

Imaging of High and Low Resolution Ebola Envelope GP Structures Composited with *in silico* Models of Difficult-to-Resolve Sections

Garry W Lynch^{1,2*}, Stefanie S Portelli¹, Siu Wai Wong¹, Maria Rocha Costa², Ana Paula Mello Lemgruber², Peter Williamson², John S Sullivan^{1,5}, Robert Booy^{1,4} and W Bret Church³

¹Faculty of Medicine, The University of Sydney, Camperdown, NSW, 2006, Australia

²Faculty of Veterinary Science, The University of Sydney, Camperdown, NSW, 2006, Australia

³Faculty of Pharmacy, The University of Sydney, Camperdown, NSW, 2006, Australia

⁴The National Centre for Immunosurveillance Research, The Children's Hospital Westmead, Westmead, NSW, 2114, Australia

⁵School of Medical Sciences, UNSW, Kensington, NSW 2052, Australia

Abstract

The Ebola surface glycoprotein, GP, facilitates receptor binding, cell infection and cytopathology, and is a principal target for immune protection. The structure of the GP central core structure and antibody target for the 1976 Ebola Zaire Mayinga strain of Ebola has been resolved by X-ray crystallography (e.g., pdb-ID: 3CSY). But other important GP regions have defied crystal structure determination. These include the apical mucin-like domain (MLD) that contains immuno-protective sites and is most distal to the membrane, and the membrane proximal C terminus region (meD) where GP is tethered to the Ebola surface. A molecular structure-based strategy in vaccine design necessitates detailed fine-structure knowledge of exposed sites for immuno-targeting. To address the lack of MLD and meD structures we have performed computational modeling of those regions using the programs iTasser, Phyre2 and CABS-flex. Candidate MLD model structures for the 1976 Mayinga strain were screened for continuity and fit with the 3CSY core and tested further for 3D compliance for its fit within the spatial boundaries of a low-resolution GP trimer cryoelectron tomograph (EMDB-ID: EMD-6003). Under these constraints only 1 of 6 initial MLD models generated by iTasser or Phyre2 met the required structural criteria of continuity, orientation and spatial fit. That best-fit structure (MLDm1p1) was additionally evaluated using the MolProbity geometric analysis, as well by the generation of highly similar and compatible structures (e.g., MLDm2c2) by the CABS-flex program. Notably, documented protective antibody sites of the MLD mapped to the MLDm1p1 apex surface, as anticipated for domain functionality or exposure to antibody attack. These studies describe best fit proximate *in silico* models of the MLD and meD regions of GP to assist molecular understanding and therapy design. Biochemical and immunologic validation and refinements of the models are continuing, to establish to their biostructural compatibilities and value to inform on GP structure-function and immuno-target potential.

Introduction

The 2014 Ebola Virus (EBOV) outbreak in West Africa had up to the end of March 2015 claimed more than 10,236 lives from 24,753 infections [1]. Now more than a year on from its emergence the crisis appears to be contained and finally a couple of vaccine constructs have been in field clinical trials, and show some evidence of protection for the EBOV 2014 strain [2,3].

The wider protective potential of vaccines in the pipeline to protect against any other Ebola strains that might emerge remains a significant unknown, and whether they could provide any protection against the more distant and highly lethal Marburg filovirus. There is therefore a high need for a variety of efficacious drugs and vaccines to pre-empt any such future outbreaks of Ebola or Marburg. Structure based drug and vaccine design has of recent times matured into being of value in the development of therapeutics for infectious disease agents. Our structure based studies of the Ebola GP protein follows from an approach developed in the study of Collateral Immunity[®] for broad-strain seasoned antibody responses and protection of Influenza [4-7]. In this present article we elaborate on an invited medical image published by this journal earlier this year on EBOV GP structure [8].

Ebola Envelope GP: The envelope surface glycoprotein, GP, protrudes from the Ebola virus surface as a homo-trimer that contains the receptor binding and membrane fusion determinants that define cell tropism, entry and infection [9,10]. Consequently a detailed understanding of GP structure and interactions as the principle target for immune defense and for vaccine development is paramount. Indeed, detailed knowledge of the high-resolution, 3.4 Ångstrom,

crystal-structure determination of the central core (pdb structure identifier: 3CSY), that contains receptor binding and fusion sites [9], has been invaluable, and particularly so for the realization of specific target sites for immune antibodies. However, that resolved structure, the most comprehensive of all performed to date, represents only about half of the whole pre-fusion protein sequence of GP. Unfortunately a number of other GP sections are without detailed structural information as they have been recalcitrant to their determination in the crystal. Within those undefined regions are important structural and functional features and for which high resolution information would be invaluable. The focus of this article is on two of those unresolved sections, namely the mucin-like domain (MLD that spans aa 313-501) of GP1 that contains immunodominant antibody sites for host immune protection [11-13], and a principal focus of this study, and the C-terminal section of GP2 (meD aa 658-676) at the GP-virion surface envelope interface.

***Corresponding author:** Garry Lynch, Faculties of Medicine and Veterinary Science, Room 551 Gunn Building (B19), University of Sydney, Camperdown, NSW, 2006, Australia, Tel: 61-286270258; E-mail: garry.lynych@sydney.edu.au

Received May 27, 2015; **Accepted** October 23, 2015; **Published** October 30, 2015

Citation: Lynch GW, Portelli SS, Wong SW, Costa MR, Lemgruber APM, et al. (2015) Imaging of High and Low Resolution Ebola Envelope GP Structures Composited with *in silico* Models of Difficult-to-Resolve Sections. J Mol Genet Med 9(4): 186 doi:10.4172/1747-0862.1000186

Copyright: © 2015 Lynch GW, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Defining protein structure

X-ray crystallography has emerged over time as the principal technology for the description of high-resolution protein structures and remains the most validated procedure. Despite its considerable value it has limitations and can provide variants of the structure not found in nature. Furthermore X-ray crystallography cannot capture all of the possible structures a protein may achieve, and many of these may be temporal and subject to their biologic context and environment (e.g., pH). With crystallography likely capturing only one of many possible structural configurations. Many proteins or sections of proteins can be recalcitrant to X-ray evaluation, as they are either refractory to crystallization or do not contribute to ordered diffraction patterns, due to being heterogeneous in composition, hydrophobic, have mobile elements, or post-translationally modified. The most notable and substantial section that had remained unresolved in the GP 3CSY structure, despite the presence of its sequence in the protein construct used for crystallisation, is the mucin-like domain (MLD). In the attached native state the MLD is the most membrane-distal segment of the envelope GP and projects out above the central, chalice-like, 3CSY core [9,10]. Characteristically this highly glycosylated mucin-like domain would provide a surface environment compatible for EBOV to mix and circulate in mucosal surfaces, by virtue of its similarity with the heterogeneous mucin family of proteins [13,14], that populate the mucosa, and with which it likely shares a degree of structural similarity. Notably the entire mucin-family is poorly represented in the structural database. The high degree of glycosylation of this glycoprotein family is in itself problematic for crystal structure refinement.

Another region of GP for which there is currently minimal structural knowledge is the membrane-associating region, which anchors GP into the EBOV envelope. Hence, we asked the question, whether useful structural information on the MLD and the membrane associating meD sections could be attained from computer-based *in silico* modeling [16]. To assist in evaluating plausible *in silico* simulations, a low-resolution cryoelectron tomograph structure determination of the GP trimer of the EBOV surface has recently been presented [16], and for our studies provides a necessary framework to which acceptable models need to be compatible.

Rational of the study approach

Success in correctly assigning and defining protein structures and as evaluated biennially with the CASP (Critical Assessment of protein Structure Prediction) challenges [18] has realistically improved the potential and value for determining structural information for proteins that defy description by other means. The GP of MLD is considered to fit within the 'highly motile', 'structurally heterogeneous' or 'inherently disordered' category of proteins, and for which no structures have been obtained. The quest is to visualize and determine the structures the GP MLD domain that are exposed for antibody binding, and that might present to cell interactions and to the immune system as a potential target or for avoidance thereof. Before technologies arise that are capable of determining the basic structure(s) of such regions the point of this study was aimed at attempting to generate feasible models of the MLD site of EBOV GP.

Templates, programs and procedures

EBOV Study Strain and sequence: For the modeled structures in this study the Ebola Mayinga 1976 strain Zaire subtype (UniProt Taxon ID: 128952) was the template reference strain sequence used for analysis to ensure direct comparison and consistency with the Ebola

strain sequence upon which the resolved core GP 3CSY structure was generated. That isolate was: Ebola virus/H.sapiens-tc/COD/1976/Yambuku-Mayinga, NCBI: NC_002549. With the specific protein studied; the 676 amino acid (aa) envelope GP protein (UniprotKB ID: Q05320; and Proteome ID: UP000007209).

Molecular Modelling Platforms: The molecular modelling of the MLD and meD segments was performed using the Phyre2 [18,19] (<http://www.sbg.bio.ic.ac.uk/phyre2>); I-Tasser [20,21] (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>); or CABS-flex [22,23] (<http://biocomp.chem.uw.edu.pl/CABSflex/>) software tools. Our model structure of the mucin-like region (MLD) was achieved using Phyre2 [19] and covered amino acids (aa) 313 to 501 across GP1 and included a 313-335 aa overlap with the 3CSY structure [9]. Modelling, rendering or alignment analyses of the GP structures was performed using RASMOL [24], PyMOL [25], and Swiss-PdbViewer, SPDBV_4.1.0_OSX computational tools. Further, our simulations for the membrane-proximal/associating meD region of GP2 (aa 658-676), was achieved using Phyre2. That short C-terminus, 13 aa, model structure was manually aligned with the 3CSY and EMD-9003 structures, and rendered together in PyMOL.

EBOV GP Reference Structures: The template scaffold structures used in this study were the high resolution 3.4 Å 3CSY derived by x-ray crystallography [9] for a truncated and modified GP sequence of the Mayinga 76 strain, and a low-resolution CET produced from Ebola virus-like particles expressing full-length EBOB GP deposited into the Electron Microscopy Data Bank (EMDB accession number 6003 (EMD-6003) (<http://www.emdatabank.org/>) [16] (note:1).

Structure Visualisation and Alignment: Best-fit simulations from the *in silico* structure predictions were mapped against the pre-fusion crystal structure 3CSY rendered in PyMOL. Successful model candidates were further aligned and rendered into the low-resolution, trimeric cryo-electron tomography (CET) structure [16], to assess additional 3D-orientation and spatial compliance.

Structure Geometry Profiling: The coordinates in pdb format of best-fit GP protein Phyre2-models experimentally obtained for the mucin-like (Phyre2 and CABS-Flex structures) and C-terminus regions (Phyre2 structure) were submitted for MolProbity analysis (<http://molprobity.biochem.duke.edu>). Assessments of molecular conflict, and rotamer and Ramachandran characteristics obtained using the MolProbity profiling tool [26,27], were compared against comparable properties generated for several other well determined protein structures. Proteins used for comparison were; the A3 collagen binding domain of von Willebrand Factor (pdb: 1AO3), Human Serum Albumin (: 1BM0), Bovine Serum Albumin (:4F5S), EBOV GP core (: 3CSY)(1976 Mayinga Strain), and EBOV GP core (: 3VE0)(1976 Boniface Strain) protein structures.

Immunomapping: Established sites of protective antibodies for the MLD region of GP were mapped onto the Phyre2 MLD model (MLDm1p1). Sites assessed were; i) the IEDB epitope ID: 156605 by the neutralising antibody C4G7 [11,28-30] and ii) a protective site of the Zaire Mayinga strain for the non-neutralising antibody H13F6, but not protective for other EBOV strains [11,12,31]. These were rendered in PyMOL onto the composite image shown in Figure 1.

Structure Refinement: For assessment and profiling of structure flexibility the MLDm1p1 model coordinates were submitted to the CABS-flex program server [22,23] for analysis (<http://biocomp.chem.uw.edu.pl/CABSflex>).

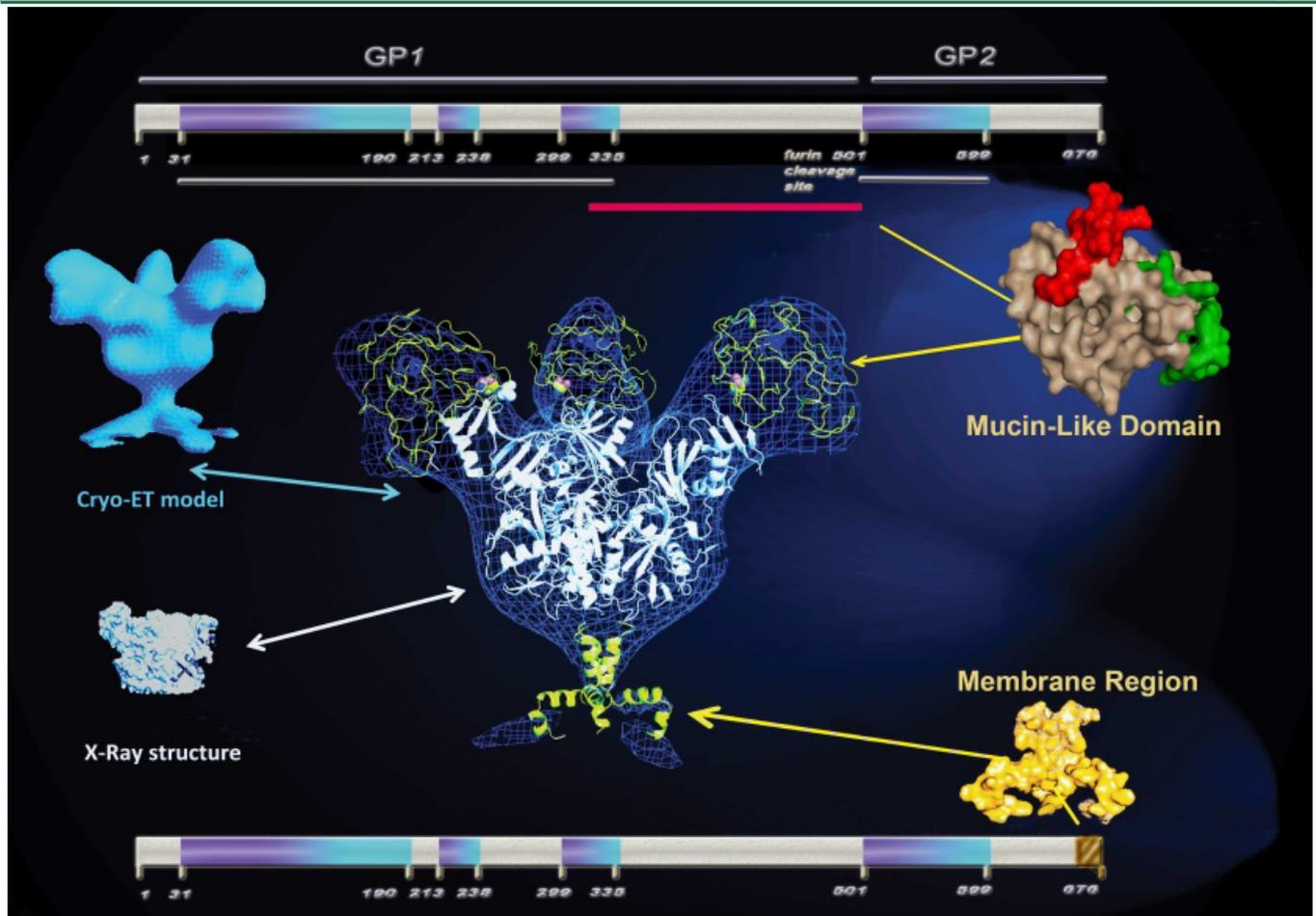


Figure 1: Compilation of resolved and modeled protein domain structures for Ebola virus (EBOV) surface envelope GP.

i) Linear sequence bar of GP amino acids (aa): Duplicate bar graph representations of the 1-676 aa of GP are shown at the top and bottom of the Figure. On the bars, the truncated sequences represented by the refined 3CSY crystal structure are indicated with gradient purple/aqua shading. Within this core structure yet to be structurally defined are the sections bounding and interspersing those 3CSY-defined regions and these are indicated by grey coloring [9]. **Top bar:** At the top of the Figure the GP1 and GP2 regions are designated as well as the furin cleavage site at aa position 501. The red line indicates the MLD-based sequence submitted for molecular simulation. Within this region is a short 10 aa region of overlap with the 3CSY structure sequence and the full mucin-like-region of 313- 501 aa. **Bottom Bar:** Depicted at the far right is a short brown and gold hatched segment showing the membrane associated section of GP sequence submitted for modeling. **ii) Resolved structures:** Shown on the left periphery of the central figure are the resolved structures, for the low-resolution cryo-electron tomograph (Cryo-ET) trimer (EMD-6003) (in blue, upper-left), and the high-resolution X-ray GP monomer core structure (pdb 3CSY) [9] (white, lower-left). **iii) MLD and meD models:** *in silico* space-filled models from Phyre2 simulations are shown to the right of the central image. This shows the designated best-fit GP1 mucin-like domain monomer model (in tan at the upper-right), upon which the regions associated with binding of the protective antibodies have been mapped and these indicated for H13F6 in red [12], and for C4G7 antibody (IED B epitope 156605) in green, respectively [33]. A trimeric model for the GP2, C-terminus membrane-stalk domain is shown in gold coloring at the lower right. **iv) Solved and modelled GP structure GP composite:** In the central image each of these structural elements are assembled as a composite, built in ribbon display within the Cryo ET (blue mesh) global framework. Together these represent the majority of the GP trimer, with the 3 separate mucin-like domains at the very top of the composite (in yellow), a central core trimer (in white) and at the bottom the membrane associating stalk domain (yellow). **Note 2:** For visual representation so that the focal point of the figure was on the *in silico* models, the resolved structures at the left are reduced in both size and color intensity from the central composite and model counterpart images.

Results

Composite imaging of resolved and *in silico* modeled GP domains and structures

In the absence of defined high resolution structures for the mucin-like (amino acids 313-501) and membrane spanning (amino acids 658-676) domains of GP we have performed *in silico* modeling of these regions and composited those models together with the published low EMD-6003 and high 3CSY structures of Ebola GP. Under the imposed structural requirements only 1 of 6 MLD computer simulated models generated by either iTasser or Phyre2 satisfied the spatial fit criteria set above; for continuity with 3CSY, and compatibility and 3D position

within the EMD-6003 trimer footprint. The successful model, ID: MLDm1p1, was modeled using Phyre2. The central composite of Figure 1 shows the X-ray crystal resolved structure, and the *in silico* MLD and meD models mapped together and within the recent cryo-electron spatial EBOV GP trimer tomograph [16]. Shown on the periphery of that image are the space-filled models of the complete low-resolution cryoelectron tomograph (Cryo-ET) trimer structure of GP (EMD-6003) and the high-resolution X-ray crystal structure of the GP trimer core (pdb 3CSY) [9]. On the right of the composite are the *in silico* space-filled models from Phyre2 simulations of a GP1 mucin-like domain monomer and the GP2 C-terminus membrane stalk domain trimer. In the centre image each

of these structural elements has been assembled as a composite, built in ribbon display, and within the Cryo ET global framework. Together these structures represent a majority (~ 85%) of the GP protein, with the 3 separate mucin-like domains at the very top of the composite (yellow), a central core trimer (white) and at the bottom the membrane domain (yellow). Structural visualizations of the refined and the *in silico*-modeled structures are from pdbviewer, Rasmol [24] or PyMOL [25] rendering. Each of the different GP domain sequence determinations were generated for the 1976 Ebola Zaire strain, and imaged to represent a single, prefusion, GP trimer, knob that projects out from the envelope surface. Noting as mentioned this composite of most of the GP is without proposed structures for the GP regions of amino acid sequences: 1-30, 190-213, 238-299 or 599-658). For direct reference of the respective regions already resolved or modeled, a linear map of the complete GP protein sequence has been included in bar format in Figure 1.

Model Validation

Attempts to validate computationally derived molecular models, without experimental or biological evidence and using only computational tools can have considerable limitations. Therefore we have taken 2 different approaches for validation of the MLD model, one based on biological data 1) using antibody binding- site mapping, and alternately 2) by computational assessment.

Relevance to known functional, structural or biologically identified sites

A necessary requirement for any proposed MLD structure is that sites for biologically protective antibodies to the viral envelope GP must be surface exposed and oriented away from the MLD-3CSY interface, which by definition would physically preclude antibody attack at the EBOV envelope surface. When tested for positioning of the sites for the binding of the protective antibodies C4G734 (marked in green) and H13F613 (marked in red) these were mapped onto the proximate MLD model (Figure 1). As shown those sites were found to map to the exterior surface and namely to the top and side surfaces of the model, as would be predicted. Whilst this does not definitively confirm this exact conformation is correct it is appropriate. Importantly, if the converse had been observed, such a structure would be considered an irrelevant configuration and excluded from further consideration.

Computational assessment, validation and molecular refinement

Molecular flexibility: Test algorithms were applied to evaluate the proximate structures modeled here. Due to the inherent characteristics of the MLD and failure to be resolved by crystallography it was deemed useful to computationally test the flexibility of our Phyre2 MLDm1p1 model, using the CAB-Flex web tool. That assessment revealed 12 related and similar structures, indicating a reasonable degree in flexibility for the MLD-m1p1. The model MLDm2c2 was selected for further study and compared with the parent Phyre2 MLDm1p1 structure (Table 1). The MLDm2c2 is highly similar, but has notable differences as evident on alignment.

Geometry profiles: The MLD models (MLDm1p1 and MLDm1c2) co-ordinates were submitted to the MolProbity program to assess their geometric rotameric and Ramachandran properties [26,27]. To put the measured geometries for the best-fit MLD and meD models in perspective, these were matched and compared with the equivalent parameters obtained in MolProbity for a number of other protein structures considered relevant for comparison (Table 1). These were

the A3 collagen binding domain of von Willebrand Factor, human albumin, bovine serum albumin, the GP core 3CSY structure the MLD (MLD-m1p1, MLD-m2c2) and meD. The respective protein structure properties and MolProbity Rotamer and Ramachandran Plot analyses are presented in Table 1.

Comparisons of the acceptable rotamer and Ramachandran plot profiles for the MLD and meD models were made against the profiles obtained for the likely semi-rigid (e.g., A3 collagen-binding domain of vwf) and flexible (e.g., albumin) proteins and x-ray crystallography defined 3CSY and 3VE0 structures. As predicted the A3 domain of vwf gave the best adherence for both rotamer and Ramachandran plot assessments. In general the *in silico* models gave geometry profiles within the range of established and defined structures. However, the MLDm1p1 had the second poorest rotamer score of 12.25% and by far the worst Ramachandran value showing 20.19% of outliers. Notably the CABS-Flex MLDn2c2 derivative of the MLDm1p1, on geometry profiling gave remarkably improved scores of 1.20% poor Rotamers and 4.84% Ramachandran outliers, comparable to the geometries of highly validated structures. This reveals a similar, but even more refined MLD structure structure in the MLDm2c2 model and better compatibility with accepted structure parameters than the MLDm1p1 model.

Conclusions

The essential boundaries and constraints placed for the modeling of MLD and meD regions of GP was that they must be compatible with the resolved low (CEM) and high (crystal) resolution structures. Fitting this basic criteria, the *in silico* MLD model MLDm1p1 structure obtained in these studies was shown to map with i) continuity to the resolved 3CSY structure, and ii) within the 3D-boundaries of the EMD-6003 trimer structure [16]. In immunology based testing of the model for biologic relevance, the G4G7 and H13F6 protective antibody sites were found to map onto the MLDm1p1, at the most exposed surface and apex of the MLD region, as would be appropriate for functionality and immune attack. Additional confidence in the Phyre2 generated MLDm1p1 and meDm1p1 models, was also achieved using computational geometric validation. This included examination of the structural flexibility of the MLDm1p1 using the program CABS-Flex, which returned 12 additional similar but varied model structures and is indicative of considerable structural flexibility for the MLD. A representative MLDm2c2 model was revealed as highly similar, but with notable structural differences to the MLDm1p1, when the two structures were aligned (Table 1). Furthermore, the geometry profiles for the *in silico* models based on MolProbity analysis, fitted within the range of values for the established and defined structures. Albeit the MLDm1p1 model had the second weakest score for both Rotamer and Ramachandran values. However, it was found that the CABS-Flex MLDn2c2 derivative of the MLDm1p1, on geometry profiling gave scores of 98.71% acceptable rotamers and 95.16% acceptable Ramachandran plotting, and was far more comparable to the geometric characteristics of highly validated structures. This has revealed a similar but valuable refinement of the MLD (i.e., the MLDm2c2), for future evaluation, and comparisons of the biological feasibility of both models. Additional structure validations are being pursued, such as compatibility with carbohydrate mapping. Whilst these current studies have mainly focussed on the MLD domain, additional models for the membrane-associating C-terminus have also been built, including a short 13 aa region. Further studies and more extended models that overlap with the 3CSY and 3VE0 core structures are now needed.

Table 1A

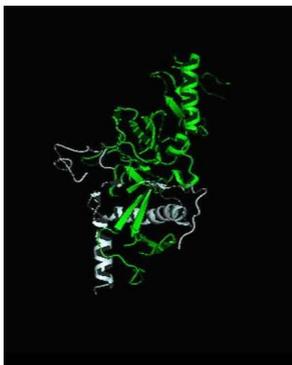
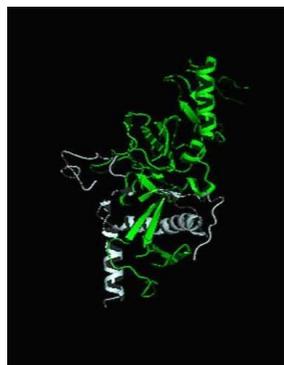
Protein	VWF A3 Domain (Coll Type I / III bind region)	Human Serum Albumin 1BM0	Bovine Serum Albumin (BSA)	3CSY 3CSYFH
EBOV Strain:	N/A	N/A	N/A	Mayinga Zaire 1976
Pdb ID	1AO3	1BM0	4F5S	3CSY
Protein size	186 aa	581 aa	581 aa	187 aa
Protein Structure				
Resolution	1.8 Å	2.5 Å	2.47 Å	3.4 Å
Rotamers Poor	1.3%	6.13%	9.93%	5.51%
Favored	96.09%	82.71%	74.98%	83.18%
Allowed	98.7%	93.87%	90.07%	94.49%
Ramachandran Outliers	0%	2.69%	0.9%	3.42%
Favored	98.37%	86.46%	92.29%	84.17%
Allowed	100%	97.31%	99.1%	96.58%

Table 1B

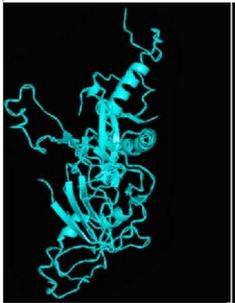
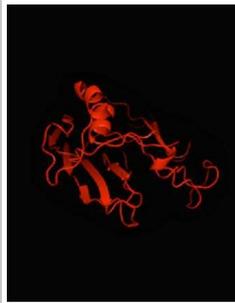
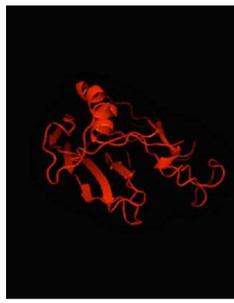
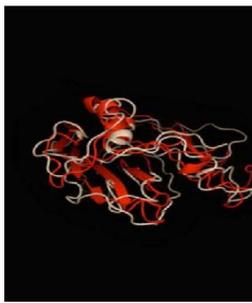
3VE0 EBOV Region	meD (Phyre2)	MLDm1p1 (Phyre2)	MLDm2c2 (CABS)	MLDm1p1 Aligned with MLDm2c2	
Strain	Boniface Sudan 1976	Mayinga Zaire 1976	Mayinga Zaire 1976	Mayinga Zaire 1976	
Pdb ID	3VE0	N/A	N/A	N/A	
Protein size	240 aa	13 aa	188 aa	2 x 188 aa	
Protein Structure					
Resolution	1.8 Å	-	-	-	
Rotamers Poor	18.8%	3.33%	12.26%	1.29%	
Favoured	65.39%	93.33%	77.42%	96.77%	
Allowed	81.2%	96.66%	87.7%	98.71%	
Ramachandran Outliers	4.16%	11.43%	20.19%	4.84%	
Favoured	83.51%	77.14%	59.14%	81.18%	
Allowed	95.84%	88.57%	79.81%	95.16%	

Table 1: Computational Geometric Comparative MolProbity analyses of GP Resolved Core structures & the Proximate MLD and meD Models.

Table 1A and 1B Show respectively the Rotamer and Ramachandran Geometry Comparison Molprobity Profiles of Table 1A) the resolved X-ray crystallography structures for the A3 collagen-binding domain of von Willebrand Factor (vwf, pdb: 1AO4), human (pdb:1BM0) and bovine (pdb: 4F5S), and the central the Ebola GP protein core structures of pdb: 3CSY (1976 EBOV Zaire strain) and in 2A) the 3VE0 (1976 EBOV Sudan strain), and their geometries compared to computational models of the 1976 EBOV Zaire sequence membrane proximal GP 658-676 aa C terminus mpcDm1p1 (modeled in Phyre2), and the 313-501 aa Mucin-like-domain MLDm1pi (modeled in Phyre2) and the MLDm2c12 (modeled in CABS). For this figure each of the crystal or modeled protein structures were visually prepared in PyMol. The very last figure in the sequence showed a combined alignment of the structurally similar Phyre2 and CABS models for the MLD. Note: The displayed structures differ in relative size between panel 1A (at 100%), and panel 1B (at 150%), to enhance visibility of the shorter protein constructs.

The GP *in silico* models shown here fit a range of structural and biomapping criteria, but it needs to be appreciated that without additional qualification, they remain as models or proxies of what those domains might be, and serve at present as interim simulations to provoke new approaches for biotesting, *en route* to the development and determination of more substantiated structures. Certainly, related similar structures generated by the CABS-Flex program and with even stronger geometric validation scores immediately offers new refinements upon which the different models can be tested for, in biological assays and further quality checks. These structures herein provide a starting point upon which experiments will be designed for the biotesting and identification of new antibody target sites. It will therefore also be valuable to know if such simulations of difficult to resolve domains can for the future be beneficial for vaccine and drug design. And whether synergistic use of molecular modeling and cryoelectron tomography (e.g., as used here) may offer a viable alternative for the fine structure determination of proteins that are resistant to crystal structure refinement. An enticing prospect for the future may be that sub 10 Ångstrom level cryoelectron tomograph determinations [34,35] may overcome this technical obstacle and be used together with *ab initio* modeling, to give correct and fully verifiable, high-definition assignments of true native structures. To achieve this would be a holygrail in the realization of difficult-to-resolve protein structures of biological and therapeutic importance.

Acknowledgments

Ana Paula Mello Lemgruber and Maria Rocha Costa are medical students and performing research internships, in the Science without Borders program, a study abroad initiative between the Brazilian Government and the University of Sydney. Terminology; "Collateral Immunity"© GW Lynch, JS Sullivan.

Disclosure

The authors report no conflicts of interest in the performance of this work. RB's institution The Children's Hospital at Westmead has received funding from major Pharmaceutical companies for attendance and presentations at scientific meetings, and honoraria for RB for delivering educational presentations, but this not person-ally accepted by RB.

References

- MSF (2015) <http://www.msf.org/article/ebola-crisis-update-23-march-2015>. Medecins Sans Frontiers.
- Agnandji ST, Huttner A, Zinser ME, Njuguna P, Dahlke C, et al. (2015) Phase 1 Trials of rVSV Ebola Vaccine in Africa and Europe - Preliminary Report. N Engl J Med.
- Zhou Y, Sullivan NJ (2015) Immunology and evolution of the adenovirus prime, MVA boost Ebola virus vaccine. Curr Opin Immunol 35: 131-136.
- Lynch GW, Selleck P, Church WB, Sullivan JS (2012) Seasoned adaptive antibody immunity for highly pathogenic pandemic influenza in humans. Immunol Cell Biol 90: 149-158.
- Lynch GW, Selleck P, Sullivan JS (2009) Acquired heterosubtypic antibodies in human immunity for avian H5N1 influenza. J Mol Genet Med 3: 205-209.
- Lynch GW, Selleck PW, Axell A-M, Downton T, Kapitza N, et al. (2008) Cross-Reactive anti-Avian H5N1 Influenza Neutralizing Antibodies in a Normal 'Exposure-Naïve' Australian Blood Donor Population. The Open Immunol J 1: 1.
- Sullivan JS, Selleck PW, Downton T, Boehm I, Axell AM, et al. (2009) Heterosubtypic anti-avian H5N1 influenza antibodies in intravenous immunoglobulins from globally separate populations protect against H5N1 infection in cell culture. J Mol Genet Med 3: 217-224.
- Lynch GW, Portelli SS, Church WB (2015) Composite Model of Full GP Structure of Ebola Virus Envelope Glycoprotein. Journal of Molecular and Genetic Medicine 9: 101.
- Lee JE, Fusco ML, Hessel AJ, Oswald WB, Burton DR, et al. (2008) Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. Nature 454: 177-182.
- Hofmann-Winkler H, Kaup F, Pöhlmann S (2012) Host cell factors in filovirus entry: novel players, new insights. Viruses 4: 3336-3362.
- Olal D, Kuehne AI, Bale S, Halfmann P, Hashiguchi T, et al. (2012) Structure of an antibody in complex with its mucin domain linear epitope that is protective against Ebola virus. J Virol 86: 2809-2816.
- Barchi JJ Jr (2013) Mucin-type glycopeptide structure in solution: past, present, and future. Biopolymers 99: 713-723.
- Corfield AP (2015) Mucins: a biologically relevant glycan barrier in mucosal protection. Biochim Biophys Acta 1850: 236-252.
- Cavasotto CN, Phatak SS (2009) Homology modeling in drug discovery: current trends and applications. Drug Discov Today 14: 676-683.
- Tran EE, Simmons JA, Bartesaghi A, Shoemaker CJ, Nelson E, et al. (2014) Spatial localization of the Ebola virus glycoprotein mucin-like domain determined by cryo-electron tomography. J Virol 88: 10958-10962.
- Park H, DiMaio F, Baker D (2015) CASP11 refinement experiments with ROSETTA. Proteins.
- Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4: 363-371.
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10: 845-858.
- Yang J, Yan R, Roy A, Xu D, Poisson J, et al. (2015) The I-TASSER Suite: protein structure and function prediction. Nat Methods 12: 7-8.
- Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res 43: W174-181.
- Jamroz M, Kolinski A, Kmiecik S (2013) CABS-flex: Server for fast simulation of protein structure fluctuations. Nucleic Acids Res 41: W427-431.
- Jamroz M, Kolinski A, Kmiecik S (2014) CABS-flex predictions of protein flexibility compared with NMR ensembles. Bioinformatics 30: 2150-2154.
- Sayle RA, Milner-White EJ (1995) RASMOL: biomolecular graphics for all. Trends Biochem Sci 20: 374.
- DeLano WL (2002) The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, Ca, USA.
- Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35: W375-383.
- Davis IW, Murray LW, Richardson JS, Richardson DC (2004) Molprobity: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic acids research 32: W615-619.
- Ponomarenko J, Vaughan K, Sette A, Maurer-Stroh S (2014) Conservancy of mAb Epitopes in Ebolavirus Glycoproteins of Previous and 2014 Outbreaks. PLoS Curr 6.
- Qiu X, Audet J, Wong G, Pillet S, Bello A, et al. (2012) Successful treatment of ebola virus-infected cynomolgus macaques with monoclonal antibodies. Science translational medicine 4: 138ra181.
- Qiu X, Fernando L, Melito PL, Audet J, Feldmann H, et al. (2012) Ebola GP-specific monoclonal antibodies protect mice and guinea pigs from lethal Ebola virus infection. PLoS Negl Trop Dis 6: e1575.
- Audet J, Wong G, Wang H, Lu G, Gao GF, et al. (2014) Molecular characterization of the monoclonal antibodies composing ZMAB: a protective cocktail against Ebola virus. Sci Rep 4: 6881.
- Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18: 2714-2723.
- Qiu X, Alimonti JB, Melito PL, Fernando L, Ströher U, et al. (2011) Characterization of Zaire ebolavirus glycoprotein-specific monoclonal antibodies. Clin Immunol 141: 218-227.
- Diebolder CA, Koster AJ, Koning RI (2012) Pushing the resolution limits in cryo electron tomography of biological structures. J Microsc 248: 1-5.
- Schur FK, Hagen WJ, de Marco A, Briggs JA (2013) Determination of protein structure at 8.5Å resolution using cryo-electron tomography and sub-tomogram averaging. J Struct Biol 184: 394-400.

Note 1: Cryoelectron micrographs (CEM) and cryoelectron tomographs (CET) are from 2D tilted and tiled CEM protein images and averaged from of up to million of individual structures, to reveal a spectrum of structures captured in liquid phase. Heterogeneity in captured structures thus dictates how discrete or broad the returned refinements are, and contrasts to far fewer and more defined X-ray crystal structures. CEM/CET strength is imaging of molecules in native water-soluble state, versus crystallography with close packing and contacts and use of non-biological, harsh solvents.