epitopes of other members of Herpesviruses by the *in vitro*, *in vivo* and *in silico* methods. By *in vitro* studies, the sites of binding of the group II and V antibodies that recognize gD epitopes were identified. The epitopes were found in positions 293-312 in the sequence in the Group II antibodies, while in the Group V antibodies, the predicted epitopes were located in positions 365 to 381 in the sequence [28]. In, The HLA- DR1, DR4, DR7, DR13, DR15 and DRB5 molecules binding to T-cell gD epitopes were identified for VZV by *in vitro* analysis [29]. But so far epitopes prediction study on VZV has been concentrated only on gD protein. Also, the first web-based vaccine was designated for the three HSV-1 genomes, one HSV-2 genome, eight other human herpesvirus genomes and forty non-human herpesvirus genomes, using the Vaxign tools. UL26.5 was found to be the prominent target for the vaccine [30]. This study has mainly concentrated on proposing an effective vaccine against chicken pox exclusively and successfully found that Us 67 and Us 68 to be a prominent target for a vaccine. Using DNASTAR, Biosun and Antheptot tools, 20 potential immunodominant epitopes for gB2, gC2, gE2, gG2 and gI2 were identified [31]. After allogeneic transplantation, the major targets for Varicella zoster virus specific CD4+ and CD8+ reconstitution taking place during herpes zoster was found to be

gB and gE which can be used as a target for subunit vaccine production[32]. Our study also confirms that the gE glycoprotein will act as a potential target for vaccine production by insilico analysis.

The peptide showing maximum binding with HLA class I molecules is VLLCLVIFL (gE) while LVIFLICTA (gE) and MVIVIVISV (gI) are peptides for HLA class II molecules. These were reconfirmed by using various existing bioinformatic tools, while reconfirmation was not performed due to the non availability of all alleles for HLA class II in the existing tools. The result obtained based on the result of the TM-score showed that gB had the best TM-score among the other glycoproteins analyzed. But the alleles coverage of gB for both HLA class I and class II is low compared to the alleles depicted in gE and gI. Hence, keeping in the mind that only glycoprotein which has the maximum binding with HLA class I and class II molecules could act as an effective vaccine. Hence considering this fact, even though gB had obtained the best TM-score it cannot be considered for proposing an effective subunit vaccine as it has very less allele predicted for both HLA class I and class II. The next best TM-score and maximum alleles was shown by gE for HLA class I and gI for HLA class II molecules can be further considered in proposing an effective subunit vaccine. It is also confirmed that gE and gI glycoprotein indeed will act as a potential target based on its subcellular location which is predicted in this study.

IV. Conclusion

Without using the cultures of Varicella zoster virus, the antigenic determinants' prediction was done by immunoinformatic studies. While the structural similarities study, prediction of the biological function of proteins along with protein alignment studies was performed using different structural Bioinformatics and molecular modeling tools. Thus with the help of the biological functions, structural similarity of protein will provide more knowledge about the predicted epitope from these proteins. The predicted epitope is recognized against the HLA I and HLA II molecule, and can be used and tested for estimating the T cell responses to innate infection, and proposing an effective DNA vaccine or subunit vaccine against chickenpox.

Conflict of Interest Statement

There are no conflicts of interest for the authors of this study.

No study sponsors were involved in this study

Reference

1. R. Stranska, R. Schuurman, Machiel de Vos, M. Anton and V. Loon, "Routine use of a highly automated and internally controlled real-time PCR assay for the diagnosis of herpes simplex and varicella-zoster virus infection", *Journal of Clinical Virology*, vol. 30, pp. 39-44, 2004.
2. C. Grose, Dean P. Edwards, William E. Friedrichs, Kristen A. Weigle and William L. McGuire, "Monoclonal antibodies against three major glycoproteins of varicella-zoster virus", *Infection and immunity*, vol. 40, pp 381-388, 1983.
3. C. Grose, W. Jackson, J.A. Traugh, "Phosphorylation of varicella-zoster virus glycoprotein gpI by mammalian casein kinase II and casein kinase I". *Journal of virology*, vol. 63, pp. 3912-3918, 1989.
4. K.S. Dingwell, C.R. Brunetti, R.L. Hendricks, Q. Tang, M. Tang, A.J. Rainbow, D.C. Johnson, "Herpes simplex virus glycoproteins E and I facilitate cell-to-cell spread *in vivo* and across junctions of cultured cells", *J. Virol.* Vol. 68, pp. 834-845, 1994.
5. L.I. Qingxue, Mir A. Ali, Kening Wang, Dean Sayre, Frederick G. Hamel, Elizabeth R. Fischer, Robert G. Bennett, and Jeffrey I. Cohen. "Insulin Degrading Enzyme Induces a Conformational Change in Varicella-Zoster Virus gE, and Enhances Virus Infectivity and Stability" *PLOS ONE*, vol. 5, e11327- e11330, 2010.
6. J. Moffat, C. Mo, J. J. Cheng, M. Sommer, L. Zerboni, S. Stamatis, A. M. Arvin, "Functions of the C-terminal domain of varicella-zoster virus glycoprotein E in viral replication *in vitro* and skin and T-cell tropism *in vivo*". *J. Virol.* Vol. 78, pp. 2406-2415, 2004.
7. K.M. Khanna, R. H. Bonneau, P. R. Kinchington, R. L. Hendricks, "Herpes simplex virus-specific memory CD8 T cells are selectively activated and retained in latently infected sensory ganglia", *Immunity* vol. 18, pp. 593- 603, 2003.
8. S.N. Mueller, C. M. Jones, W. Chen, Y. Kawaoka, M. R. Castrucci, W. R. Heath, F. R. Carbone. "The early expression of glycoprotein B from herpes simplex virus can be detected by antigen-specific CD8 T cells". *J. Virol.* Vol. 77, pp. 2445-2451, 2003.
9. L. Maresova, T. Pasiaka, T. Wagenaar, W. Jackson, C. Grose, "Identification of the authentic varicella-zoster virus gB (gene 31) initiating methionine overlapping the 3' end of gene 30". *Journal of medical virology*, vol. 70, pp. S64-S70, 2003.
10. E.A. Montalvo, C. Grose, "Neutralization epitope of varicella zoster virus on native viral glycoprotein gp18 (VZV glycoprotein gpIII)", *Virology*, vol. 149, pp. 230-4, 1986.
11. L. Maresova, T. Pasiaka, T. Wagenaar, W. Jackson, C. Grose, "Identification of the authentic varicella-zoster virus gB (gene 31) initiating methionine overlapping the 3' end of gene 30". *Journal of medical virology*, vol. 70, pp. S64-S70, 2003.

12. Susan E. Vlecka, Stefan L. Olivera, Jennifer J. Brady, Helen M. Blau, Jaya Rajamania, Marvin H. Sommer, Ann M. Arvina, "Structure–function analysis of varicella-zoster virus glycoprotein H identifies domain-specific roles for fusion and skin tropism" *PNAS*, vol. 108, pp.18412-18417, 2011.
13. Z.H. Wang, M.D. Gershon, O. Lungu, Z. Zhu, S. Mallory, A.M. Arvin, A.A.Gershon, "Essential role played by the C-terminal domain of glycoprotein I in envelopment of varicella-zoster virus in the trans-Golgi network: interactions of glycoproteins with tegument". *J. Virol.* Vol.75, pp.323–340, 2001.
14. S. Mallory, M. Sommer, A. M. Arvin. "Mutational analysis of the role of glycoprotein I in varicella-zoster virus replication and its effects on glycoprotein E conformation and trafficking", *J. Virol.*, vol. 7, pp. 8279–8288, 1987.
15. C.Mo, J. Lee, M. Sommer, C. Grose, A. M. Arvin, "The requirement of varicella zoster virus glycoprotein E (gE) for viral replication and effects of glycoprotein I on gE in melanoma cells". *Virology*, vol. 304, pp.176–186, 2004.
16. J.K.Olson, C. Grose, "Complex formation facilitates endocytosis of the varicella-zoster virus gE:gI Fc receptor" , *J. Virol.* vol. 72, pp.1542–1551, 1998.
17. Z.H. Wang, M. D. Gershon, O. Lungu, Z. Zhu, S. Mallory, A. M. Arvin, A. A. Gershon, "Essential role played by the C-terminal domain of glycoprotein I in envelopment of varicella-zoster virus in the trans-Golgi network: interactions of glycoproteins with tegument" . *J. Virol.* vol. 75, pp.323–340, 2001.
18. Z. Yao, W. Jackson, B. Forghani, C. Grose, "Varicella-zoster virus glycoprotein gpI/gpIV receptor: expression, complex formation, and antigenicity within the vaccinia virus-T7 RNA polymerase transfection system" , *J. Virol.*vol. 67, pp. 305–314, 1993.
19. H.Ito, M. H. Sommer, L. Zerboni, H. He, D. Boucaud, J. Hay, W. Ruyechan, A. M. Arvin, "Promoter sequences of varicella-zoster virus glycoprotein I targeted by cellular transactivating factors Sp1 and USF determine virulence in skin and T cells in SCIDhu mice *in vivo*" . *J. Virol.*vol. 77,pp.489–498,2003.
20. J.Moffat, H. Ito, M. Sommer, S. Taylor, A. M. Arvin, "Glycoprotein I of varicella-zoster virus is required for viral replication in skin and T cells" , *J. Virol.* vol. 76, pp.8468–8471, 2002.
21. C.Grose, "Glycoproteins of Varicella-Zoster Virus and Their Herpes Simplex Virus Homologs", *Reviews of Infectious Diseases*, vol. 13, pp.S960-3, 1991
22. P.R. Kinchington, J. Remenick, J. M. Ostrove, S. E. Straus, W. T. Ruyechan, J. Hay, "Putative glycoprotein gene of varicella-zoster virus with variable copy numbers of a 42-base-pair repeat sequence has homology to herpes simplex virus glycoprotein C" . *J. Virol*, vol. 59, pp.660–668, 1986.
23. P.R.Kinchington, P. Ling, M. Pensiero, A. Gershon, J. Hay, W. T. Ruyechan, "A Possible role for glycoprotein gpV in the pathogenesis of varicella-zoster virus, *In* C. Lopez (ed.), Immunobiology and prophylaxis of human herpesvirus infections" . Plenum Press, New York, N.Y. pp. 83-91, 1990.
24. J.I.Cohen, K. E. Seidel, "Absence of varicella-zoster virus (VZV) glycoprotein V does not alter growth of VZV *in vitro* or sensitivity to heparin". *J. Gen. Virol.* vol. 75, pp. 3087–3093, 1994.
25. C. Grose. "Glycoproteins of Varicella zoster virus and their Herpes simplex virus homologs" . *Rev. Infect.Dis.* vol. 13, pp. S960- S963, 1991.
26. P. Bertolazzi, C Guerra, G Liuzzi, "A global optimization algorithm for protein surface alignment". *BMC Bioinformatics* , vol. 11,pp. 488, 2010.
27. Arne Elofsson "A study on protein sequence alignment quality". *Proteins: Structure, Function, and Genetics* ,vol. 46,pp. 330-339, 2002.
28. R. J. Eisenberg, D. Long, M. Ponce de Leon, J. T. Matthews, P. G. Spear, M. G. Gibson, L. A. Lasky, P. Berman, E. Golub, G. H. Cohen. "Localization of epitopes of Herpes Simplex Virus Type 1 glycoprotein D". *J.Virol.* vol. 53, pp. 634-644, 1985.
29. X. Zhang, Florence A. Castelli, X. Zhu, M. Wu, B. Maillere, L. BenMohamed. "Gender-dependent HLA- DR-restricted epitopes identified from Herpes simplex virus type I glycoprotein D". *Clinical and Vaccine Immunology.* vol. 15,pp. 1436-1449, 2008.
30. Z. Xiang, Y. He. "Genome-wide prediction of vaccine targets for human herpes simplex viruses using Vaxign reverse vaccinology", *BMC Bioinformatics*, vol. 14, pp. S2, 2012.
31. M. Pan, X. Wang, J. Liao, D. Yin, S. Li, Y. Pan, Y.Wang, G. Xie, S. Zhang, Y. Li1, "Prediction and Identification of Potential Immunodominant Epitopes in Glycoproteins B, C, E, G, and I of Herpes Simplex Virus Type 2", *Clinical and Development Immunology*, 2012.
32. P. Kleemann, E. Distler, E. M. Wagner, S. Thomas, S. Klobuch, S. Aue, E. Schnürer, H. Schild, M. Theobald, B. Plachter, S. Tenzer, R. G. Meyer, and W. Herr. "Varicella-zoster virus glycoproteins B and E are major targets of CD4+ and CD8+ T cells reconstituting during zoster after allogeneic transplantation". *Haematologica.* Vol. 97, pp. 874-882, 2012.

Table 1: Physiochemical properties of different glycoproteins of Varicella zoster virus with accession number

Putative proteins	Gene	Accession no	Expected Molecular weight (kDa)	pI value	Aliphatic index	Antigenicity value	Virus-PLoc – Predicted location
gE	68	CAA27951	69.956	5.52	74.591	0.5077	Plasma membrane
gB	31	CAA27914	98.066	8.47	79.954	0.5167	Plasma membrane
gH	37	CAA27920	93.651	6.39	95.91	0.5149	Plasma membrane
gI	67	CAA27950	39.364	8.56	100.452	0.4205	Interact with DNA
gC	14	CAA27897	61.352	8.89	81.339	0.6697	Plasma membrane

Table 2: Allele coverage of selected peptide binding to depicted HLA class I molecules of Varicella zoster virus Dumas strain glycoproteins

Glycoprotein	T cell epitopes	Start position	CTLPred	Atom research	NetMHC	IEDB	SYFPEITHI
gE	VLLCLVI FL	599	A2,A*0201,A*0205,A24 A3,A2.1,B14,B*2705,B*3701,B*3901,B*3902,B*5301,B*5401,B*51,B62,B7,B8,Cw*0301,Cw*0401, Cw*0602,H2-Db,H2-Db,H2-Kd	A02, A6802, C08 ,C05 ,C16 C03 ,C07 ,C06 A24 ,A28 ,C15 A42 ,A19 ,B40 B39 ,B52 ,C04 B1510 ,C18 ,C02 B46 ,B49 ,B63 B14 ,B1501 ,B38 B44 ,B1503 ,A32 B27 ,A23 ,A29 B50 ,B12 ,B17, A05 ,B24 ,C01 C14 ,C12 ,B18 B45 ,B37 ,A33 B51 ,B08 ,B13 B81 ,A03 ,A31 A26 ,A69 ,A74 A34 ,A11 ,A01 A30 ,B07 ,B42 A6801 ,B53 ,B57 B1516 ,B58 ,A25 B35 ,B55 ,B56	A0101,A0201,A0202,A0203,A0206,A0211,A0212 ,A0216,A0219,A0250,A0301,A1101,A2301,A2402 ,A2403,A2501,A2601,A2602,A2603,A2902,A3001 ,A3002,A3101,A3201,A3301,A6801,A6802,A6901 ,A8001,B0702,B0801,B0802,B0803,B1501,B1502, B1503,B1509,B1517,B1801,B2703,B2705,	B*07,A*02,B*38,B*58,A*32,B848,B*18,A*02,A*01,A*29,B*57,B*51,B*44, A*24,A*03,A*31, B*35,B*39,B*08,B*53,A*25,B*15,A*11,A*30,B*40,A*68,B*46,B*27,A*30 ,B*44,A*23,B*58, A*26	H2-Db,H2-Kd,H2-Kk,H2-Ld,A*01,A*02,A*003, A*11,A*24,A*26,A*68, B*07,B*08,B*13,B*14, B*15,B*18,B*27,B*35, B*37,B*38,B*39,B*40, B*41,B*44,B*45,B*47, B*49,B*51,B*53,B*58,
gB	TSSVEFA ML	495	B*3501,B*3701 B*3902,B*5801 B60,B7,Cw*0301, H2-Db,H2-Kb,H2-Ld	A01,C08 ,A32 B14 ,B1510 ,A29 B57 ,B39 ,B38 A26 ,C16 ,A6802 C04 ,B52 ,B53 C03 ,C05 ,B08 B58 ,C06 ,C07 B1501 ,B51 ,A28 B81 ,A19 ,A42 C15 ,B1516 ,B35 A30 ,B27 ,C18 C02 ,B46 ,B63 B1503 ,B42 ,B07 B13 ,A23 ,B12 B17 ,A05 ,C12	A0101, A0201,A0202,A0203,A0206,A0211,A0212,A0216 , A0219,A0250,A0301,A1101,A2301,A2402,A2403 ,A2501,A2601,A2602,A2603,A2902,A3001,A3002 ,A3101,A3201,A3301,A6801,A6802,A6901,A8001 ,B0702,B0801,B	B*07,A*02,B*38,B*58,A*32,B*48,B*18,A*02A*01,A*29,B*57,B*51,B*44, A*24,A*03,A*31, B*35,B*39,B*08,B*53,A*25,B*15,A*11,A*30,B*40,A*68,B*46,B*27,A*30 ,B*44, A*23,B*58,A*26,	H2-Db,H2-Kb,H2-Kd,H2-Kk,H2-Ld,A*01, A*02,A*03,A*11,A*24, A*26,A*68,B*07,B*08, B*13,B*14,B*15,B*18, B*27,B*35,B*37,B*38, B*39,B*40,B*41,B*44, B*45,B*47,B*49,B*51, B*53,B*57,B*58,

				C14 ,B24 ,C01 B40 ,A25 ,B44 A24 ,B49 ,B55 B56 ,B45 ,A33 B50 ,A34 ,A74 A31 ,B37 ,B18 A6801 ,A03 ,A11 A69 ,A02	0802,B0803,B15 01,B1502,B1503, B1509,B1517,B1 801,B2703,B270 5,B3501,B3801, B3901,B4001,B4 002,B4402,B440 3,B4501,B4601, B4801,B5101,B5 301,B5401,B570 1,B5801,B7301		
gH	FPNGTV VTL	831	A*0205,A24,B*35 01,B*3801,B*3902 ,B*5101,B*5102,B *5103,B*5301,B*5 401,B*51,B60,B7, B*0702,B8,Cw*04 01,Cw*0602,w*07 02,H2-Ld	B35,B81 ,B07 B53 ,B42 ,B51 C08 ,B1501 ,B57 A6802 ,B52 ,B55 A02 ,B56 ,B1510 C16 ,C05 ,B58 B1516 ,C03 ,C04 B13 ,A28 ,C06 C07 ,B38 ,A42 C15 ,A19 ,B40 B63 ,B39 ,B49 C18 ,C02 ,B46 A32 ,B45 ,B50 A23 ,A26 ,B1503 B44 ,B08 ,B17 A05 ,B12 ,C01 C14 ,C12 ,B24 B37 ,B14 ,A30 A29 ,A01 ,A69 B18 ,A24 ,B27 A33 ,A6801 ,A74 A34 ,A31 ,A25 A03 ,A11	A0101,A0201,A 0202,A0203,A02 06,A0211,A0212 ,A0216,A0219,A 0250,A0301,A11 01,A2301,A2402 ,A2403,A2501,A 2601,A2602,A26 03,A2902,A3001 ,A3002,A3101,A 3201,A3301,A68 01,A6802,A6901 ,A8001,B0702,B 0801,B0802,B08 03,B1501,B1502, B1503,B1509,B1 517,B1801,B270 3,B2705,B3501, B3801,B3901,B4 001,B4002,B440 2,B4403,B4501, B4601,B4801,B5 101,B5301,B540 1,B5701,B5801, B7301	B*07,A*02,B*38,B *58,A*32,B*48,B* 18,A*02,A*01,A*2 9,B*57,B*51,B*44, A*24,A*03,A*31, B*35,B*39,B*08,B *53,A*25,B*15,	H2-Db,H2-Kd,H2- Kk,H2- Ld,A*01,A*02,A*03,A *11,A*24,A*26,A*68,B *07,B*08,B*13,B*14,B *15,B*18,B*27,B*35,B *37,B*38,B*39,B*40,B *41,B*44,B*47,B*49,B *51,B*53,B*58,

gI	FYIQVT NAL	63	A24,B*3701,B*3801,B*3901,B*3902,B*5102,B*5301,B*5401,B*51,Cw*0301,Cw*0401,Cw*0702,H2-Db,H2-Kb,H2-Kd	A24 ,A23 ,C08 C05 ,C16 ,C03 C07 ,B39 ,B1510 A28 ,B38 ,C06 A42 ,A19 ,C15 C04 ,B52 ,B1501 C18 ,B46 ,C02 B81 ,B40 ,B07 B1503 ,B63 ,B14 A32 ,B53 ,B35 A05 ,B17 ,B12 C12 ,C14 ,B24 C01 ,B27 ,A02 B42 ,B49 ,B13 A6802 ,A29 ,B44 B51 ,B50 ,A26 B08 ,B45 ,B57 B37 ,A30 ,B58 B18 ,A01 ,B55 B1516 ,B56 ,A25 A33 ,A31 ,A74 A34 ,A69 ,A11 A6801 ,A03	A0101,A0201,A0202,A0203,A0206,A0211,A0212,A0216,A0219,A0250,A0301,A1101,A2301,A2402,A2403,A2501,A2601,A2602,A2603,A2902,A3001,A3002,A3101,A3201,A3301,A6801,A6802,A6901,A8001,B0702,B0801,B0802,B0803,B1501,B1502,B1503,B1509,B1517,B1801,B2703,B2705,B3501,B3801,B3901,B4001,	B*07,A*02,B*38,B*58,A*32,B*48,B*18,A*02,A*01,A*29,B*57,B*51,B*44,A*24,A*03,A*31,B*35,B*39,B*08,B*53,A*25,B*15,A*11,A*30,B*40,A*68,	H2-Db,H2-Kd,H2-Kk,H2-Ld,A*01,A*02,A*03,A*11,A*24,A*26,B*07,B*08,B*13,B*14,B*15,B*18,B*27,B*35,B*37,B*37,B*38,B*39,B*40,B*41,B*44,B*47,B*49,B*50,B*51,B*53,B*58
gC	FLYENIQ CV	201	A2,A*0201,A*0205,A3,A2.1,B14,B*2702,B*2705,B*3901,B*5101,B*5102,B*5103,B*5301,B*5401,B*51,B62,H2-Db-17	A02, A6802 ,B14 ,B1510,C03, B1503,A28 ,C08, C16,C05 ,B40, B39,B38 ,B51, A03 C06 ,C04 ,B08 C15 ,A42 ,A19 C07 ,B57 ,C02 B46 ,C18 ,A23 A32 ,B63 ,B52 B1501,B49 ,A69 A33 ,B58 ,B17 A05 ,B12 ,C14 C12 ,B24 ,C01 B1516 ,B81 ,A34 A74 ,B27 ,A6801 B42 ,A24 ,A31 A26 ,B50 ,B13 B45 ,B37 ,A29 B53 ,B44 ,B35	A0101,A0201,A0202,0203,A0206,A0211,A0212,A0216,0219,A0250,A0301,A1101,A2301,A2402,2403,A2501,A2601,A2602,2603,A2902, A3001,A3002,A3101,A3201,A3301,A6801,A6802,A6901,A8001,B0702,B0801,B0802,B0803,B1501	B*07,B*15,A*02,B*38,A*32,B*18,A*02,A*01,A*29,A*24,A*03,A*31,B*35,B*39,B*08,A*25,A*11,A*30,A*68,B*27,A*30,B*44,A*23,B*58,A*26,	H2-Db,H2-Kd,H2-Kk,H2-Ld,A*01,A*02,A*03,A*11,A*24,A*26,A*68,B*07,B*08,B*13,B*14,B*15,B*18,B*27,B*37,B*38,B*39,B*40,B*41,B*44,B*45,B*47,B*49,B*50,B*51,B*53,B*58,R TI.A1,

				A11 ,A30 ,A01 B18 ,B07 ,B55 A25 ,B56			
--	--	--	--	--	--	--	--

Table 3: Allele coverage of selected peptide binding to depicted HLA class II molecules of Varicella zoster virus Dumas strain glycoproteins

Glycoprotein	T cell epitopes	HLA class II alleles	netMHCII
gE	LVIFLICTA	DRB10306, DRB10307, DRB10308, DRB10311, DRB10402, DRB10404, DRB10408, DRB10410, DRB10423, DRB10801, DRB10802, DRB10804, DRB10806, DRB10813, DRB10817, DRB11101, DRB11102, DRB11104, DRB11106, DRB11107, DRB11114, DRB11120, DRB11121, DRB11128, DRB11301, DRB11302, DRB11304, DRB11305, DRB11307, DRB11311, DRB11322, DRB11323, DRB11327, DRB11328, DRB11501, DRB11502, DRB11506	DRB10301, DRB10401, DRB10404, DRB10405, DRB10701, DRB10802, DRB10901, DRB11101, DRB11302, DRB11501, DRB30101, DRB40101, DRB50101
gB	VVVFQPLLS	DRB10101, DRB10102, DRB10306, DRB10307, DRB10311, DRB10402, DRB10404, DRB10408, DRB10410, DRB10423, DRB11101, DRB11102, DRB11104, DRB11106, DRB11114, DRB11120, DRB11121, DRB11128, DRB11301, DRB11302, DRB11304, DRB11305, DRB11307, DRB11311, DRB11321, DRB11322, DRB11323, DRB11327, DRB11328, DRB11501, DRB11502, DRB11506	DRB10101, DRB10301, DRB10401, DRB10404, DRB10405, DRB10701, DRB10802, DRB10901, DRB11101, DRB11302, DRB11501, DRB30101, DRB40101, DRB50101
gH	MMIFTTWTA	DRB10402, DRB10404, DRB10408, DRB10423, DRB11102, DRB11114, DRB11120, DRB11121, DRB11301, DRB11302, DRB11322, DRB11323, DRB11327, DRB11328, DRB11501, DRB11506	DRB10101, DRB10301, DRB10401, DRB10404, DRB10405, DRB10701, DRB10802, DRB10901, DRB11101, DRB11302, DRB11501, DRB30101, DRB40101, DRB50101
gI	MVIVIVISV	DRB10101, DRB10102, DRB10301, DRB10305, DRB10306, DRB10307, DRB10308, DRB10309, DRB10311, DRB10402, DRB10404, DRB10405, DRB10408, DRB10410, DRB10423, DRB10701, DRB10703, DRB10802, DRB10804, DRB10806, DRB10813, DRB10817, DRB11101, DRB11106, DRB11107, DRB11128, DRB11301, DRB11305, DRB11307, DRB11311, DRB11321, DRB11327, DRB11328, DRB11501, DRB11502, DRB11506	DRB10101, DRB10301, DRB10401, DRB10404, DRB10405, DRB10701, DRB10901, DRB11101, DRB11302, DRB11501, DRB30101, DRB40101, DRB50101,
gC	IQINLITI	DRB10306, DRB10307, DRB10308, DRB10311, DRB10804, DRB10806, DRB10817, DRB11101, DRB11102, DRB11104, DRB11106, DRB11114, DRB11120, DRB11121, DRB11128, DRB11301, DRB11302, DRB11304, DRB11305, DRB11307, DRB11311,	DRB10101, DRB10301, DRB10401, DRB10404, DRB10405, DRB10701, DRB10802, DRB10901, DRB11101, DRB11302, DRB11501, DRB30101, DRB40101, DRB50101

		DRB11321, DRB11322, DRB11323, DRB11327, DRB11328, DRB11501, DRB11502, DRB11506	
--	--	---	--

Table 4: Validation of VZV glycoprotein protein by Errat and Pro Q

Glycoprotein	Errat (%)			
	Phyre 2	Raptor X	CPH Model	PS2-V2
gE	46.78	41.14	72.79	0.51
gB	60.73	64.26	78.60	60.61
gH	44.25	64.00	60.81	24.19
gI	21.49	47.50	61.19	0.53
gC	15.56	21.54	-	-

Table5: Structural similarity predicted by TM-Align

Glycoprotein	PDB ID	Aligned Length	TM score (normalized by chain 1)	TM score (normalized by chain 2)
gE	D2giya1	176	0.98	0.95
gB	C3nw8B	690	0.99	0.99
gH	C3m1cA	690	0.95	0.94
gI	C2gj7F	78	0.41	0.74
gC	C3f8uB	127	0.76	0.32

0.0 < TM-score < 0.30, random structural similarity

0.5 < TM-score < 1.00, in about the same fold

Table6: Biological function of the protein annotated by Cofactor to identify functional site homology

Glycoprotein	Description	PDB hit	TM-score	RMSD ^a	IDEN ^a	Cov.	Predicted active site residues
gE	similar structure	2giyA (HSV-1gE)	0.96	0.52	0.32	0.97	-
	similar enzyme	1zvyA (Hydrolase)	0.47	2.99	0.11	0.59	N/A
	Gene ontology	2giyA	0.96	0.52	0.32	0.97	-
	similar Binding site	1kn4L (catalytic Ab D2.3 complex)	0.47	3.48	0.14	0.61	44,46,148,150,156,158
gB	similar structure	2gumb3(HSV-1 gB)	0.45	0.96	0.54	0.46	-
	similar enzyme	2jfdA(Transferase)	0.15	7.07	0.04	0.233	94,189
	Gene ontology	3nw8D	0.99	0.61	0.50	1.00	-
	similar Binding site	3nw8B (HSV-1 gB)	0.99	0.33	0.50	0.99	60,62,188,190

gH	similar structure	3m1cA (regulator gH-gL)	0.94	1.04	0.25	0.95	-
	similar enzyme	1cb8A (chondroitin AC lyase)	0.33	7.70	0.05	0.52	NA
	Gene ontology	3m1cA	0.94	1.04	0.25	0.95	-
	similar binding site	3m1cA	0.94	1.04	0.25	0.95	571,573,583,584,585,587,606,608
gI	similar structure	2giyA (HSV-1 gE ectodomain)	0.75	2.04	0.15	0.88	-
	similar enzyme	2e1qA (human xanthine oxidoreductase mutant)	0.49	3.93	0.09	0.90	N/A
	Gene ontology	2giyA (gE-gI/Fc complex)	0.75	2.04	0.15	0.88	-
	similar binding site	2gj74	0.74	2.05	0.16	0.88	9,14,16,20,21,22,24,41,42,44,45,46
gC	similar structure	3f8uD (Tapasin/ ERp57 heterodimer)	0.76	2.66	0.14	0.91	-
	similar enzyme	2bk8A (M1 domain from Titin)	0.48	2.16	0.13	0.57	NA
	Gene ontology	2petA	0.73	3.00	0.17	0.95	-
	similar binding site	1h3uB(Glycoform)	0.55	3.44	0.13	0.70	51,53,70,71,101,105,107

TM-score = measure of global structural similarity between query and template protein.

RMSD^a =the RMSD between residues that are structurally aligned

IDEN^a = percentage sequence identity in the structurally aligned region

Cov. =the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein