Article ID: WMC005042 ISSN 2046-1690

The Glycoprotein Mucin-Like Domain (MLD) in the Zaire ebolavirus (EBOV) may be responsible for the manifestations of Post-Ebola Virus Disease Syndrome (PEVDS)

Peer review status:
No

Corresponding Author:
Dr. Robert A Ricketson,
Research Consultant, Hale O'mana'o Clinical and Biomedical Research, 230 East Second Street, 73034 - United States of America

Submitting Author:
Dr. Robert A Ricketson,
Research Consultant, Hale O'mana'o Clinical and Biomedical Research, 516 N Broadway, 73034 - United States of America

Other Authors:
Ms. Leah Roberts,
Independent research Consultant, Hale O\'mana\'o Biomedical Research - United States of America

Article ID: WMC005042

Article Type: Research articles

Submitted on:18-Jan-2016, 05:58:17 AM GMT **Published on:** 18-Jan-2016, 09:19:26 AM GMT

Article URL: http://www.webmedcentral.com/article_view/5042

Subject Categories:BIOINFORMATICS

Keywords: Post-Ebola Virus Disease Syndrome, PEVDS, Ebola mucin-like domain, Ebola MHC II epitopes, Ebola glycoprotein

How to cite the article: Ricketson RA, Roberts L. The Glycoprotein Mucin-Like Domain (MLD) in the Zaire ebolavirus (EBOV) may be responsible for the manifestations of Post-Ebola Virus Disease Syndrome (PEVDS). WebmedCentral BIOINFORMATICS 2016;7(1):WMC005042

Source(s) of Funding:

The authors have not received any funding from any source.

Competing Interests:

None

Additional Files:

The Glycoprotein Mucin-Like Domain (MLD) in the Zaire ebolavirus (EBOV) may be responsible for the manifestations of Post-Ebola Virus Disease Syndrome (PEVDS)

Author(s): Ricketson RA, Roberts L

Abstract

The incidence of various ocular, CNS, and musculoskeletal complaints in the convalescent period following recovery from Ebola Virus Disease (PEVDS) has become an intense area of interest in the aftermath of the recent West Africa outbreak 2014-2015. A review of past outbreaks involving EBOV, SUDV, TAFV, and BDBV revealed similar, but poorly documented symptoms in the convalescent period. Additionally, viable virus has been identified in the recovery phase in sperm and ocular fluid up to 9 months after recovery with no evidence of viremia. A prior study by Yang in 2000 demonstrated the mucin-like domain was responsible for the vascular permeability and inflammation seen in EVD with Zaire ebolavirus infections but not in Reston ebolavirus infections.

To study the mucin-like domain further, a comparative multi-sequence amino acid analysis of the poorly conserved MLD was performed and all 16 B-cell epitopes previously identified comparing all virulent species of Ebolavirus against RESTV was used to identify regions that could explain the differential virulence and vascular access to regions of immune privileged regions. Within the E1 epitope (301-316) of the MLD, the charged motif R (X1) EELSF demonstrated significant homology (86%) within the virulent species but poorly conserved in RESTV (25%). In E5-E8, charged tandem grouping of charged residues in the motif Glu-336, Asp-337, and His-338 of EBOV were conserved as to charge in the virulent group but these charged residues were replaced by the uncharged Pro-336, Thr-337, and Arg-338 in the RESTV sequence. The C-terminus GLINT motif in epitope E16 of the MLD at the GP2 junction was moderately conserved between all species including RESTV and therefore not felt to contribute to the overall differential virulence.

Within the MHC class I predictions, there did not appear to be a statistically significant difference between the virulent species and RESTV. However, the MHC Class II predicted epitopes did identify a statistically significant difference between the virulent species and RESTV. Three epitopes of acceptable binding affinities in the virulent group were significantly different from RESTV ($p>.05$)

This study suggests that the residues located within the B-cell E1 epitope (301-316) and three MHC Class II epitopes may be responsible for vascular access to these immune privileged site and persistent symptoms observed in the convalescent PEVDS. Further studies are required into the potential role of VP40 in viral persistence and reemergence within these immune privileged areas; therefore potentially provide a therapeutic strategy in PEVDS and viral persistence associated with reemergence.

Introduction

The Zaire ebolavirus outbreak 2014-2015 has had a total of 28,005 reported confirmed, probable, and suspected cases with 11,287 reported deaths as of August 31, 2015. Of the survivors, Carod-Artal, et al and others have reported persistent symptoms in over 16,500 termed the "**Post-Ebola Virus Disease Syndrome" (PEVDS)**. These clinical manifestations in the convalescent stage have included chronic joint and myalgias, fatigue, anorexia, hearing loss, headache, sleep disturbances, short-term memory loss, depressed affect, and ocular symptoms. Documentation of convalescent symptomatology in prior outbreaks has been limited. Visual problems in as many as 50% of survivors in Kenema, Sierra Leone were recently reported by Nanyonga in October, 2014 [44].

To study these convalescent symptoms in Liberia, in June of 2015, a study was launched following people in Liberia who have survived Ebola virus disease (EVD) within the past two years every 6 months for 5 years called **PREVAIL**. PREVAIL is sponsored by the Ministry of Health of Liberia and the National Institute of Allergy and Infectious Diseases **(NIAID)** through clinics sponsored by the **NEI** [45].

The documented ocular manifestations have been primarily related to the Zaire ebolavirus. Upon review of previous reports in these outbreaks, **Zaire ebolavirus (EBOV)** outbreaks in both Yambuku (1976) and Kikwit (1995) reported various ocular manifestations in the acute and convalescent stages of Ebola Virus Disease (EVD) including, subconjunctival hemorrhages, blurred vision, blindness, acute anterior uveitis, ciliary injection, photophobia, hyperlacrimation, presence of keratin precipitates, iridocyclitis, and posterior uveitis.

In a report by Kibadi[1] and Bwaka[2] et al in 1999 regarding the Kikwit outbreak in 1995 (CFR 77%) of the 20 survivors, 15% developed uveitis. PCR analysis of the fluid in one patient was positive 15 weeks after onset of EVD and 9 weeks after the initial viremia cleared demonstrating persistence of the virus in the intraocular fluid. Conjunctival hemorrhages are common; as many as 48% in Kikwit 1995 and 58% in the initial Yambuku outbreak of 1976.[3-5]

Varkey[6] et al recovered via anterior chamber paracentesis viable EBOV-Makona virus in ocular fluid from a patient 9 weeks after laboratory evidence demonstrated both blood and urine were negative for Zaire ebolavirus (EBOV) by quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) at the time of his discharge. The patient also developed color change that resolved several months after treatment[6].

In a prior outbreak of the **Bundibugyo ebolavirus (BDBV)**, Roddy et al [16] reported the acute symptoms of conjunctivitis appeared to have resolved late but completely at less than 10 days. There was no mention of a convalescent PEVDS uveitis in that report, but the musculoskeletal complaints associated with PEVDS with BDBV were a predominant, persistent symptom in that study. More recently, ocular manifestations in the 2007 Uganda outbreak of Bundibugyo ebolavirus (BDBV) in survivors was reported by Clark et al. In their study, published in Lancet, Volume 15, No. 8, p905–912, August 2015. In their report, 70 survivors of the 2007 Bundibugyo Ebola virus and 223 contacts were identified, including 49 probable and confirmed cases. 157 seronegative contacts in the outbreak were used in the comparative group. Compared to the seronegative contact group, survivors of the Bundibugyo ebolavirus outbreak were noted to be at significantly increased risk of ocular manifestations of EVD (retro-orbital pain [RR 4·3, 95% CI 1·9–9·6; p< 0·0001], blurred vision [1·9, 1·1–3·2; p=0·018]), as well as hearing loss (2·3, 1·2–4·5; p=0·010), difficulty swallowing (2·1, 1·1–3·9; p=0·017), difficulty sleeping (1·9, 1·3–2·8; p=0·001), arthralgias (2·0, 1·1–3·6; p=0·020), and various constitutional symptoms characteristic of PEVDS. Chronic health problems (prevalence ratio [PR] 2·1, 95% CI 1·2–3·6; p=0·008) and limitations due to memory loss or confusion (PR 5·8, 1·5–22·4; p=0·010) were also identified more frequently by survivors of Bundibugyo ebolavirus than in the seronegative contact group.

The only report of a confirmed **Taï Forest Ebolavirus (TAFV)** infection in humans occurred in 1994 [17]. The TAFV glycoprotein is most closely related to the Bundibugyo ebolavirus. The patient, a 34-year-old researcher investigating an EVD outbreak in chimpanzees, suddenly became ill with fever, deafness (attributed presumably to malarial prophylaxis with quinine), followed by non-bloody diarrhea, anorexia, nausea, vomiting, a non-pruritic generalized rash, temporary loss of memory, anxiety, confusion, and irritability. She recovered 6 weeks later. One month later, in the convalescent period, she developed a pronounced significant loss of elasticity of her hair and alopecia lasting approximately 3 months. As this late manifestation of alopecia is only reported in this instant case, the drug array and not necessarily the virus which caused her eventual hair loss needs to be considered. There were no ocular manifestations of EVD in this single patient with confirmed EVD due to the Taï Forest Ebolavirus (TAFV).

On review of reports in the literature of prior outbreaks caused by the **Sudan ebolavirus (SUDV)**, there is no documentation of a clinical finding suggestive of PEVDS in the convalescent stage. SUDV emerged in 1976, soon after the first outbreak of EBOV, and 6 outbreaks of SUDV have occurred since 1976 with the last known outbreak in 2014. In a report of one large outbreak from August, 2000, through January of 2001 in Uganda (Gulu, Masindi, and Mbarara districts), there were 425 presumptive cases (51% laboratory confirmed) with a case-fatality rate of calculated at 53%. Reported symptoms of EVD in that outbreak identified as caused by SUDV included headache, vomiting, anorexia, diarrhea, weakness or severe fatigue, abdominal pain, body aches or joint pains, difficulty swallowing, difficulty breathing, and hiccups. No ocular manifestations were documented.

The **Reston ebolavirus (RESTV)** is the only known non-African species of the Ebolavirus genera and is not historically known to be pathogenic in humans but is capable of causing EVD in *Macaca fascicularis,* a crab-eating macaque in the Philippines. Seven outbreaks from 1989-2009 have occurred. In 2009, 7 hog-farm workers were identified to be seropositive for RESTV infection. None became symptomatic.

PATHOGENESIS

The specific cause of the uveitis seen in survivors of EVD is not understood. Implicated in this is immune evasion, and Yang et al in 2007[7] reported the mucin-like domain (MLD) of glycoprotein (GP) is involved in shielding the viral glycoprotein in addition to its effect of causing increased vascular cell permeability[8]. Presumably, this is the mechanism by which EBOV is able to gain entry to the anterior chamber and cause uveitis.

The Ebola glycoprotein is 676 aa in length and contains GP1, the mucin-like domain (MLD), GP2-delta, and GP2. The GP1 receptor binding region (position 53-201, length 181 aa) contains the critical residues for binding at positions 57, 63, 64, 88, 95, and 170. GP1 is responsible for binding to the receptor(s) on target dendritic and endothelial cells.

GP2 (position 502-676) is classified a class I viral fusion protein and contains the short fusion peptide at position 524-539 and is responsible for penetration of the virus into the cell cytoplasm through fusion of the membrane of the endocytosed virus with the endosomal membrane of the target cell. The immunosuppressive region (584-602) been shown to down-regulate the expression of host surface molecules involved in immune surveillance and cell adhesion.

The MLD, position 301-48 (**Figure 1**) is located at the C-terminus of GP1 domain and has been shown to be both highly inflammatory and responsible for vascular cell cytotoxicity resulting in subsequent hemorrhage seen in EVD. Yang, et al was able to demonstrate the MLD was capable of inducing endothelial cell and blood vessel disruption within 48 hours of infection by the Zaire ebolavirus but not the Reston ebolavirus (RESTV). RESTV has not been shown to be pathogenic in humans despite evidence of infection. By inference, the mucin-like domain of the Zaire ebolavirus glycoprotein is the most likely domain responsible for viral entry into the anterior chamber via vascular cell disruption and subsequent inflammatory reaction.

Earlier this year we reported[9] that 16/20 nonsynonymous single nuclear polymorphisms the GP segment (303-501) I the current strain of the Zaire ebolavirus were located in the MLD (1-40 in the isolated sequence), notoriously poorly conserved amongst the several strains. GP1 and GP2 were highly conserved upon multisequence amino acid analysis of EBOV-Makona and the glycoprotein amino acid sequences of EBOV from selected prior outbreaks.

Materials and Methods

Multisequence Alignment

A DELTA-Blast of the EBOV-Makona strain full glycoprotein sequence (accession number NP_066246.1) from the current outbreak was submitted for multisequence analysis online at NCBI. This sequence was compared to selected BLAST hits from prior outbreaks the Zaire ebolavirus (Figure 4). The same sequence was used to compare the MLD with selected sequences from the Sudan ebolavirus (SUDV- accession number AGL73425.1), Bundibugyo ebolavirus (BDBV – accession number AGL73460.1), Tai Forest ebolavirus (TAFV-accession number YP_003815426.1), and the Reston ebolavirus (RESTV- accession number ACT22802.1). A multisequence analysis of these selected sequences was performed using CLUSTAL-Omega algorithm using default parameters within Jalview.

Secondary Structure Prediction

Representative sequences of each species (Zaire ebolavirus|P87671|, Zaire ebolavirus (Makona) |AHX24667.2|, Sudan ebolavirus|AAB37096.1|, Bundibugyo ebolavirus|YP_003815435|, Tai Forest ebolavirus|B8XCN9|, and Reston ebolavirus|QQ66799|) were obtained from the National Center for Biotechnology and Information database via Basic Local Alignment Search Tool BLASTp search [23, 24]. Secondary structure information and interactive modeling were then obtained by submitting the GP1 though the C-terminal amino acid sequence of the mucinlike domain (position 1-320) via the interactive web structure homology-modelling server interface at the Protein Model Portal (PMP)][25 26]. The PHYRE2 v. 2.0 [22], ModWeb v. r175 (Modeller), and Swiss-Model portals [25-34] were selected for secondary structure prediction and interactive modelling in UCSF Chimera version 10.2.

Epitope Analysis

Generally, for MHC class I T-cell epitope predictions, selection of predicted binders can be done based on the percentile rank (< = 1%) or the MHC binding affinity (IC50 < 500 nM) [16, 17] to cover most of the immune responses. Binding affinity (IC50) thresholds of 500 nM identifies peptide binders recognized by T cells which can be used to select peptides[18].

IC50 binding affinities correlate better with immunogenicity than a specific percentile rank. Paul et al has recommended that MHC-specific absolute binding affinity thresholds correlate better with immunogenicity rather than specifically using the 500

nM threshold[19].

In our study, we expanded the percentile rank score to < =2.5 so as to identify acceptable IC50 binding affinities results outside the recommended percentile rank of < = 1% using the MHC-specific absolute binding affinity threshold criteria (Table 1). We additionally excluded those epitopes whose IC50 binding affinities fell outside MHC-specific absolute binding affinity thresholds (Table 1) regardless of percentile rank. No epitopes were excluded in the filtered analysis whose IC50 binding affinity was at or below the 500 nM IC50 threshold.

Table 1 below show the allele-specific thresholds for the 38 most common HLA-A and HLA-B alleles, representative of the nine major supertypes.

A heatmap of both filtered and unfiltered comparative epitopes arrays was generated online using the Matrix2png, version 1.2.1 (February 2011) [21,22] web interface. The color bounds were set such that the minimum value associated with percentile rank and best binding affinity was red and the maximum value was colored green. Missing values in the filtered results were shaded black. Statistical analysis was performed using the single-factor ANOVA in Excel with alpha=0.05.

MCH Class I Allele Specific Affinity Cut-offs

Table 1: Alleles sorted by name

Allele	Population frequency of allele	Allele specific affinity cutoff (IC50 nM)
A*0101	16.2	884
A*0201	25.2	255
A*0203	3.3	92
A*0206	4.9	60
A*0301	15.4	602
A*1101	12.9	382
A*2301	6.4	740
A*2402	16.8	849
A*2501	2.5	795
A*2601	4.7	815
A*2902	2.9	641
A*3001	5.1	109
A*3002	5	674
A*3101	4.7	329
A*3201	5.7	131
A*3301	3.2	606
A*6801	4.6	197
A*6802	3.3	259
B*0702	13.3	687
B*0801	11.5	663
B*1402	2.8	700
B*1501	5.2	528
B*1801	4.4	732
B*2705	2	584
B*3501	6.5	348
B*3503	1.2	888
B*3801	2	944
B*3901	2.9	542
B*4001	10.3	639
B*4002	3.5	590
B*4402	9.2	904
B*4403	7.6	780
B*4601	4	926
B*4801	1.8	887
B*5101	5.5	939
B*5301	5.4	538
B*5701	3.2	716
B*5801	3.6	446

MHC class II

For MHC class II T cell epitope predictions, selection of predicted binders has been done based on both the percentile rank and/or MHC binding affinity. The IEDB currently recommends making selections based on a consensus percentile rank of the top 10% as well as selecting peptides supported by experimental data5 predicted to bind at 1,000nM. In our analysis of the MHC class II epitopes, we used sequence in each series with the best percentile rank. Binding affinities of >1000 nM were excluded in the filtered analysis. A heatmap of both filtered and unfiltered comparative epitopes arrays was generated online using Matrix2png, version 1.2.1 (February 2011) [21, 21] web interface. The color bounds were set that the minimum value associated with percentile rank was red and minimum value was colored green. Missing values in the filtered results are shaded black. Statistical analysis was performed using the single-factor ANOVA in Excel with alpha=0.05.

Modeling

The major difficulty with modeling the MLD is glycosylation, the most common of post-translational protein modifications, is that the heterogeneity introduced by glycosylation can hinder crystallization of a glycoprotein thus limiting the use of X-ray crystallography, one of the major techniques using in protein structure solution[41, 42]. Consequently, only the N-terminus of the MLD is included as reference in this report.

Interactive modeling was obtained though the protein model portal [43] with 3VE0 and 3CSY used as the primary templates. The resulting PDB files were converted to PQR through the PDB2PQR, version 2.0.0 web interface [36, 37, 38, 39] using default parameters. Protonation states were assigned using PROPKA. The electrostatic properties were simulated with APBS 1.4.1 locally in UCSF Chimera version 10.2. The DX files resulting from APBS were used for the visualization. The residues E1 epitope in the NH2-terminus of the mucin-like domain along with the COOH-terminus of GP1 ($^{295-320}$) were isolated in UCSF Chimera, stable release 1.10.2 [40].

Results

ultisequence Analysis

The Zaire ebolavirus (tax ID 186538) GP1 segment representing the receptor binding region (position 1-304) was highly conserved within the EBOV taxonomy group (**Table 2**) in 130 hits with >90% coverage, ranging from 81-100%. Bundibugyo ebolavirus (BDBV-tax ID 186538) had 6 hits with >90% coverage with a percent identity of 78-79% with EBOV GP1 segment. The Sudan ebolavirus (SUDV-tax ID 186540) yielded 16 BLAST hits with >90 coverage, but was poorly conserved compared to EBOV at 67%. Similarly, the Tai Forest Ebolavirus

(TAFV-tax ID 186541) resulted in only 5 hits, with a percentage identity at 78% with >90% coverage. Reston ebolavirus (RESTV-tax ID 186539) resulted in 17 BLAST hits with over 90% coverage and a percent identity of 68-71%.

The *Zaire ebolavirus* (tax ID 186538) GP2 segment (position 502-676) containing GP2-delta, fusion peptide, and through the immunosuppressive C-terminus was highly conserved within the Zaire ebolavirus taxonomy group, with 130 BLAST hits and >90% coverage the range was 77-100%.

Bundibugyo ebolavirus (BDBV-tax ID 186538) had a limited 3 hits with >90% coverage and a percent identity of 90% with EBOV GP2 query.

The *Sudan ebolavirus* (SUDV-tax ID 186540) yielded 11 BLAST hits with >90 coverage, but was significantly less conserved compared to the BDBV-EBOV conservation at 78%.

Tai Forest Ebolavirus (TAFV-tax ID 186541) resulted in only 2 hits, with a percentage identity at 87% with >90% coverage. The Reston ebolavirus (RESTV-tax ID 186539) resulted in 7 BLAST hits with over 90% coverage and a percent identity of 82-84%.

The *Zaire ebolavirus* (tax ID 186538) segment representing the mucin-like domain of interest in this study up to the GP2 sequence (position 301-501) was submitted as the query for DELTA-BLAST analysis and again filtered as to each individual species using the default parameters in the CLUSTAL-Omega online BLAST program at NCBI (**Figure 4**). Within EBOV at 90% coverage, the MLD query resulted in 130 hits, and resulted in a percentage identity ranging from 86-100%. Bundibugyo ebolavirus (BDBV-tax ID 186538) had 11 BLAST hits with >90% coverage and was poorly conserved with a percent identity of only 25-27%. The *Sudan ebolavirus* (SUDV-tax ID 186540) yielded 11 BLAST hits with only 21-22% identity against EBOV. *Tai Forest ebolavirus* (TAFV-tax ID 186541) resulted in only 3 hits, with a percentage identity of only 29%. Finally, the *Reston ebolavirus* (RESTV-tax ID 186539) resulted in 23 BLAST hits with over 90% coverage and was also poorly conserved with respect to EBOV, ranging from 23-31% identity.

TABLE 2 DELTA-BLAST results of the individual domains within all species of Ebolavirus

GP1		
Taxon ID	**BLAST hits**	**% Identity**
Zaire ebolavirus [186538]	130	86-100%
Bundibugyo ebolavirus [565995]	6	78-79%
Tai forest ebolavirus [186541]	5	78%
Sudan ebolavirus [186540]	16	67%
Reston ebolavirus [186539]	17	68-71%
GP2		
Taxon ID	**BLAST hits**	**% Identity**
Zaire ebolavirus [186538]	130	77-100%
Bundibugyo ebolavirus [565995]	6	90%
Tai forest ebolavirus [186541]	2	87%
Sudan ebolavirus [186540]	11	78%
Reston ebolavirus [186539]	7	82-84%
Mucin-like Domain (MLD)		
Taxon ID	**BLAST hits**	**% Identity**
Zaire ebolavirus [186538]	130	86-100%
Bundibugyo ebolavirus [565995]	11	25-27%
Tai forest ebolavirus [186541]	3	29%
Sudan ebolavirus [186540]	11	21-22%
Reston ebolavirus [186539]	23	23-31%

Secondary Structure Prediction

The secondary structure of the GP1 segment from position 4-300 (**Figure 5)** demonstrated significant similarity. Protein disorder overall was the least in the Zaire ebolavirus (18%) compared to all other species (23%). RESTV was predicted to have the greatest protein disorder at 23%. The structural α-helices overall represented 15-18% of the submitted sequence and β-strand percentage was similar between the species, ranging from 38-43%. Consistent α-helices were noted in regions 1-30, 50-55, 88-90, 190-195, and 258-262. The β-strands were predicted at regions 35-40, 60-72, 95-110 (two), and 2 highly conserved β-strand segments spanning the regions 135-185 and 210-245. Two additional conserved β-strands were predicted at 270-280 and 285-290 in all species at the C-terminus of GP-1. Protein disorder was consistently identified at from position 1-10 and variably within the 45-60 position in GP1.

As expected, protein disorder prediction was greatest within the C-terminus GP1-mucin-like domain transition from position 260-320. Two conserved β -strands at regions 300-315 and 320-330 within the N-terminus of the mucin-like domain were identified. The secondary structure of the E1 epitope from 295-300 contained an α-helix or β-strand secondary structure of uncertain reliability as this also was located within the region of predicted high protein disorder.

Within GP1 (**Figure** 4) , the 15 residues (**Table 3**) previously reported by Jeffers et al in 2002 that are involved in cellular entry and GP2 expression are highly conserved in the virulent species (86.7%) . In RESTV 86.7% are also conserved as to sequence. At position 64 (cellular entry), there is a Lysine substitution in all species. At Lys-95 in EBOV, there is a Glutamine substitution in SUDV. At Val-170 in EBOV, an Ile-170 substitution is identified with BDBV and TAFV. At position Val-204 in EBV, several residues have been substituted in RESTV (Met-204) and in BDBV and TAFV (Ala-204). At position 296, only one sequence (EBOV) identified a nonsynonymous substitution (Thr-296). All five of these positions are

involved with cellular entry but not GP2 expression. Despite the nonsynonymous mutation, the secondary structure is conserved in all species. Only position 296 is involved in the mucin-like domain. The critical residues were 100% conserved.

Figure 4. Multiple Sequence Alignment of GP1

Table 3. Identification of Covalent Modification between the Various Species of Ebolavirus in GP1

RESIDUE POSITION	PROCESS	CONSERVATION
Asn-40	Loss of cellular entry	100% conserved; β-strand conserved
Cys-53	Loss of cellular entry	100% conserved; α-helix conserved
Leu-57	Loss of cellular entry	100% conserved; β-strand
Leu-63	Loss of cellular entry	100% conserved; β-strand
Arg-64->Lys-64	Loss of cellular entry	SUDV, BDBV, TAFV, and RESTV Lys-64 substitution; β-strand
Phe-88	Loss of cellular entry	100% conserved; α- helix
Lys-95->Gln-95	Loss of cellular entry	SUDV substitution Gln-95
Cys-108	Loss of cellular entry	100% conserved
Leu-121	Loss of cellular entry	100% conserved
Cys -135	GP2 expression loss; Loss of cellular entry	100% conserved
Cys -147	Loss of cellular entry	100% conserved
Val-170->Ile-170	Loss of cellular entry	BDBV, TAFV substitution; β-strand structure conserved
Val-204->Ala-204,Met-204,	Loss of cellular entry	RESTV substitution (Met-204); SUDV substitution (Ala-204)
Asn -257	Loss of cellular entry	100% conserved; α- helix
Asn-296-Thr-296	Loss of cellular entry	Thr-296 (single EBOV sequence); α- helix
Reference: *"Covalent modifications of the ebola virus glycoprotein." Jeffers S.A., Sanders D.A., Sanchez A.J. Virol. 76:12463-12472(2002)*[35]		

Epitope Analysis

B-Cell Epitopes

All epitopes from the IEDB were obtained by separately querying with EBOV, SUDV, BDBV, TAFV, and RESTV. A total of 106 epitopes were obtained spanning GP1 (1-301), the MLD (301-492), and GP2 (495). Sixteen linear epitopes (**Table 4**) including flanking sequences from the mucin-like domain (301-495) were obtained across all species from human hosts.

Table 4. B-cell Epitopes within the Zaire ebolavirus mucin-like domain

Epitope ID	Object Type	Description	Starting Position	Ending Position
28398	Linear peptide	IRSEELSFTVVSNGA	301	315
227646	Linear peptide	TVVSNGAKNISGQSP	309	323
227409	Linear peptide	NISGQSPARTSSDPG	317	331
52394	Linear peptide	QSPARTSSDPGTNTT	321	335
187509	Linear peptide	RTSSDPGTNTTTEDH	325	339
227099	Linear peptide	DPGTNTTTEDHKIMA	329	343
227422	Linear peptide	NTTTEDHKIMASENS	333	347
187314	Linear peptide	EDHKIMASENSSAMV	337	351
27448	Linear peptide	IMASENSSAM**V**Q**V**HS	341	355
227141	Linear peptide	ENSSAMVQVHSQGRE	345	359
227307	Linear peptide	KPGPDNSTHNTPVYK	381	395
187300	Linear peptide	DNSTHNTPVYKLDIS	385	399
227248	Linear peptide	HNTPVYKLDISEATQ	389	403
227687	Linear peptide	VYKLDISEATQVEQH	393	407
187296	Linear peptide	DISEATQVEQHHRRT	397	411
187541	Linear peptide	TGEESASSGKLGLIT	469	483

A pairwise multisequence analysis of selected sequences from prior outbreaks from 1976-2015 (n=38) of EBOV (n=18), SUDV (n=9), BDBV (n=6), TAFV (n=2), and RESTV (n=3) were utilized. The overall percent identity with respect to EBOV (mean-97.75%) within the mucin like domain was not significantly different with respect to SUDV (mean=55.7%), BDBV (mean=65.6%), TAFV (mean=64.6%), and RESTV (mean=57.84%) reflecting the overall poor conservation across the MLD as compared to GP1 and GP2 (**Table 2**).

The individual epitopes varied significantly as compared to EBOV-Makona (**Table 4**) The RESTV epitope conservation with respect to EBOV was 15.62% (0-38.7%), SUDV 21.43 (5.6-51.2%), BDBV (0-40%), and TAFV (0-46.67%). E1 epitope (IRSEELSFTVVSNGA, 301-316) had the highest conservation at 40.57%, followed by E10 (ENSSAMVQVHSQGRE, 32.3%), E8 (EDHKIMASENSSAMV, 29.43%), E9 (IMASENSSAMVQVHS, 28.23%), and E16 (TGEESASSGKLGLIT, 24.7%). The conservation from E8-E10 is due to the overlapping residues spanning the epitopes.

Table 4. B-cell epitope analysis overall identities with Zaire ebolavirus

EPITOPE	EBOV	SUDV	BDBV	TAFV	RESTV	Mean	Median	STDEV
IRSEELSFTVVSNGA	94.87	51.20	40.00	46.67	24.40	40.57	46.67	24.884
TVVSNGAKNISGQSP	93.20	16.20	25.45	6.67	38.73	21.76	25.45	34.069
NISGQSPARTSSDPG	93.00	12.60	20.00	0.00	34.27	16.72	20	36.299
QSPARTSSDPGTNTT	95.80	12.60	26.67	6.67	28.87	18.70	26.67	35.721
RTSSDPGTNTTTEDH	93.67	19.20	26.67	26.67	4.44	19.25	26.67	34.499
DPGTNTTTEDHKIMA	93.67	5.60	20.00	33.30	0.00	14.73	20	37.612
NTTTEDHKIMASENS	96.20	18.07	25.87	46.67	6.67	24.32	25.87	35.305
EDHKIMASENSSAMV	99.00	31.87	19.20	53.30	13.33	29.43	31.87	34.685
IMASENSSAMVQVHS	98.33	32.06	25.87	35.00	20.00	28.23	32.06	31.878
ENSSAMVQVHSQGRE	97.40	37.87	31.53	28.13	33.33	32.72	33.333	29.14
KPGPDNSTHNTPVYK	96.20	6.67	0.00	6.67	0.00	3.34	6.67	41.664
DNSTHNTPVYKLDIS	98.50	19.80	0.00	6.67	0.00	6.62	6.67	41.879
HNTPVYKLDISEATQ	99.00	20.00	6.67	6.67	6.67	10.00	6.67	40.217
VYKLDISEATQVEQH	92.40	13.13	6.67	0.00	6.67	6.62	6.67	38.643
DISEATQVEQHHRRT	97.33	13.13	6.67	0.00	6.67	6.62	6.67	40.833
TGEESASSGKLGLIT	97.87	32.80	26.62	13.30	25.90	24.66	26.62	33.5

However, there appears to be a greater conservation (**Figures 6 and 7**) of the R(X) EELSF motif in E1 within the pathogenic species as compared to RESTV. The C-terminus of E1 is highly charged with Arg-302 conserved in SUDV and the negatively charged Glu-304, Glu-305 tandem residues. Overall E1 conservation as compared to EBOV-Makona ranged from 51.2% in SUDV, 40% (BDBV), 46.67% TAFV) to 24.4% in RESTV. The EELSF motif is specifically conserved within in the pathogenic species (100%) In RESTV, only Glu-304 and Lys- 306 are conserved within that motif (25%).

From E2-E4, both EBOV and RESTV share a QSPA motif not significantly conserved with SUDV, BDBV, or TAFV. It is unlikely, therefore, that this region of conservation between EBOV and RESTV is a contributing factor to differential virulence. The high conservation shown by RESTV in E5 is due to the overlapping of the QSPA motif in E4.

From E5-E8, there is another relatively well conserved, charged tandem grouping of residues in the motif Glu-336, Asp-337, and His-338 as compared to RESTV. In the pathogenic species, Glu-336 is replaced by Arg-338 in SUDV and Lys-338 in BDBV. RESTV contain no charged residues within that motif. Asp-337 is replaced by Lys-337 in SUDV. The high conservation shown by RESTV in E5 is due to the overlapping of the QSPA motif in E4. RESTV

conservation falls significantly (0-13.3%) between E5-E8.

From E9-E10, BDBV and EBOV share the same VQV(X4)RE motif. The conservation between pathogenic species and RESTV is similar (28-32%) and may not contribute to the differential virulence.

E11-15 are also poorly conserved (3.3-10%) between all species with respect to EBOV and, therefore, probably do not contribute to the differential virulence.

In E16, at the junction of the MLD and GP2, the GLIT motif at the C-terminus of E16 is highly conserved between all species and probably also does not contribute to the differential virulence.

MHC Class I Analysis.

The unfiltered data of the MHC Class II was analyzed using ANOVA in Excel using the percentile rank as the query as previously discussed. The percentile rank correlates with the IC50 to determine which epitope is predicted to be the best binder. The reference 167 epitope sequences were from the Zaire ebolavirus, strain Mayinga (1976). There was no significant variance between the species (*DF=166; F=1.16; F_{crit} =1.21); p=0.067*). Figure 8 shows the heatmap of the filtered percent rank scores (rank score < 2.5) in the aligned sequences.

MHC Class II Epitope Analysis

The unfiltered data of the MHC Class II was analyzed using ANOVA in Excel using the percentile rank as the query. The percentile rank correlates with the IC50 to determine which epitope is predicted to be the best binder. The reference 168 epitope sequences were from the Zaire ebolavirus, strain Mayinga (1976). The overall variation was significant between the species (*DF=166; F=2.06; F_{crit}=1.22) with P< .05 (P=1.16E-10*). The mean rank score of each predicted rank score was 18.5(range 1.6-37.5) with a median rank score calculated at 19.1.

The regions predicted to have the greatest variability (mean=>18.5) between all species were from residues VSNGAKNISGQSPAR-> TTEDHKIMASENSSA, HSQGREAAVSHLTTL-EAAVSHLTTLATIST, TISTSPQSLTTKPGP-TSPQSLTTKPGPDNS, VYKLDISEATQVEQH-TQVEQHHRRTDNDST, and the COOH-terminal residues TDNDSTASDTPSATT->GEESASSGKLGLITN.

A separate ANOVA of the virulent species was also significant (F=1.46, Fcrit=1.22; P=.000657). Twenty two epitopes were identified where the variance was least in the NH2-terminus of the MLD, particularly in the regions EELSFTVVSNGAKNIS->LSFTVVSNGAKNISG and EDHKIMASENSSAMV->MVQVHSQGREAAVSH (mean variance =5.12). One epitope was excluded (ASENSSAMVQVHSQG) as was felt to be out of range (single amino acid variance=267) compared to the adjacent epitope variance. The epitope sequences PDNSTHNTPVYKLDI-STHNTPVYKLDISEA (mean variance=13.4) and TTLATISTSPQSLTT->LATISTSPQSLTTKP (mean variance=14.4) were additional regions identified to be of low variability between the virulent species.

Significant variance was also identified between EBOV and RESTV and was analyzed separately (F=1.69, Fcrit=1.29, P=.0004) at p< .05.

An ANOVA was then performed between the virulence species and RESTV. A significant variance was identified between the twenty two epitopes shared within the virulent group and RESTV using the rank score criteria (*F=0.99, F_{crit}=2.06; p=0.51*).Three epitopes of acceptable binding affinities in the virulent group were significantly different from RESTV (p>.05) (Table 5).

Epitope Sequence	*Variance*	*EBOV Rank Score*	*RESTV Rank Score*
EELSFTVVSNGAKNIS	0.01445	0.42	0.25
LSFTVVSNGAKNISG	0.06845	0.63	0.26
EDHKIMASENSSAMV	36.55125	2.16	10.71
DHKIMASENSSAMVQ	2.8322	2.16	4.54
HKIMASENSSAMVQV	2.645	2.16	4.46
KIMASENSSAMVQVH	1.9208	2.16	4.12
IMASENSSAMVQVHS	2.1632	2.27	4.35
MASENSSAMVQVHSQ	3.40605	4.47	7.08
SENSSAMVQVHSQGR	128.8013	0.72	16.77
ENSSAMVQVHSQGRE	0.01445	0.72	0.55
NSSAMVQVHSQGREA	0.01125	0.7	0.55
SSAMVQVHSQGREAA	0.0098	0.69	0.55
SAMVQVHSQGREAAV	0.0128	0.71	0.55
AMVQVHSQGREAAVS	2.02005	2.56	0.55
MVQVHSQGREAAVSH	0.55125	2.65	1.6
TTLATISTSPQSLTT	93.5712	1.74	15.42
TLATISTSPQSLTTK	145.6925	2.74	19.81
LATISTSPQSLTTKP	0.41405	2.97	2.06
ATISTSPQSLTTKPG	471.5521	32.24	1.53
PDNSTHNTPVYKLDI	0.3872	7.88	7
DNSTHNTPVYKLDIS	0.08405	10.27	9.86
NSTHNTPVYKLDISE	136.6205	22.62	6.09

Table 5 Comparative Analysis of MHC II Binding Predictions between Virulent and RESTV species in Ebolavirus

Twenty two epitopes (length=14) from the virulent species of Ebolavirus (EBOV, SUDV, BDBV, and TAFV) ound to have significantly low variance were compared using single-factor ANOVA to the rank score binding capabilities of RESTV. The reference sequences were from Zaire ebolavirus, strain Mayinga, 1976. Strong MHC II binding capacity was identified in 4 epitopes shared with the virulent species (shaded green; rank score range 0.72-2.74) but not predicted to have strong IC50 binding capacity in RESTV (rank score 10.71-19.81)). In contrast, the epitope ATISTSPQSLTTKPG (shaded yellow) was predicted to have low binding capability (32.24) as compared to RESTV (rank score =1.53). One epitope of significant variance (shaded blue) is predicted to be of low binding capacity in all species and probably does not contribute to pathogenesis.

The epitopes EDHKIMASENSSAMV, SENSSAMVQVHSQGR, and TTLATISTSPQSLTT, and TLATISTSPQSLTTK are predicted to have good binding capabilities (rank score range 0.72-2.74), but RESTV does not (rank score range 10.71-19.81) The epitope ATISTSPQSLTTKPG correlated within the virulent group was not predicted to have good MHC class II binding capabilities (rank score=32.24) but the corresponding epitope in RESTV is predicted to strongly bind MHC II molecules (rank score=1.53). This additionally raises the question of molecular mimicry within the virulent species as compared to RESTV. The heat map of the entire MLD MHC Class II epitope analysis is shown in Figure 8.

Interactive Modeling

All models generated for positions 1-320 were modelled at >90% accuracy involving at least 249 residues covering >75% of all selected templates (3CSY and 3VE0) with verified PDB documentation.

The isolated E1 epitope, IRSEELSFTVVSNGA (EBOV-Makona, is located within a β-strand from residues 301-315 in EBOV, BDBV, and TAFV. The residues of the C-terminus of the GP1 segment from Lys-294 through Lys-295 are predicted to lie within a short α-helix and contribute significantly to positive charged surface zone in four virulent species but are less conspicuous in RESTV. The E1 epitope, in contrast, is predominately negatively charged as a result of the tandem glutamic acid residues at GLU-304 and GLU-305 and neighboring residues in the virulent species as compared to RESTV. Total predicted charge across the E1 epitope sequence in each were obtained using a 35 amino acid sequence and submission to PDB2PQR [GEWAFWETKKNFSQQLHGENLHFQILSTHTNNSS (RESTV), GEWAFWENKKNFTKTLSSEELSFVPVPETQNQVLD (TAFV), GEWAFWENKKNFTKTLSSEELSVIFVPRAQDPGSN (BDBV), GEWAFWENKKNLSEQLRGEELSFETLSLNETEDDD (SUDV), and GEWAFRETKKNLTRKIRSEELSFTAVSNGPKNISG (EBOV-Makona)].

Secondary tertiary structures were obtained as previously described with the final total charges assigned to that segment obtained from the PQR output file (**Table 6**).

Over the entire E1 epitope, the charges ranged from -2.4 (SUDV) to -0.6 (BDBV). As expected, the majority of the negative charge was across the EELSF motif, ranging from -1.0 (RESTV) to -2.0 (EBOV, BDBV), suggesting a great binding capability attributed to the additional Glu residue in the virulent species (**Table 7**).

Table 6 E1 B-cell E1 Predicted Electrostatic Charges

ID	Sequence	Total charge
EBOV-Makona	IRSEELSFTAVSNG	-1.0
SUDV	LRGEELSFETLSLN	-2.4
BDBV	LSSEELSVIFVPRA	-0.6
TAVF	LSSEELSFVPVPVP	-2.0
RESTV	LHGENLHFQILSTH	-1.0

Table 7. E1 B-cell E1 Motif Electrostatic Charges

ID	Motif Sequence	Total charge
EBOV-Makona	EELSF	-2.0
SUDV	EELSF	-1.6
BDBV	EELSV	-2.0
TAVF	EELSF	-1.6
RESTV	ENLHF	-1.0

To examine the contributory effect of neighboring residues around the E1 epitope, the amino acid sequence from 250-310 in each species was selected based upon the predicted model for each. The global charge for the region approximating the E1 B-cell epitope of the mucin-like domain appears to be neutral in the lesser virulent species (BDBV, TAFV, and RESTV), relatively negatively charged for EBOV-Makona, and positively charged for SUDV. This suggests a greater antigenic potential for EBOV and SUDV compared to BDBV, TAFV, and RESTV (FIGURE 9).

Discussion

The Post-Ebola Virus Disease Syndrome (PEVDS) has come to the forefront of many researchers following the recent two year Zaire ebolavirus devastating outbreak in West Africa. At the time of this report, Guinea, Liberia, and Sierra Leone, the countries most affected during the outbreak, have been declared free of transmission after 42 days of no new cases. With the transition to the dry season in the early spring of 2016, the spot reports of virus persistence in privileged bodily fluids, and the discrepancy between the case-fatality rates (CFR) associated with the reported number of deaths during the outbreak it is quite expected we will see reemergence of the active disease in West Africa. In that most of the directed therapeutic strategies are directed towards blocking viral entry through GP1/GP2 inhibition, that approach may not be effective in preventing the cytokine storm at higher viral loads. As the mucin-like domain has already been shown to be responsible for many of the immune-related symptoms, this study was designed to investigate this region.

This study has potentially identified one B-cell linear epitope, IRSEELSFTAVSNG that contains a highly conserved EELSF, negatively charged, motif with the

virulent species. RESTV in comparison has only 24% homology with this predicted epitope. Theoretical homology modeling in this study also appears to show a considerable contribution to the surface potential surrounding the E1 B-cell epitope by the surrounding residues. The highly charged Lys-296 within the C-terminus of GP1 appears to be prominently located near the mucin-like domain and could contribute to the shielding effect of the MLD.

Bioinformatic analysis demonstrated four MHC Class II T-cell epitopes with statistically significant high binding affinities (range=0.72-2.16) associated with the four virulent species identified as EDHKIMASENSSAMV, SENSSAMVQVHSQGR, TTLATISTSPQSLTT, and TLATISTSPQSLTTK within the mucin-like domain. The corresponding binding affinity to these epitopes in RESTV however, was predicted to have a statistically significant low binding potential (range=10.71-19.82) when compared to the virulent species. There was no significantly significance within the MHC Class I epitopes between the virulent and nonvirulent species of Ebolavirus (p=0.067).

Another consideration not specifically addressed in this paper is in regards to those epitopes that have shared low-binding affinities in the virulent groups as compared to RESTV. In this study, we investigated this potential in MHC II exogenous processing of the glycoprotein mucin-like domain to CD4 T-cell recognition. This pathway may be more important as Ebola preferentially attacks macrophages and dendritic cells. Here we identified one potential candidate sequence, ATISTSPQSLTTKPG , located in the C-terminal region of the mucin-like domain, with low binding affinity (rank score=32.2) as compared to the corresponding RESTV epitope (rank score=1.53) and could represent potential molecular mimicry. Molecular mimicry by Ebolavirus is potentially one mechanism by which viral persistence could be maintained through immune tolerance. To evaluate this potential, the binding potential of the MCH I and MHC II was obtained by submitting the amino acid sequence of the mucin-like domain to the IEDB database and filtered to those with high binding affinities. Those regions of low binding affinity could be regions expressing peptide sequences with homologies not recognizable as foreign thereby partially and/or completely evading the immune response leading to viral persistence. Those epitopes of high binding affinities would be recognized by MHC I class molecules as foreign endogenous antigens, presented to cytotoxic CD8 T-cells and eliminated. In the consideration of differential virulence, a shared epitope of the virulent species recognized by MHC I or class II but not demonstrating a similar high binding affinity could by inference partially explain the differential virulence in Ebola Virus Disease expression as compared to RESTV. Conversely, persistence resulting in PEVDS could be explained by a lower binding affinity by one species over another, i.e., molecular mimicry.

In consideration of the homology between all species within GP1 and GP2 and the critical interacting residues within GP1 as reported by Jeffers, et al, these B-cell and T-cell epitopes within the mucin-like domain may ultimately be demonstrated to be responsible clinical findings in PEVDS. Laboratory confirmation in this study will be required to validate these findings. If future studies ultimately are shown to be significantly associated with the clinical findings with PEVDS, therapeutic strategies may then be more specifically directed towards the mucin-like domain.

References

1. Kibadi K1, Mupapa K, Kuvula K, Massamba M, Ndaberey D, Muyembe-Tamfum JJ, Bwaka MA, De Roo A, Colebunders R.; Late ophthalmologic manifestations in survivors of the 1995 Ebola virus epidemic in Kikwit, Democratic Republic of the Congo.J Infect Dis. 1999 Feb;179 Suppl 1:S13-4
2. Bwaka MA1, Bonnet MJ, Calain P, Colebunders R, De Roo A, Guimard Y, Katwiki KR, Kibadi K, Kipasa MA, Kuvula KJ, Mapanda BB, Massamba M, Mupapa KD, Muyembe-Tamfum JJ, Ndaberey E, Peters CJ, Rollin PE, Van den Enden E, Van den Enden E.Ebola hemorrhagic fever in Kikwit, Democratic Republic of the Congo: clinical observations in 103 patients. J Infect Dis. 1999 Feb;179 Suppl 1:S1-7
3. Moshirfar M, Fenzl CR, Li Z. What we know about ocular manifestations of Ebola. Clinical Ophthalmology (Auckland, NZ). 2014;8:2355-2357. doi:10.2147/OPTH.S73583.
4. Enzo Maria Vingolo, Giuseppe Alessio Messano, Serena Fragiotta, Leopoldo Spadea, and Stefano Petti, “Ocular Manifestations of Ebola Virus Disease: An Ophthalmologist’s Guide to Prevent Infection and Panic,” BioMed Research International, Article ID 487073, in press.
5. S Wiwanitkita,V Wiwanitkit. Ocular problem in Ebola virus infection: A short review. Saudi Journal of Ophthalmology, Volume 29, Issue 3, July–September 2015, Pages 225–226
6. Jay B. Varkey, M.D., Jessica G. Shantha, M.D., Ian Crozier, M.D., Colleen S. Kraft, M.D., G. Marshall Lyon, M.D., Aneesh K. Mehta, M.D., Gokul Kumar, M.D., Justine R. Smith, M.B., B.S., Ph.D., Markus H. Kainulainen, Ph.D., Shannon Whitmer, Ph.D., Ute Ströher, Ph.D., Timothy M. Uyeki, M.D., M.P.H., M.P.P., Bruce S. Ribner, M.D., M.P.H., and Steven Yeh, M.D.Persistence of Ebola Virus in Ocular Fluid during Convalescence. N Engl J Med 2015; 372:2423-2427June 18,

2015DOI: 10.1056/NEJMoa150030
7. Yang, Z. Y. et al. Identification of the Ebola virus glycoprotein as the main viral determinant of vascular cell cytotoxicity and injury. Nature Med. 6, 886-889
8. Martinez, O et al. Impact of Ebola Mucin-Like Domain on Antiglycoprotein Antibody Responses Induced by Ebola Virus-Like Particles. The Journal of Infectious Diseases Volume 204, Issue suppl 3Pp. S825-S832
9. RA Ricketson, LA Roberts. Pitfall in Screening for Ebola Virus Disease: The Variaable Febrile and Human Subjective Response. Published online https://www.ecronicon.com/ecmi/pdf/ECMI-01-00005.pdf January 31, 2015
10. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191
11. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. "UCSF Chimera - A Visualization System for Exploratory Research and Analysis." J. Comput. Chem. 25:1605-1612 (2004).
12. N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan D. Eramian, M. Shen, U. Pieper, A. Sali. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, 2006.
13. M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.
14. A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.
15. A. Fiser, R.K. Do, & A. Sali. Modeling of loops in protein structures, Protein Science 9. 1753-1773, 2000
16. Clinical Manifestations and Case Management of Ebola Hemorrhagic Fever Caused by a Newly Identified Virus Strain, Bundibugyo, Uganda, 2007–2008; PLOS One; December 28, 2012DOI: 10.1371/journal.pone.0052986
17. Formenty et al ; http://jid.oxfordjournals.org/content/179/Supplement_1/S48.long
18. Davis SJ, Crispin M. Solutions to the Glycosylation Problem for Low- and High Throughput Structural Glycoproteomics. Owens RJ, Nettleship JE, editors. Functional and Structural Proteomics of Glycoproteins2011. p. 127-58
19. Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. 2008. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol. 4(4):e1000048.
20. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B. 2010. Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinformatics. 11:568.
21. The role of the proteasome in generating cytotoxic T cell epitopes: Insights obtained from improved predictions of proteasomal cleavage. M. Nielsen, C. Lundegaard, O. Lund, and C. Kesmir. Immunogenetics., 57(1-2):33-41, 2005
22. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. 2015 Jun ;10(6):845-58. Doi: 10.1038/nprot.2015.053. Epub 2015 May 7. PubMed PMID: 25950237
23. http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins
24. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410. PubMed
25. Haas J., Roth S., Arnold K., Kiefer F., Schmidt T., Bordoli L. and Schwede T.(2013). The Protein Model Portal - a comprehensive resource for protein structure and model information.
26. http://www.proteinmodelportal.org/?pid=modelling_interactive
27. https://modbase.compbio.ucsf.edu/modweb/
28. http://swissmodel.expasy.org/
29. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res, 25, 3389-3402.
30. Remmert, M., Biegert, A., Hauser, A. and Soding, J. (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods, 9, 173-175.
31. Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis, 18, 2714-2723.
32. Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol, 234, 779-815.
33. Benkert, P., Biasini, M. and Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics, 27, 343-350.
34. Mariani, V., Kiefer, F., Schmidt, T., Haas, J. and Schwede, T. (2011) Assessment of template based protein structure predictions in CASP9. Proteins, 79 Suppl 10, 37-58.
35. Jeffers SA, Sanders DA, Sanchez A. Covalent modifications of the ebola virus glycoprotein. J Virol. 2002 Dec;76(24):12463-72. PubMed PMID: 12438572; PubMed Central PMCID: PMC136726.
36. http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0/
37. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Research 32 W665-W667 (2004)
38. J.W. Ponder and D.A. Case. Force fields for protein simulations. Adv. Prot. Chem. 66, 27-85 (2003)
39. D.A. Case, J.T. Berryman, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L.

Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman (2015), AMBER 2015, University of California, San Francisco.
40. https://www.cgl.ucsf.edu/chimera
41. Davis SJ, Crispin M. Solutions to the Glycosylation Problem for Low- and High Throughput Structural Glycoproteomics
42. Owens RJ, Nettleship JE, editors. Functional and Structural Proteomics of Glycoproteins2011. p. 127-58
43. http://www.proteinmodelportal.org/].
44. http://www.who.int/features/2014/post-ebola-syndrome/en/].
45. http://www.nih.gov/news/health/jun2015/niaid-17.htm