research papers

Techniques and tactics used in determining the structure of the trimeric *ebolavirus* glycoprotein

Jeffrey E. Lee,[a] Marnie L. Fusco,[a] Dafna M. Abelson,[a] Ann J. Hessell,[a] Dennis R. Burton[a] and Erica Ollmann Saphire[a,b]*

[a]Department of Immunology and Microbial Science, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA, and [b]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

Correspondence e-mail: erica@scripps.edu

The trimeric membrane-anchored *ebolavirus* envelope glycoprotein (GP) is responsible for viral attachment, fusion and entry. Knowledge of its structure is important both for understanding *ebolavirus* entry and for the development of medical interventions. Crystal structures of viral glycoproteins, especially those in their metastable prefusion oligomeric states, can be difficult to achieve given the challenges in production, purification, crystallization and diffraction that are inherent in the heavily glycosylated flexible nature of these types of proteins. The crystal structure of *ebolavirus* GP in its trimeric prefusion conformation in complex with a human antibody derived from a survivor of the 1995 Kikwit outbreak has now been determined [Lee *et al.* (2008), *Nature (London)*, **454**, 177–182]. Here, the techniques, tactics and strategies used to overcome a series of technical roadblocks in crystallization and phasing are described. Glycoproteins were produced in human embryonic kidney 293T cells, which allowed rapid screening of constructs and expression of protein in milligram quantities. Complexes of GP with an antibody fragment (Fab) promoted crystallization and a series of deglycosylation strategies, including sugar mutants, enzymatic deglycosylation, insect-cell expression and glycan anabolic pathway inhibitors, were attempted to improve the weakly diffracting glycoprotein crystals. The signal-to-noise ratio of the search model for molecular replacement was improved by determining the structure of the uncomplexed Fab. Phase combination with Fab model phases and a selenium anomalous signal, followed by NCS-averaged density modification, resulted in a clear interpretable electron-density map. Model building was assisted by the use of B-value-sharpened electron-density maps and the proper sequence register was confirmed by building alternate sequences using N-linked glycan sites as anchors and secondary-structural predictions.

1. Introduction

Ebolavirus (EBOV) causes a severe hemorrhagic fever with 50–90% lethality and outbreaks of the virus have increased fourfold in the last decade. The *ebolavirus* glycoprotein (GP) is the only virally expressed protein on the virion surface and is critical for attachment to and fusion with host cells. Hence, EBOV GP is the critical target of neutralizing antibodies and is an important component of vaccines. The *Zaire ebolavirus* (ZEBOV) surface glycoprotein contains 676 amino acids and is post-translationally cleaved by furin (Volchkov *et al.*, 1998) into two disulfide-linked subunits: GP1 and GP2. Three GP1–GP2 units form a 450 kDa trimeric spike on the virus surface (Sanchez *et al.*, 1998). GP1 contains an N-terminal signal sequence (Sanchez *et al.*, 1993, 1998), a putative receptor-

binding site (Kuhn *et al.*, 2006; Manicassamy *et al.*, 2005) and a heavily glycosylated mucin-like domain. GP2 contains an internal fusion loop, two heptad-repeat regions separated by a CX_6CC disulfide motif, an ~30-residue transmembrane anchor and a four-residue C-terminal cytoplasmic tail (Feldmann *et al.*, 2001).

Although the structures of viral and mammalian glycoproteins such as EBOV GP are of biological interest, crystallization of these proteins is complicated by several factors. It can be difficult to express mammalian proteins in high yields in stable and soluble forms. Furthermore, multiple rounds of construct design and redesign may be necessary to identify protein variants that are suitable for structural studies. In addition, glycoproteins often require their oligosaccharide chains for proper folding and stability. However, the chemical and conformational heterogeneity of these glycans generally inhibit the formation of a well ordered lattice.

We have determined the crystal structure of the prefusion trimeric ZEBOV GP complexed with a neutralizing antibody (KZ52) identified in a human survivor of the 1995 Kikwit outbreak (Lee *et al.*, 2008). Determination of the structure of this protein was particularly challenging, as EBOV GP expresses poorly with heterogeneous glycosylation. Moreover, crystals diffracted weakly and a lack of native methionines made selenium incorporation and initial phasing difficult. We generated over 140 different constructs of GP, grew ~50 000 crystals and harvested and screened 800 of the largest ones in order to identify one crystal that diffracted to 3.4 Å resolution. Here, we describe the strategies that were used successfully and unsuccessfully to control N-linked glycosylation, phase the large macromolecular complex and trace the backbone in low-resolution maps. Our experiences in this endeavor may provide key technical suggestions for similar problematic crystallographic situations.

2. Materials and methods

2.1. Protein preparation

2.1.1. Protein-construct design and expression screening. The ZEBOV glycoprotein primary sequence (GenBank accession code AAG40168.1) was subjected to a series of bioinformatics algorithms in order to identify the location and lengths of predicted secondary-structural elements (*Network Protein Sequence Analysis*; Combet *et al.*, 2000), potential N-linked and O-linked glycosylation sites (*NetNGlyc/ NetOGlyc*; Gupta *et al.*, 2004; Julenius *et al.*, 2005), regions of low complexity which may be associated with disorder (*DISOPRED*2; Ward *et al.*, 2004), the location of signal peptides (*SignalP*; Emanuelsson *et al.*, 2007) and transmembrane anchors (*TMHMM/Phobius*; Kall *et al.*, 2007). The full-length DNA sequence of ZEBOV (Mayinga strain; GenBank accession code U23187) was codon-optimized for expression in *Homo sapiens* and whole-gene synthesized by Blue Heron Biotechnology (Bothell, Washington, USA). The GP DNA was subsequently cloned into the *Bgl*II/*Sal*I restriction sites of the pDISPLAY (Invitrogen, Carlsbad, California, USA)

multiple cloning-site region, with a stop codon introduced before the internal transmembrane segment of the vector. Deletions of the mucin-like domains were engineered using overlap extension PCR (Heckman & Pease, 2007).

All ZEBOV GP constructs were tested for expression in human embryonic kidney (HEK) 293T cells (American Type Culture Collection No. CRL-1573, Manassas, Virginia, USA) using six-well culture plates (Corning, Lowell, Massachusetts, USA). All cell-culture media and supplements were purchased from GIBCO/Invitrogen (Carlsbad, California, USA). 2.0×10^6 cells were grown in 2 ml Dulbecco's Modified Eagle's Medium (DMEM), $1\times$ Pen/Strep, $1\times$ GlutaMAX and $5\%(v/v)$ FBS and incubated at 310 K with 5% CO_2 for 4 h to allow cell attachment. MiniPrep DNA (Qiagen, Hilden, Germany) was transiently transfected into HEK293T cells using FuGene HD (Roche Diagnostics, Indianapolis, Indiana, USA) according to the manufacturer's protocol. Transfected cells were subsequently incubated at 310 K with 5% CO_2 for 4 d. The supernatant was harvested and dilute recombinant protein was detected by nonreducing Western blots. A more detailed protocol is given in Lee *et al.* (2009). Briefly, protein was separated on a 10–15% gradient SDS–Tris–HCl polyacrylamide gel (Bio-Rad Laboratories, Hercules, California, USA; samples were not heated or reduced) and transferred onto an activated Immobilon-P membrane (Millipore, Billerica, Massachusetts, USA). The transferred membrane was probed with either anti-hemagglutinin (HA) 16B12 (Covance, Princeton, New Jersey, USA) or KZ52 (Maruyama, Parren *et al.*, 1999; Maruyama, Rodriguez *et al.*, 1999) primary antibodies. An alkaline-phosphatase-conjugated secondary antibody was incubated with the transferred membrane prior to development with SIGMA *FAST* BCIP/NBT (Sigma–Aldrich, St Louis, Missouri, USA) according to the manufacturer's protocol.

2.1.2. Preparation of GPΔmuc$_{312–463}$Δtm–KZ52 complex. From small-scale expressions of the various EBOV GP constructs, the highest expressing and most homogeneous variant was GPΔmuc$_{312–463}$Δtm (see §3.1 for a more detailed description). Large-scale expression of ZEBOV GPΔmuc$_{312–463}$-Δtm was performed using HEK293T cells transfected by standard calcium phosphate precipitation (Kingston *et al.*, 2003) in ten-layer CellSTACKS (Corning; 6360 cm^2 surface area; 1.3 l medium; Lee *et al.*, 2008, 2009). The DNA/calcium phosphate mixture was added to 70% confluent cells grown in 1.3 l DMEM plus $1\times$ Pen/Strep and $5\%(v/v)$ FBS. The supernatant was harvested 4 d post-transfection by centrifugation and filtered with a 0.22 μm Acrodisk (Pall Corp, East Hills, New York, USA) prior to being concentrated to 150 ml using a Centramate tangential flow filtration system with an Omega membrane cassette (molecular-weight cutoff 30 kDa; Pall Corp.). The concentrated glycoprotein was purified on a 2 ml bed-volume anti-HA-agarose immunoaffinity column (Roche Applied Sciences) by gravity at a flow rate of 1 ml min^{-1}. Bound GPΔmuc$_{312–463}$Δtm was washed extensively with $1\times$ Dulbecco's phosphate-buffered saline (PBS; GIBCO/Invitrogen) and $0.05\%(v/v)$ Tween-20 (Sigma–Aldrich), and eluted from the column by competition with

research papers

1 mg ml^{-1} synthetic hemagglutinin (HA) peptide (sequence: YPYDVPDYA) dissolved in PBS. Collected fractions were separated on a 10–15% gradient SDS–Tris–HCl polyacrylamide gel and all fractions containing GPΔmuc$_{312-463}\Delta$tm were pooled. Purified GPΔmuc$_{312-463}\Delta$tm was enzymatically deglycosylated using peptide N-glycosidase F (PNGaseF; New England Biolabs, Ipswich, Massachusetts, USA) at a final concentration of 1000 units ml^{-1} at room temperature for 18 h with 10% glycerol added for protein stability.

Glycosylated and deglycosylated GPΔmuc$_{312-463}\Delta$tm were complexed individually with Fab KZ52 to facilitate crystallization. Protocols for the expression of native immunoglobulin G (IgG) KZ52 and the purification of Fab fragments for crystallization have been described previously (Lee *et al.*, 2008). Briefly, IgG KZ52 ($\sim$3 mg ml^{-1}) was digested for 2 h with a final concentration of 2%(v/v) activated papain (Sigma–Aldrich), digestion was terminated using 50 mM iodoacetamide (Sigma–Aldrich) and samples were buffer-exchanged into 1$\times$ PBS using an Amicon Ultrafree-4 centrifugal concentrator (molecular-weight cutoff 10 kDa; Millipore). Cleaved Fc and uncleaved IgG were loaded onto a 5 ml Protein A affinity column (GE Healthcare, Piscataway, New Jersey. USA). The flowthrough, containing Fab KZ52, was collected, buffer-exchanged into 50 mM sodium acetate pH 4.7 and 20 mM NaCl (buffer A) and subsequently loaded onto a Mono S 5/5 column (GE Healthcare). Two Fab isoforms were separated on a gradient of 0–30% buffer A + 1 M NaCl over 80 column volumes. The higher molecular-weight isoform of Fab KZ52 was mixed in an 1.5 molar excess with either fully glycosylated or PNGaseF-treated GPΔmuc$_{312-463}\Delta$tm and incubated on ice for 1 h. Prior to crystallization, the glycoprotein–antibody complexes were purified on a Superdex 200 10/300 GL (GE Healthcare) column equilibrated with 10 mM Tris–HCl pH 7.5 and 150 mM NaCl. Interestingly, both trimeric and monomeric species of the EBOV GPΔmuc$_{312-463}$-Δtm–KZ52 complex were resolved on the Superdex-200 column, although only trimeric species of GPΔmuc$_{312-463}\Delta$tm were noted in the absence of KZ52. It is possible that the GP trimer interface is somewhat unstable in the presence of KZ52, although the reasons why are as yet unclear. Based on the chromatogram and SDS–PAGE analysis, the trimeric and monomeric GPΔmuc$_{312-463}\Delta$tm–Fab fractions were pooled separately, but only the trimeric complex was used in subsequent studies.

2.2. Crystallization and diffraction

Glycosylated and deglycosylated GPΔmuc$_{312-463}\Delta$tm–KZ52 were concentrated to $\sim$10 mg ml^{-1} using Amicon Ultrafree-0.5 centrifugal concentrators (10 kDa molecular-weight cutoff). OptiMix I, II and III and PEG sparse-matrix screens (Fluidigm Corp., South San Francisco, California, USA) were set up using the Topaz system (Fluidigm Corp.), which uses free-interface liquid diffusion to effect crystallization. The crystallization chips were stored at 295 K and were examined at t = 0, 24, 48, 96 and 168 h post-setup using an AutoInspeX II workstation (Fluidigm Corp.).

The top two crystal hits were translated to traditional hanging-drop vapour diffusion by mixing 1.5 µl protein solution and 1.5 µl precipitant solution and equilibrating against 1 ml of the same precipitant solution. Crystals were grown in an incubator maintained at 295 K. Crystal form A grew as large rod-shaped crystals (0.4 $\times$ 0.2 $\times$ 0.2 mm) over a two-week period in 8.5%(w/v) PEG 6000, 0.1 M sodium acetate pH 4.8 and 1.0 M NaI. Crystal form B formed large rhombohedral crystals (0.2 $\times$ 0.2 $\times$ 0.2 mm) over a three-week period in 8.5%(w/v) PEG 10 000, 0.1 M Tris–HCl pH 8.5, 0.6 M sodium acetate and 10%(v/v) PEG 200. Crystals were looped and soaked with a variety of cryoprotectants [40%(v/v) glycerol, 45%(v/v) glucose, 100%(v/v) Paratone-N, 40%(v/v) ethylene glycol, 40%(v/v) MPD, 45%(v/v) PEG 200 or 40%(v/v) PEG 400] prior to being flash-cooled in liquid nitrogen. Crystals were exposed to X-rays on a home rotating-anode FR-D X-ray generator (Rigaku, Woodlands, Texas, USA) equipped with a MAR 345 image plate (Rayonix/MAR USA; Evanston, Illinois, USA) and on beamlines 4.2.2, 5.0.2, 8.2.1, 8.2.2, 8.3.1 and 12.3.1 at the Advanced Light Source (ALS; Berkeley, California, USA), and beamlines 9-2 and 11-1 at the Stanford Synchrotron Radiation Laboratory (SSRL; Menlo Park, California, USA).

2.3. Attempts to improve crystal diffraction

2.3.1. Insect-cell expression. GPΔmuc$_{312-463}\Delta$tm was produced in *Trichoplusia ni* insect cells (High Five; Invitrogen) by stable and baculovirus-based expression, according to the manufacturer's protocols. Briefly, to create a stable cell line, GPΔmuc$_{312-463}\Delta$tm DNA was subcloned into the pMIB vector (Invitrogen) and transfected into 60% confluent High Five cells using Cellfectin (Invitrogen) in T-25 cm^2 flasks. 2 d post-transfection, the cells were split to $\sim$20% confluency and incubated overnight with selection media [Express Five serum-free media (SFM; Invitrogen), 1$\times$ GlutaMAX, containing 60 µg ml^{-1} blasticidin (Invitrogen)]. The selection medium was changed every 4 d and expression was tested after 2–3 weeks. Selected High Five cells were subsequently adapted for growth in suspension by transferring 4 $\times$ 10^5 cells to 100 ml Express Five SFM, 1$\times$ GlutaMAX, 10 µg ml^{-1} blasticidin and 10 U ml^{-1} heparin (Invitrogen) in a small shaker flask. At a concentration of 2 $\times$ 10^6 cells ml^{-1}, High Five cells were expanded into 1 l Express Five SFM, 1$\times$ GlutaMAX and 10 µg ml^{-1} blasticidin in 2 l shaker flasks. For large-scale expression, 2 l stable insect cells were grown at 289 K for 4 d prior to harvest.

Baculovirus-based expression was performed using the Sapphire vector (Orbigen/Allele Biotech, San Diego, California, USA). A baculovirus stock was amplified in Sf9 insect cells (Orbigen/Allele Biotech) according to the manufacturer's protocol using HyQ SFX-Insect Medium (Thermo Fisher Scientific/HyClone, Waltham, Massachusetts, USA) with 2$\times$ GlutaMAX and 10 µg ml^{-1} blasticidin and titred using the FastPlax kit (EMD Biosciences/Novagen, San Diego, California, USA). Small-scale expression was optimized in 100 ml shaker flasks by varying the amounts of virus and the

length of expression. For large-scale production, 2 l High Five cells (2×10^6 cells ml^{-1}) in HyQ SFX-Insect Medium, 2× GlutaMAX and 10 µg ml^{-1} blasticidin were infected at a multiplicity of infection (MOI) of 5. For both stable and baculovirus-based expression of GPΔmuc$_{312-463}\Delta$tm, supernatants were harvested by centrifugation 4 d post-scale-up or post-transfection and concentrated to 150 ml using a Centramate tangential flow concentration system (molecular-weight cutoff 30 kDa). Concentrated protein was loaded onto an Ni–NTA matrix (Qiagen), which was equilibrated in 50 mM Tris–HCl pH 8.0, 300 mM NaCl and 20 mM imidazole. Insect cell-produced GPΔmuc$_{312-463}\Delta$tm was eluted with a step gradient of 50, 100, 250, 375 and 500 mM imidazole in 50 mM Tris–HCl pH 8.0 and 300 mM NaCl. All fractions that contained GPΔmuc$_{312-463}\Delta$tm were pooled according to SDS–PAGE analysis. Subsequently, GPΔmuc$_{312-463}\Delta$tm was deglycosylated at room temperature with EndoF3 (Calbiochem/EMD) or EndoH (New England Biolabs) in 100 mM sodium citrate pH 5.5 prior to Fab KZ52 complexation and crystallization, as described in §§2.1 and 2.2.

2.3.2. Glycan anabolic pathway inhibitors. Kifunensine (Toronto Research Chemicals, Toronto, Canada) was added at a final concentration range of 1–8 µg ml^{-1} to 70% confluent HEK293T cells in a ten-layer CellSTACK and incubated for 2 h before calcium phosphate transfection of the pDISPLAY vector encoding GPΔmuc$_{312-463}\Delta$tm to allow inhibitor uptake. Transfection, purification and crystallization procedures were also performed as described in §§2.1 and 2.2.

2.3.3. Point mutations. Single-, double-, triple-, quadruple- and quintuple-site point mutations were generated according to the manufacturer's protocol using the QuikChange II or QuikChange Multi site-directed mutagenesis kits (Stratagene, La Jolla, California, USA). All glycan mutants were confirmed by DNA sequencing and subsequently expressed, purified and crystallized as described in §§2.1 and 2.2.

2.3.4. Chaotrope-assisted enzymatic deglycosylation. 3 ml of a 0.1 mg ml^{-1} sample of previously PNGaseF-treated T42V/T230V GPΔmuc$_{312-463}\Delta$tm was further deglycosylated overnight in the presence of 2500 units of PNGase F and a final concentration of 1.5 M urea. The reaction was incubated at 310 K for 3 h and subsequently buffer-exchanged into 10 mM Tris–HCl pH 7.5, 150 mM NaCl and 10%(v/v) glycerol using an Amicon Ultrafree-4 centrifugal concentrator (molecular-weight cutoff 10 kDa) prior to complexation with Fab KZ52 and purification by size-exclusion chromatography as described in §2.1.

2.4. Structure determination of the GPΔmuc$_{312-463}\Delta$tm–KZ52 complex

2.4.1. Data collection: 4.0 Å resolution crystal. The urea PNGaseF-treated T42V/T230V GPΔmuc$_{312-463}\Delta$tm–KZ52 complex was crystallized in 13%(w/v) PEG 4000, 0.1 M Tris–HCl pH 8.4, 0.4 M sodium malonate by hanging-drop vapor diffusion over a three-week period at 295 K. A single crystal was sequentially soaked in 10, 20, 30 and 40%(v/v) glycerol in 15%(w/v) PEG 4000, 0.4 M sodium malonate and 0.1 M Tris–

HCl pH 8.4 prior to being flash-cooled in a bowl of liquid nitrogen. Data were measured remotely on beamline 11-1 at SSRL using an Area Detector Systems Corporation (ADSC, Poway, California, USA) Quantum 315 CCD detector. The incident X-ray beam was collimated to 75 × 75 µm and the diffraction of the crystal was monitored visually. Once the resolution limits had deteriorated by >1 Å, as judged visually by the fading of reflections on the display, the crystal was translated to expose a fresh region for diffraction. Hence, a complete data set was obtained by merging diffraction from two segments of one single crystal. The first data segment was indexed and integrated using $d*TREK$ (Pflugrath, 1999), the resulting orientation matrix was used to index the second segment and the two segments were merged prior to absorption correction and scaling. Data statistics are presented in Table 1.

2.4.2. Data collection: 3.4 Å resolution selenomethionine-incorporated crystal. Selenomethionine-containing KZ52 Fab was produced by expressing IgG KZ52 in Chinese hamster ovary (CHO) cells (American Type Culture Collection, catalog No. CCL-61) cultured in methionine-free DMEM supplemented with 60 mg l^{-1} L-selenomethionine (SeMet; Sigma–Aldrich). The secreted IgG KZ52 was purified by Protein A chromatography (Thermo Scientific/Pierce, Rockford, Illinois, USA) according to the manufacturer's protocol. IgG was cleaved to Fab and complexed and crystallized with T42V/T230V GPΔmuc$_{312-463}\Delta$tm as described in previous sections. The urea PNGaseF-treated T42V/T230V GPΔmuc$_{312-463}\Delta$tm–SeMet KZ52 complex was crystallized in 8.25%(w/v) PEG 10 000, 0.1 M Tris–HCl pH 8.5, 0.4 M sodium acetate and 10%(w/v) PEG 200 by vapor diffusion over a three-week period at 295 K. Hanging-drop crystals were gently cross-linked by placing 2 µl 100%(v/v) glutaraldehyde on a microbridge in the reservoir for 20 min. Cross-linked crystals were subsequently cryoprotected in sequential soaks of 10, 20, 30 and 40%(v/v) glycerol in 10%(w/v) PEG 10 000, 0.1 M Tris–HCl pH 8.6, 0.4 M sodium acetate and 10%(v/v) PEG 200 prior to flash-cooling in a bowl of liquid nitrogen. SAD data sets were collected from two crystals that diffracted to 3.4 and 4.0 Å resolution, respectively, at the peak wavelength ($\lambda = 0.98030$ Å) with inverse-beam geometry in 10° data wedges on ALS beamline 5.0.2. The data from the two crystals were independently indexed, integrated and scaled with $d*TREK$ (Pflugrath, 1999; Table 1).

2.4.3. Small-angle X-ray scattering studies. Small-angle X-ray scattering (SAXS) studies on the T42V/T230V GPΔmuc$_{312-463}\Delta$tm–KZ52 complex in 10 mM Tris–HCl pH 7.5, 150 mM NaCl and 10%(v/v) glycerol were undertaken on the SIBYLS beamline 12.3.1 at ALS. Prior to data collection, SAXS protein samples were analyzed by dynamic light scattering and size-exclusion chromatography to confirm the monodispersity of the sample. A series of long and short X-ray exposures were collected at a wavelength of 1.1271 Å using a MAR 165 CCD (Rayonix/MAR USA) using three different protein concentrations (6, 3 and 1.5 mg ml^{-1}) to control for possible protein aggregation and concentration effects. Long exposures were used to collect the weakly scattered higher

Table 1
Data-collection, phasing and refinement statistics.

Values in parentheses are for the outer resolution shell.

	GPΔmuc$_{312-463}\Delta$tm–KZ52 Native	GPΔmuc$_{312-463}\Delta$tm–KZ52 SeMet crystal 1	GPΔmuc$_{312-463}\Delta$tm–KZ52 SeMet crystal 2	Fab KZ52
Data collection				
Protein source	CHO/293T†	CHO/293T†	CHO/293T†	CHO
Space group	H32	H32	H32	$P3_221$
Unit-cell parameters				
$a = b$ (Å)	275.4	273.1	275.6	129.5
c (Å)	425.0	409.7	415.4	185.0
$\alpha = \beta$ (°)	90	90	90	90
γ (°)	120	120	120	120
Wavelength (Å)	0.97945	0.98030	0.98030	1.11588
Resolution (Å)	44.7–4.0 (4.14–4.0)	48.4–3.4 (3.5–3.4)	48.8–4.0 (4.14–4.0)	48.0–2.5 (2.6–2.5)
No. of reflections	169885	626036	368030	433797
No. of unique reflections	50930	151305	45017	62060
R_{merge}‡ (%)	9.3 (47.1)	14.2 (56.8)	17.7 (56.1)	11.8 (62.7)
Completeness (%)	97.2 (99.3)	96.2 (94.7)	87.8 (91.0)	99.0 (98.8)
Redundancy	3.3 (3.4)	4.1 (3.6)	8.2 (8.0)	7.0 (6.4)
$I/\sigma(I)$	6.5 (2.2)	6.2 (1.7)	6.3 (2.8)	7.5 (2.0)
Refinement				
No. of heavy-atom sites		20		n/a
No. of protein atoms		23122		6707
No. of glycan atoms		417		n/a
Resolution (Å)		48.4–3.4		48.0–2.5
No. of reflections (work/test)		146996/4405		108605/5633
R_{work}/R_{free}§ (%)		26.2/30.6		20.3/25.2
Average B values (Å^2)				
Fab KZ52 [variable/constant]		128.6 [108.5/152.4]		66.8 [60.9/73.2]
GP1 [base/head/glycan cap]		113.9 [106.0/110.8/129.8]		n/a
GP2		104.0		n/a
Glycans		142.3		n/a
Ramachandran statistics				
Favoured region (%)		—		95.8
Outliers (%)		2.4		0.6
MolProbity score		3.10		2.62

† GPΔmuc$_{312-463}\Delta$tm was produced using HEK293T cells and KZ52 was expressed in CHO cells. ‡ $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ and $\langle I(hkl)\rangle$ represent the diffraction-intensity values of the individual measurements and the corresponding mean values. The summation is over all unique measurements. § $R_{work} = \sum_{hkl} ||F_{obs}| - |F_{calc}||/\sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. For R_{free} the sum extends over a subset of reflections excluded from all stages of refinement.

angle data, while short exposures were selected to maximize accurate small-angle measurement and minimize CCD detector overloads near the beam stop. All data were buffer-subtracted. Scattering profiles from long and short exposures were merged using the program *PRIMUS* (Konarev *et al.*, 2003). The radiation sensitivity of the samples was assessed by superimposing SAXS profiles from successive short exposures. The real-space pair distribution function $P(r)$ was determined from scattering data using *GNOM* (Svergun, 1992) and *ab initio* model calculations were performed using *GASBOR* (Svergun *et al.*, 2001) with no input other than an expected 2900 residues of scattering mass and threefold symmetry.

2.4.4. Improvement of search model: structure of Fab KZ52. Mono S-purified Fab KZ52 was concentrated to 15 mg ml^{-1} and crystallized by hanging-drop vapor diffusion at 295 K. Large 0.7 × 0.3 × 0.2 mm rod-shaped crystals were grown directly in Wizard III condition No. 45 [1.5 *M* ammonium sulfate, 0.1 *M* Tris–HCl pH 8.5 and 12%(*v/v*) glycerol]. Hanging-drop crystals were gently cross-linked by placing 2 µl 100%(*v/v*) glutaraldehyde on a microbridge in the reservoir for 20 min. Cross-linked crystals were gently soaked in sequential steps of 20 and 30%(*w/v*) glucose in 1.5 *M* ammo-

nium sulfate, 0.1 *M* Tris–HCl pH 8.5 and 12%(*v/v*) glycerol prior to flash-cooling in liquid nitrogen. These crystals diffracted to ~2.5 Å resolution and a complete data set was collected on ALS beamline 8.3.1 using an ADSC Q210 CCD detector. The data were indexed, integrated and scaled with *d*TREK* (Table 1). Analysis of Matthews coefficients (Matthews, 1968) suggested the presence of two Fab fragments per asymmetric unit. Molecular replacement was performed in *CNSsolve* (v.1.2; Brünger *et al.*, 1998) using the variable and constant domains of the anti-HIV-1 neutralizing human antibody b12 (PDB code 1n0x; Saphire *et al.*, 2007) as initial search models. Cross-rotation and translation functions clearly identified one constant domain. This constant domain was fixed and a subsequent translation search clearly revealed the position of the second constant domain. The molecular-replacement searches were then repeated for the variable domains until both complete Fab models in the asymmetric unit had been generated. Coordinates corresponding to this molecular-replacement solution were subsequently subjected to rigid-body and torsion-angle simulated annealing from a starting temperature of 5000 K using all data with no σ cutoffs in *CNSsolve* (Brünger *et al.*, 1990, 1998) and σ_A-weighted

$mF_o - DF_c$ and $2mF_o - DF_c$ Fourier maps were calculated. The program *Coot* (v.0.3.3; Emsley & Cowtan, 2004) was used to manually rebuild the initial model to the correct KZ52 sequence and alternated with rounds of amplitude-based maximum-likelihood torsion-angle simulated annealing and individual *B*-value refinement with an overall anisotropic temperature and bulk-solvent correction. Water molecules

were included into the model during the later rounds of refinement based on the presence of positive 3σ peaks in the σ_A-weighted $mF_o - DF_c$ difference electron-density maps and at least one hydrogen bond to a protein, peptide or solvent atom. After the addition of water molecules, alternating rounds of crystallographic conjugate-gradient minimization refinement with TLS refinement (Winn *et al.*, 2001) and model
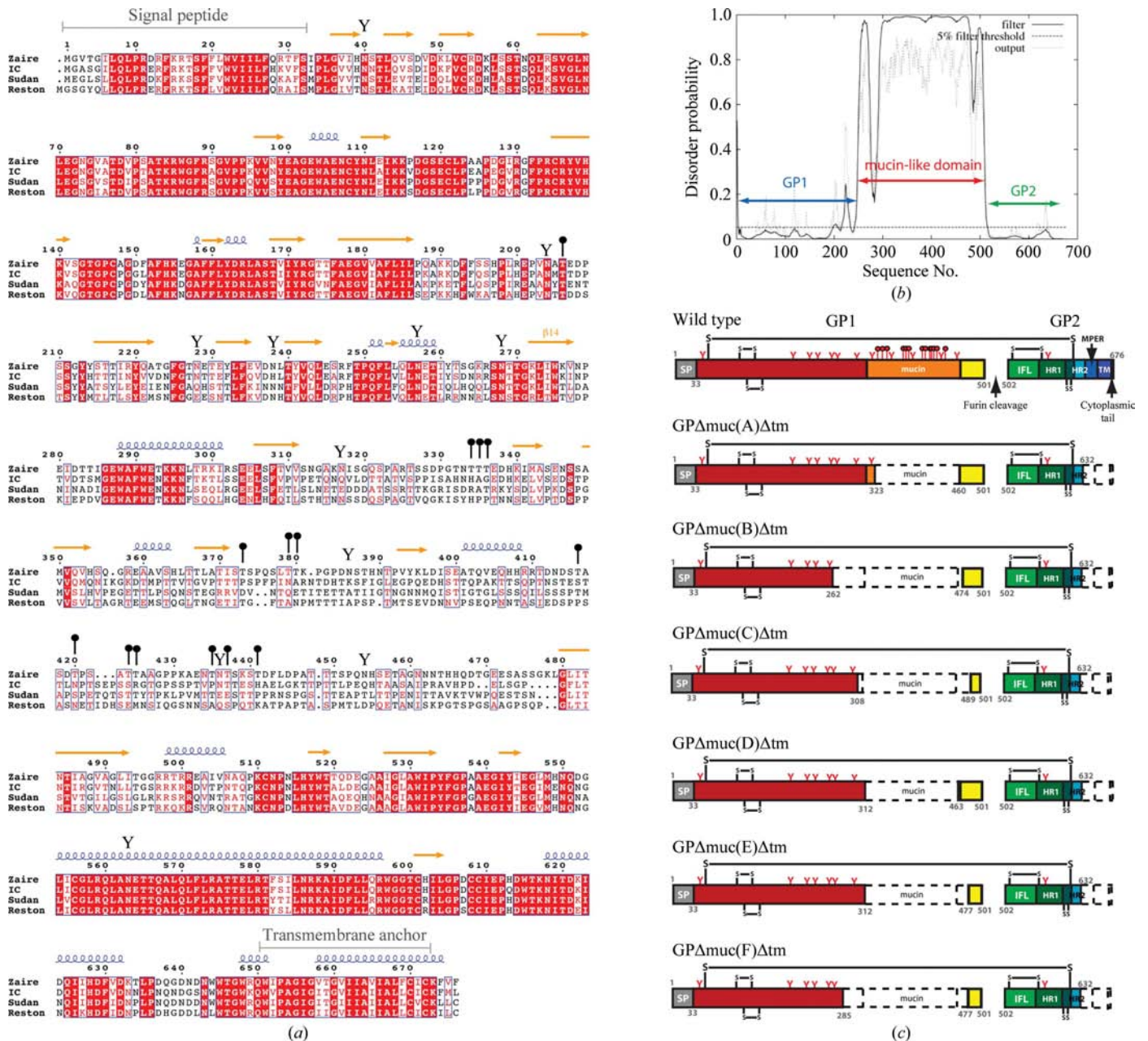
Figure 1
ZEBOV GP construct design. (*a*) *ClustalW* alignment of the glycoprotein primary sequence from the *Zaire*, *Sudan*, *Ivory Coast* and *Reston ebolavirus* species. Consensus secondary-structural prediction, using the *Network Protein Sequence Analysis* server (Combet *et al.*, 2000), is shown for *Zaire ebolavirus*. Helices and β-strands are shown as coils and arrows, respectively. N- and O-linked glycans predicted by the *NetNGlyc* (Gupta *et al.*, 2004) and *NetOGlyc* servers (Julenius *et al.*, 2005) are shown by Y and lollipop symbols, respectively. (*b*) Disordered domain prediction for ZEBOV GP using the *DISOPRED*2 server (Ward *et al.*, 2004). (*c*) Schematic representation of the ZEBOV GP and GP variants generated for expression screening. For clarity, we show only selected examples of mucin-like domain deletions. Disulfide bridges (-S-S-), signal peptide (SP), internal fusion loop (IFL), heptad-repeat region 1 (HR1), heptad-repeat region 2 (HR2), membrane-proximal external region (MPER), transmembrane anchor (TM) and cytoplasmic tail are labeled accordingly. N- and O-linked glycans are shown as red-colored Ys and lollipops, respectively.
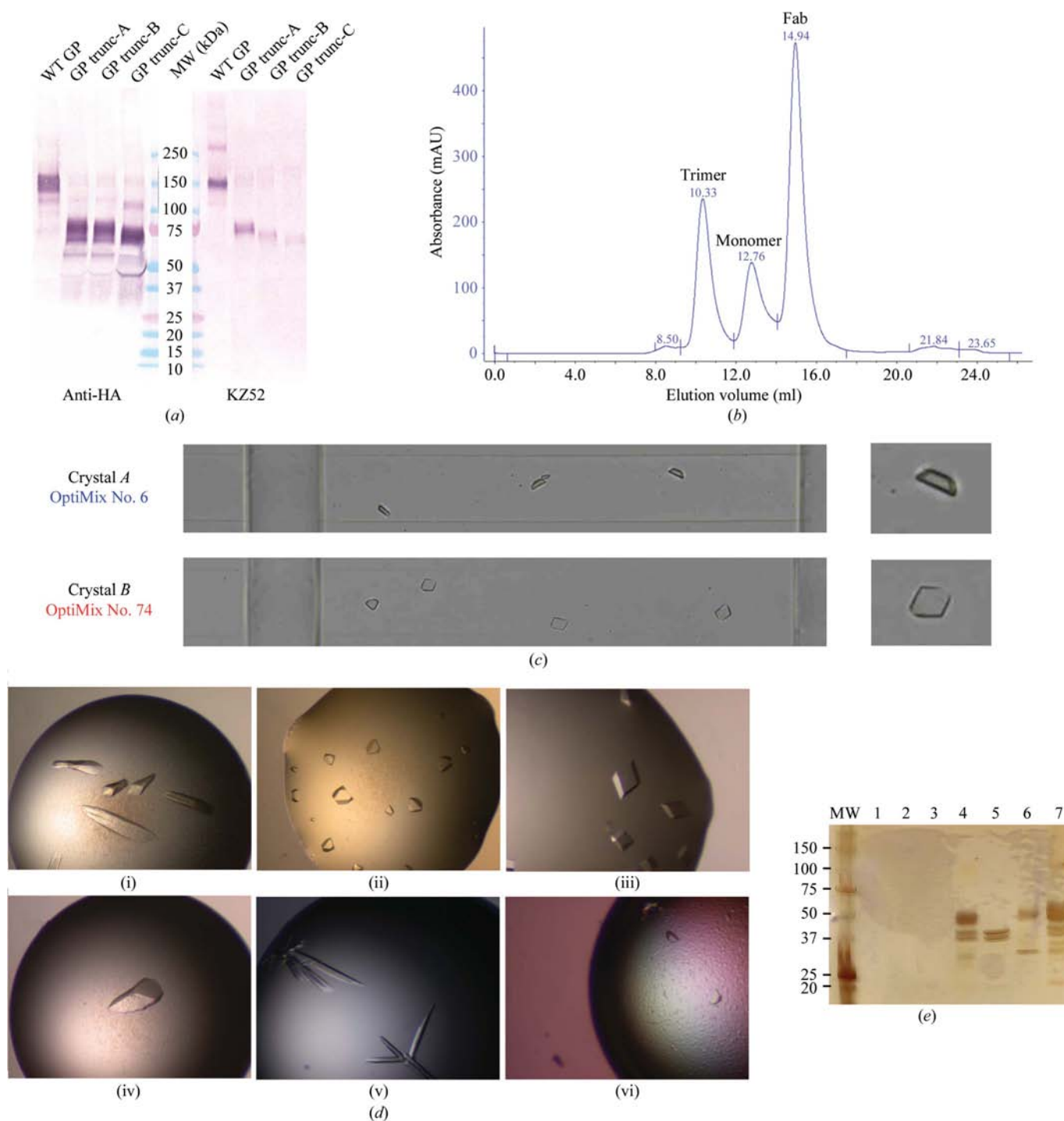
Figure 2
Expression, purification and crystallization of ZEBOV GPΔmuc$_{312-463}$Δtm. (a) Small-scale expression screening of selected GP truncation variants using a transient transfection HEK293T system. Conditioned media containing ZEBOV GP variants were harvested 4 d post-transfection and 10 μl supernatant was separated by nonreducing SDS–PAGE and probed with anti-HA (linear) or KZ52 (conformational) primary monoclonal antibodies. (b) ZEBOV GPΔmuc$_{312-463}$Δtm–KZ52 Superdex 200 10/300 GL chromatogram. (c) The top two selected free-interface diffusion crystallization hits. Crystal form A [OptiMix1 No. 6; 10% ethylene glycol, 10%(w/v) PEG 10 000 and 0.6 M sodium acetate] and crystal form B [OptiMix2 No. 74; 10%(w/v) PEG 6000, 0.1 M PIPES pH 6.5 and 0.6 M NaI] are shown. (d) Various crystal morphologies of ZEBOV GPΔmuc$_{312-463}$Δtm–KZ52 were obtained by hanging-drop vapour diffusion in various precipitants and additives and at various pH values: (i) rod crystals, pH 4.2, (ii) pyramidal crystals, pH 4.8, (iii) rhomboid crystals, pH 6.5, (vi) rod crystals, pH 5.0, with cetyl-trimethylammonium bromide (CTAB) additive, (v) rod crystals with cesium chloride and (vi) trapezoidal crystals with ethylene glycol. (e) A single crystal (∼0.2 × 0.2 × 0.2 mm) grown in 8.5%(w/v) PEG 10 000, 0.4 M sodium acetate, 0.1 M Tris–HCl pH 8.5 and 10%(v/v) PEG 200 was washed three times in mother liquor and then dissolved in nonreducing SDS–PAGE sample buffer. The washes (lanes 1–3), dissolved crystal (lane 4), Fab KZ52 (lane 5), PNGaseF-treated GPΔmuc$_{312-463}$Δtm (lane 6) and PNGaseF-treated GPΔmuc$_{312-463}$Δtm–KZ52 (lane 7) were analyzed by silver-stained SDS–PAGE. Note that Fab KZ52 migrates at a lower molecular weight (∼40 kDa) than the expected 50 kDa for typical antibody fragments and also exists as two isoforms (doublet bands) prior to MonoS ion-exchange purification.

rebuilding were performed using the programs *phenix.refine* (Adams *et al.*, 2002, 2004; Afonine *et al.*, 2005) and *Coot*. Refinement statistics are shown in Table 1.

2.4.5. Phase determination of GPΔmuc$_{312–463}$Δtm–KZ52. Molecular-replacement searches using our independently determined Fab KZ52 model were performed using *Phaser* (McCoy *et al.*, 2007), *CNSsolve* (Brünger *et al.*, 1998), *MOLREP* (Vagin & Teplyakov, 1997) and *EPMR* (Kissinger *et al.*, 1999, 2001). A single Fab KZ52 solution from *Phaser* (McCoy *et al.*, 2007), residing on the crystallographic threefold axis, was then used to generate a Fab KZ52 trimeric assembly for a second molecular-replacement search using *Phaser* and *CNSsolve*. Both programs scored a clear and identical hit for the trimeric Fab arrangement. An anomalous Fourier electron-density map using the trimeric KZ52 assembly model phases and the SAD Se peak data set was generated using *CNSsolve*. The maps were contoured at 3σ using *Xfit* (McRee, 1993) and a total of 20 selenium peaks were picked manually. Unimodal Hendrickson–Lattmann coefficients (HLA/HLB) were calculated using *SFTOOLS* in the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) from trimeric Fab KZ52 model phases calculated in *SFALL* (Agarwal, 1978) and scaled using *SIGMAA* (Read, 1986). Selenium anomalous data from two crystals and model Fab KZ52 phases were input into the program *SHARP* (Vonrhein *et al.*, 2007) for heavy-atom refinement and phasing. The phases calculated from *SHARP* were histogram-matched and averaged over 1500 cycles using threefold noncrystallographic symmetry (NCS) and density modification using the *CCP*4 suite program *DM* (Cowtan, 1994). Clear secondary-structural elements and solvent boundaries were observed in the initial experimental electron-density maps.

2.5. Model building and refinement

The program *RESOLVE* (Terwilliger, 2000, 2003) was used to automatically trace fragments into the NCS-averaged density-modified electron-density map. The chain directions of these fragments were subsequently used as starting points in manual chain extension and building using the program *Coot* (v.0.3.3; Emsley & Cowtan, 2004). In this process, alanine residues were only built into clear density of at least 1.3σ and idealized polyalanine helical segments were generated and fitted into helical density as rigid bodies. All β-strand and helical fragments were refined in real space with tight secondary-structural torsional restraints using *Coot*. This initial polyalanine model was subsequently refined with tight NCS restraints using torsion-angle simulated annealing (5000 K starting temperature) and a maximum-likelihood amplitude target in *CNSsolve* (v.1.2; Adams *et al.*, 1997; Brünger *et al.*, 1990, 1998). All data between 48 and 3.4 Å with no σ cutoffs were used in the refinements. After the refinement of the initial polyalanine fragments, a cross-building protocol was used to reduce model bias. The model was split into two coordinate files containing either Fab KZ52 plus GP1 or Fab KZ52 plus GP2. Model phases were then calculated for each of these files using *SFALL* and scaled with *SIGMAA*.

Updated model phases from Fab KZ52–GP1 or Fab KZ52–GP2 and selenium anomalous phases were recombined in *SHARP* (Vonrhein *et al.*, 2007) and subjected to density modification. Separate electron-density maps corresponding to phases from KZ52–GP1 and KZ52–GP2 were then generated. Fragments corresponding to GP1 were built into electron-density maps calculated from GP2 and KZ52 model phases, while fragments corresponding to GP2 were built into electron-density maps derived from GP1 and Fab KZ52 model phases. In addition, a series of *B*-value-sharpened electron-density maps were generated in *FFT* (Ten Eyck, 1973) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) with *B* values of −25, −50, −75, −100, −125, −150, −175 and −200 Å^2 to improve side-chain electron-density features. All *B*-value-sharpened electron-density maps were visually inspected for signs of improvement and maps with applied *B* values of −75 and −100 Å^2 were used to provide improved side-chain details. The *GlyProt* (Bohne-Lang & von der Lieth, 2005) server was used to generate idealized biantennary Man$_{3–5}$(GlcNAc)$_2$ cores, which were subsequently fitted into electron density. Rounds of simulated annealing with torsion-angle dynamics in the resolution range 48.4–3.4 Å with no σ cutoff using the programs *CNSsolve* and *phenix.refine* (Brünger *et al.*, 1990, 1998; Adams *et al.*, 2002; Afonine *et al.*, 2005) were alternated with manual refitting of the model with NCS-averaged σ$_A$-weighted *m*F_o − *D*F_c and 2*m*F_o − *D*F_c Fourier electron-density maps. The progress of rebuilding was monitored by the concomitant drop of R_{work} and R_{free}. For the final round of refinement, riding H atoms were added using *phenix.reduce* and their positions were energy-minimized without the X-ray term (25 iterations) prior to simulated-annealing refinement with NCS restraints and TLS refinement using *phenix.refine* (Adams *et al.*, 2002, 2004; Afonine *et al.*, 2005). Side-chain rotamer and peptide torsion angles were calculated and analyzed throughout the model-building process using the programs *PROCHECK* (Laskowski *et al.*, 1993) and *MolProbity* (Davis *et al.*, 2004). Oligosaccharide torsion angles and nomenclature were validated using the *pdb-care* (Lutteke & von der Lieth, 2004) server. The coordinates and structure factors for the T42V/T230V GPΔmuc$_{312–463}$Δtm–SeMet Fab KZ52 complex and unbound Fab KZ52 were deposited in the Protein Data Bank (Berman *et al.*, 2000) with accession codes 3csy and 3inu, respectively.

3. Results and discussion

3.1. Protein preparation

3.1.1. Construct design. ZEBOV GP is predicted to contain an overall secondary-structural content of 16% helices, 18% β-strands and 62% extended coil (Fig. 1*a*). There are also 11 N-linked and 14 O-linked glycosylation sites in ZEBOV GP, with the majority of oligosaccharides residing on a glycan-rich mucin-like domain at the C-terminal end of GP1, an N-terminal signal peptide (residues 1–32) and a C-terminal transmembrane anchor (residues 650–672). Disorder-prediction servers describe the mucin-like domain as having low

complexity (residues 250–500; Fig. 1*b*). In order to improve the expression of soluble and homogeneous protein amenable to crystallization, the disordered mucin-like domain and the hydrophobic membrane-proximal external region, trans-membrane and cytoplasmic (residues 632–676) domains were excised. Importantly, the mucin-like domain does not appear to be required for viral attachment or entry: pseudotyped virions with deletions of this domain are equally or somewhat more infectious than those carrying wild-type GP (Yang *et al.*, 2000; Manicassamy *et al.*, 2005; Medina *et al.*, 2003). Hence, deletions of this domain are unlikely to significantly alter the structure of the regions of GP critical for receptor binding or fusion. However, the boundaries of the mucin-like domain are not well defined and therefore construction of multiple dele-tion variants was required (Fig. 1*c*). In general, the start and end residues of the glycoprotein deletion were chosen to minimize the disruption of all predicted secondary-structural elements (*i.e.* α-helices and β-sheets). We made a total of $\sim$20 different mucin-like domain-deletion constructs.

3.1.2. Protein screening, large-scale expression and deglycosylation. All ZEBOV GP constructs were screened for expression on a small scale using the pDISPLAY vector and transient transfection of HEK293T cells in six-well culture plates. The pDISPLAY vector was chosen for its strong cytomegalovirus promoter, high copy numbers (suitable for making milligram quantities of DNA for use in large-scale expression) and high-affinity hemagglutinin (HA) tag for efficient protein capture and detection in dilute solutions. The use of HEK293T allows rapid screening of large numbers of constructs without having to wait to select stable transfectants or to build up baculoviral stocks. A more general protocol for the HEK293T screening and expression system is presented elsewhere (Lee *et al.*, 2009). The detection of dilute protein in the secreted media was performed by Western blot analysis using dual antibodies against conformational and linear epitopes. The conformation-dependent antibody KZ52 used in these studies was identified in a human survivor of the 1995 Kikwit outbreak (Maruyama, Parren *et al.*, 1999; Maruyama, Rodriguez *et al.*, 1999) and binds to a conformational GP1/ GP2-containing epitope. The linear antibody 16B12 used in these studies recognizes the influenza virus A HA purification tag. The use of these two antibodies allowed us to distinguish properly folded and intact GP from misfolded GP and released GP1. Among the constructs expressed, GPΔmuc$_{312–463}$-Δtm bound KZ52 as well as wild-type GP did (Fig. 2*a*), suggesting that this mucin-like domain deletion has not altered the overall fold. Large-scale recombinant expression of ZEBOV GPΔmuc$_{312–463}$$\Delta$tm using ten-layer CellSTACKs produced about 2 mg glycoprotein in total. >75% of the secreted protein was captured using anti-HA immunoaffinity resin. Purified ZEBOV GPΔmuc$_{312–463}$$\Delta$tm migrates at a molecular weight of $\sim$75 kDa based on nonreducing SDS–PAGE analysis. Comparison with the theoretical molecular weight of nonglycosylated GPΔmuc$_{312–463}$$\Delta$tm (52.4 kDa) suggests $\sim$20 kDa of attached carbohydrates on the purified GPΔmuc$_{312–463}$$\Delta$tm, consistent with the molecular weight of the seven predicted N-linked glycosylation sites.

Heterogeneous glycosylation of proteins expressed in mammalian cells is usually a major detriment to crystallization (Grueninger-Leitch *et al.*, 1996; Kwong *et al.*, 1999). Glycans can be cleaved between the asparagines and the innermost GlcNAc oligosaccharide under native conditions using peptide *N*-glycosidase F (PNGaseF). PNGaseF efficiently removed $\sim$20 kDa of glycans from an overnight digestion (see §3.2.4 for additional deglycosylation details). After deglycos-ylation and further purification by size-exclusion chroma-tography, we are typically left with $\sim$1 mg purified GPΔmuc$_{312–463}$$\Delta$tm or other GP-deletion variants. We chose to screen for crystallization by microfluidic free-interface diffusion, so that multiple constructs could be screened, each on a small purification scale. We set up microfluidic crystal-lization trials for both glycosylated and PNGaseF-treated mucin-deleted GP variants. However, no crystal hits were obtained for any mucin-deleted GP variant from a screen of 384 crystallization conditions.

3.1.3. Strategies for crystallization: antibody–GP com-plexes. The use of antibody fragments in cocrystallization has facilitated the determination of a number of challenging structures (Kovari *et al.*, 1995), including the KvAP ion channel (Jiang *et al.*, 2003), the HIV-1 gp120–CD4 complex (Kwong *et al.*, 1998), cytochrome *c* oxidase (Ostermeier *et al.*, 1997) and HIV-1 p24 (Prongay *et al.*, 1990). A number of conformation- and linear epitope-dependent antibodies are available for ZEBOV GP (Maruyama, Parren *et al.*, 1999; Takada *et al.*, 2003; Wilson *et al.*, 2000; Druar *et al.*, 2005; Shahhosseini *et al.*, 2007). Deglycosylated ZEBOV GPΔmuc$_{312–463}$$\Delta$tm was complexed with seven different antibodies that bound conformation-dependent epitopes. We reasoned that the use of conformation-dependent antibodies may improve the stability of the glycoprotein for crystal-lization. One of the more promising antibodies is KZ52, as this antibody requires both the GP1 and GP2 subunits for recog-nition (Maruyama, Parren *et al.*, 1999; Maruyama, Rodriguez *et al.*, 1999). All GPΔmuc$_{312–463}$$\Delta$tm–antibody complexes were screened by microfluidic free-interface diffusion. However, crystallization hits (39 positive conditions) were only obtained from the ZEBOV GPΔmuc$_{312–463}$$\Delta$tm–KZ52 complex (Fig. 2*b*) within a 48 h period. The top two crystal hits (Fig. 2*c*), characterized by the largest crystals of best morphology, were translated to traditional hanging-drop vapour-diffusion-based crystallization. Moderate-sized crystal wedges (0.15 × 0.15 × 0.15 mm) were obtained within a 48–72 h period by varying the drop size and the concentrations of the precipitant and additive components of the crystallization condition. We were able to obtain crystals in a variety of precipitants and at a variety of pH values (Fig. 2*d*). Washing and dissolving these crystals confirmed the unambiguous presence of both GPΔmuc$_{312–463}$$\Delta$tm and the KZ52 antibody fragment (Fig. 2*e*). In addition, it appears that certain glyco-forms of GPΔmuc$_{312–463}$$\Delta$tm may be forming the crystal lattice, as GPΔmuc$_{312–463}$$\Delta$tm from the dissolved crystal consisted of more homogeneous and lower molecular-weight protein than that present in the crystallization drop. Exposure of the GPΔmuc$_{312–463}$$\Delta$tm–KZ52 crystals on a Rigaku FR-D

X-ray home source for 20 min did not reveal any diffraction. Synchrotron X-rays (SSRL and ALS) improved the diffraction of these crystals to ~30 Å resolution. Room-temperature and cryocooled crystals diffracted to the same resolution limits, indicating that the poor diffraction was not a consequence of the cryoprotectant or freezing, but rather of internal disorder of the crystals. Decreasing the precipitant concentration and increasing the drop size allowed the growth of larger crystals (0.25 × 0.25 × 0.25 mm). These crystals diffracted to 15–20 Å resolution, but the quality of diffraction was highly variable between crystals. Further attempts to improve crystal diffraction by varying the pH, introducing additives, heavy atoms and detergents, optimizing cryoprotectants, using macroseeding, microseeding and streakseeding, controlling the humidity using the free-mounting system (Proteros/MSC) and/or cryo-annealing the crystals failed to improve the diffraction limits to better than 15 Å resolution.

3.2. Tactics to improve crystal diffraction

Although our glycoprotein sample had been treated with PNGaseF, matrix-assisted laser desorption ionization–time of flight (MALDI–TOF; Applied Biosystems, Foster City, California, USA) mass spectrometry revealed the presence of ~7.5 kDa of glycans remaining on GPΔmuc$_{312–463}\Delta$tm. We thought that incomplete deglycosylation of the protein could be impeding the formation of strong crystal contacts and the growth of well ordered crystals (Kwong et al., 1999). Hence, we employed a multi-pronged approach that involved a combination of (i) insect-cell expression, (ii) glycan anabolic pathway inhibitors, (iii) point mutants to delete N-linked glycosylation sites and (iv) chaotrope-assisted enzyme deglycosylation to control the extent of glycosylation.

3.2.1. Insect-cell expression.
Glycoproteins from mammalian cells are usually processed to generate complex-type oligosaccharides with terminal N-acetylneuraminic (sialic) acid. Proteins expressed in insect cells are typically paucimannose-type or oligomannose-type structures (Altmann et al., 1999; Hsu et al., 1997; Luckow, 1995), which are more homogeneous, uncharged, smaller and often more amenable to crystallization. SDS–PAGE analysis of GPΔmuc$_{312–463}\Delta$tm from stably transfected or baculovirus-infected T. ni (High Five) cells revealed a single migrating band with an apparent molecular weight of ~52−55 kDa depending on the purification tag attached (Fig. 3a). The theoretical molecular weight of nonglycosylated GPΔmuc$_{312–463}\Delta$tm is 51.4 kDa with a C-terminal 6×His tag and 52.4 kDa with an N-terminal HA tag, suggesting that insect cells produced a minimally glycosylated product containing ~1–4 kDa of carbohydrate. High-Five-expressed GPΔmuc$_{312–463}\Delta$tm appears to be properly cleaved by insect furin into GP1 and GP2 subunits, as evidenced by reducing SDS–PAGE analysis, and appears to be properly folded as evidenced by strong recognition by the conformational KZ52 antibody. The glycans attached to High-Five-expressed GPΔmuc$_{312–463}\Delta$tm are easily and almost completely removed by treatment with EndoF3 or EndoH glycosidase, as shown by SDS–PAGE analysis (data not shown). Complexes of both fully glycosylated and EndoF3-treated High Five GPΔmuc$_{312–463}\Delta$tm–KZ52 crystallized under conditions similar to those of HEK293T-expressed protein. The insect-cell-produced GPΔmuc$_{312–463}\Delta$tm led to a significant improvement in diffraction (6 Å), although we were unable to extend the diffraction of these crystals to better than 6 Å. One major drawback of insect cells or baculoviruses, however, is the length of time and the amount of labor needed to either select stable colonies or build up baculoviral stocks for every new construct. Moreover, bacu-
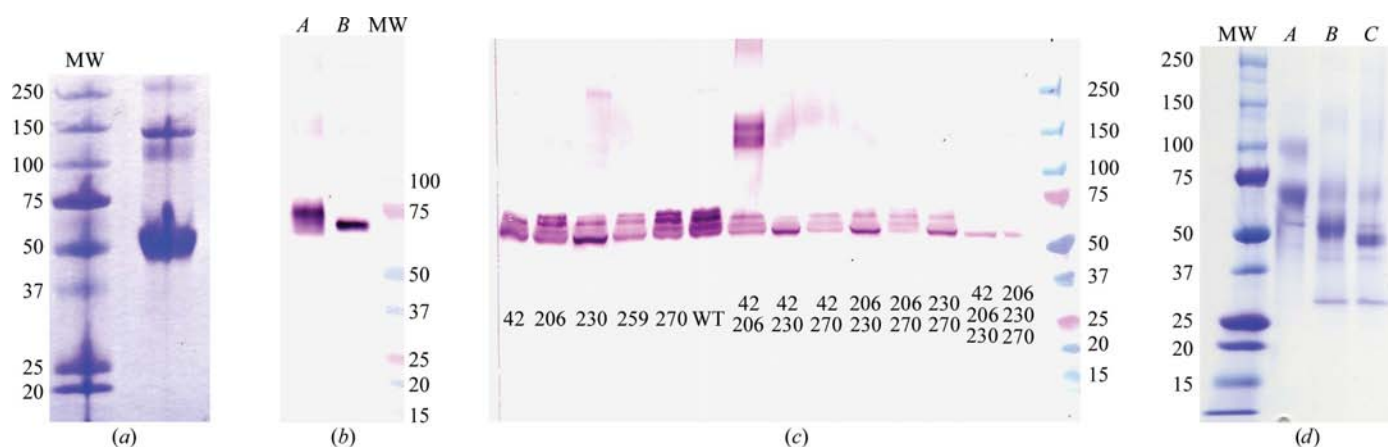

Figure 3
Control of N-linked glycosylation. (a) Ni-affinity purified GPΔmuc$_{312–463}\Delta$tm produced in T. ni insect cells (High Five cells). A fraction of the sample remains as a glycoprotein trimer (~150 kDa) even in the presence of SDS. (b) Nonreducing Western blots of kifunensine-treated HEK293T-produced GPΔmuc$_{312–463}\Delta$tm. Lane A, HEK293T-expressed fully glycosylated GPΔmuc$_{312–463}\Delta$tm; lane B, kifunensine-treated (1 µg ml^{-1} final concentration) GPΔmuc$_{312–463}\Delta$tm. The effect of the inhibitor on glycosylation was monitored by nonreducing immunoblot probed with a KZ52 primary mAb. (c) Nonreducing Western blots of selected single-site, double-site and triple-site sugar mutants. All samples were deglycosylated using PNGaseF and probed with a KZ52 primary antibody. The numbers on the gel refer to the sites of threonine residues in the N-X-T/S glycosylation motif that were mutated to valine. (d) Coomassie-stained SDS–PAGE analysis of the effect of urea-assisted PNGaseF deglycosylation. Lane A, HEK293T-expressed fully glycosylated GPΔmuc$_{312–463}\Delta$tm; lane B, PNGaseF-deglycosylated T42V/T230V GPΔmuc$_{312–463}\Delta$tm; lane C, PNGaseF deglycosylation of T42V/T230V GPΔmuc$_{312–463}\Delta$tm in the presence of 2.0 M urea.

lovirus infection of insect cells surprisingly produced less protein (∼0.8 mg per litre of culture) than HEK293T cells (1.5 mg per litre of culture). Hence, yields from transiently transfected HEK293T cells were comparable to, if not better than, the more established baculovirus/insect-cell expression methods.

3.2.2. Glycan anabolic pathway inhibitors. The use of anabolic inhibitors of glycosylation pathways represents an alternative method to minimize N-linked glycosylation. Kifunensine has been reported to be a potent α-mannosidase I inhibitor (Elbein *et al.*, 1990) and will result in the addition of oligomannose-type N-linked glycans [Man$_{5-9}$(GlcNAc)$_2$; Chang *et al.*, 2007]. Moreover, kifunensine is generally nontoxic to adherent HEK293T cells and does not adversely affect secretion or protein expression yields. The Man$_{5-9}$(GlcNAc)$_2$ glycans are much larger than those derived from insect cells, but like insect cell-produced glycoprotein can be removed using EndoH or EndoF glycosidases (Chang *et al.*, 2007). We have found that kifunensine concentrations as low as 1 µg ml^{-1} produced highly homogeneous glycoproteins, as judged by Western blot analysis (Fig. 3*b*). However, when glycosylated and deglycosylated kifunensine-treated ZEBOV GPΔmuc$_{312-463}\Delta$tm–KZ52 were crystallized, no improvement in diffraction was observed.

3.2.3. Point mutations. Given that insect cell-based expression and glycan-processing inhibitors failed to impart a significant improvement on diffraction, we made systematic point mutations to delete N-linked glycan sites by altering either end of the N-X-(T/S) sequon: Asn to Asp in addition to Ser to Ala or Thr to Val at each predicted N-linked site in GP1. We noted that alternate point mutations within a given N-linked sequon had different effects on folding and protein stability. For example, N238D resulted in properly folded protein, while T240V at the other end of the same sequon resulted in poorly folded insoluble protein. Importantly, single-site elimination of the glycan attached to either Asn40 (*via* a T42V mutation) or Asn228 (*via* a T230V mutation) improved GPΔmuc$_{312-463}\Delta$tm homogeneity (Fig. 3*c*). A double mutation eliminating both these sites (T42V/T230V) resulted in a more homogeneous sample and fortuitously increased protein yields by 50%. Crystals of T42V/T230V GPΔmuc$_{312-463}\Delta$tm–KZ52 were grown under previously described conditions and exhibited improved resolution limits, for human cell-produced protein, of ∼7–15 Å.

In addition to the T42V/T230V mutant, we also made 20 double, 15 triple, three quadruple and six quintuple glycosylation-site point mutants in order to determine whether an almost completely deglycosylated ZEBOV GPΔmuc$_{312-463}$-Δtm could be expressed. Quadruple (N40D/T230V/N238D/T259V) and quintuple [N40D/T230V/N238D/N204(D/A)/T259V] point mutants were expressed as homogeneous soluble proteins that were recognized by KZ52. However, these mutants were more unstable and when crystallized resulted in diffraction that was worse than that of the T42V/T230V double mutant. Although site-directed mutagenesis eliminates certain individual glycosylation sites without detriment, simultaneous introduction of many of these mutations can destabilize the protein, even when the mutations are fairly conservative. However, we show that it may not be necessary to remove all potential sugar sites by mutagenesis: deletion of one or two key sites may be sufficient to improve diffraction.

3.2.4. Chaotrope-assisted enzymatic deglycosylation. The use of chaotropes may perturb the local structure to allow PNGaseF to access these sterically hindered sites. In general, low concentrations of urea (<3 M) do not usually cause irreversible protein denaturation. Indeed, PNGaseF itself is stable in 2.5 M urea at 310 K for 24 h and still possesses ∼40% activity in 5 M urea (Maley *et al.*, 1989). A concentration series of urea (0.5, 1.0, 1.5, 2.0 and
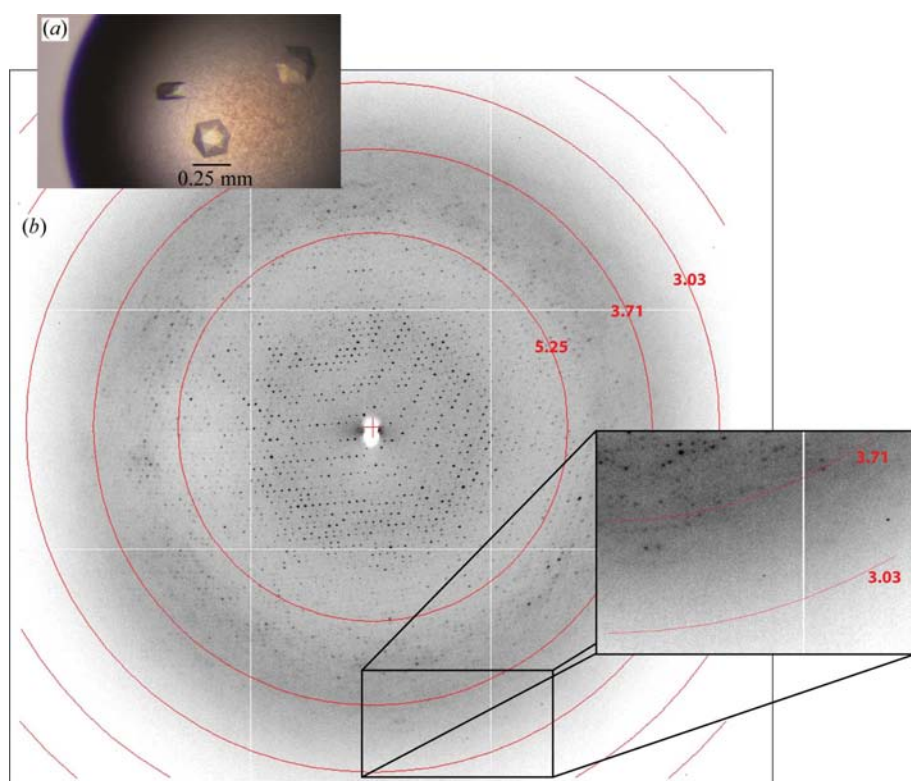

Figure 4
Diffraction of ZEBOV T42V/T230V GPΔmuc$_{312-463}\Delta$tm–SeMet KZ52 crystals. (*a*) Translated T42V/T230V GPΔmuc$_{312-463}\Delta$tm–SeMet Fab KZ52 crystals obtained by hanging-drop vapor diffusion in 8.5%(*w/v*) PEG 10 000, 0.4 M sodium acetate, 0.1 M Tris–HCl pH 8.5 and 10%(*v/v*) PEG 200. (*b*) Diffraction image of GPΔmuc$_{312-463}\Delta$tm–SeMet KZ52 collected on beamline 5.0.2, Advanced Light Source (Berkeley, California, USA) using an ADSC Quantum 315 CCD detector. Reflections are clearly visible to at least 3.5 Å resolution (inset box).

2.5 M) was incubated with PNGaseF and ZEBOV T42V/T230V GPΔmuc$_{312–463}\Delta$tm. Analysis by SDS–PAGE and MALDI–TOF mass spectrometry revealed the removal of an additional $\sim$2.5 kDa of glycans in 2 M urea (Fig. 3d), suggesting that urea treatment exposed one previously inaccessible site to PNGaseF. After deglycosylation, urea was removed by buffer exchange and the protein was analyzed by immunoblotting. The urea-assisted deglycosylated ZEBOV T42V/T230V GPΔmuc$_{312–463}\Delta$tm retained the ability to bind KZ52, suggesting that urea did not cause any irreversible global unfolding. Removal of the one extra oligosaccharide chain allowed more consistent diffraction of crystals to between 6 and 8 Å resolution and, in combination with the T42V/T230V double mutant, led to the identification of one crystal (0.3 × 0.3 × 0.3 mm) that diffracted to 4.0 Å resolution (Table 1).

3.3. Techniques for phase determination

The collection of a complete 4.0 Å resolution data set from the T42V/T230V GPΔmuc$_{312–463}\Delta$tm–KZ52 complex was a key turning point in the course of this project, as it brought us within the resolution range of a traceable map, allowing us to focus on obtaining initial phases. Traditionally, *de novo* structure determination involves either soaking heavy atoms into crystals or incorporating selenomethionine into samples in order to overcome the phase problem. Unfortunately, large weakly diffracting mammalian complexes offer technical obstacles to experimental phasing (see, for example, Fu *et al.*, 1999; Lowe *et al.*, 1995; Thygesen *et al.*, 1996). High-molecular-weight complexes typically require either a large number of heavy atoms or a large complex of heavy atoms in order to achieve a sufficient signal-to-noise ratio. In addition, larger protein complexes may be difficult to crystallize and only a small fraction of crystals grown will diffract well, further limiting the availability of crystals that are suitable for heavy-atom screening. Certainly for this structure the rarity of crystals that diffracted to better than 4.0 Å resolution ($\sim$1/250) complicated the empirical search for heavy-atom compounds for derivatization. Hence, we employed a multi-pronged approach involving Se-SAD, molecular replacement (MR) and phase combination to solve the phase problem.

3.3.1. Selenomethionine incorporation. Selenomethionine (SeMet) has become the anomalous scatterer of choice for experimental structure determination by multiwavelength anomalous dispersion (Hendrickson *et al.*, 1990). However, there is only one native methionine (Met548) left in the ZEBOV sequence after the removal of the initiating
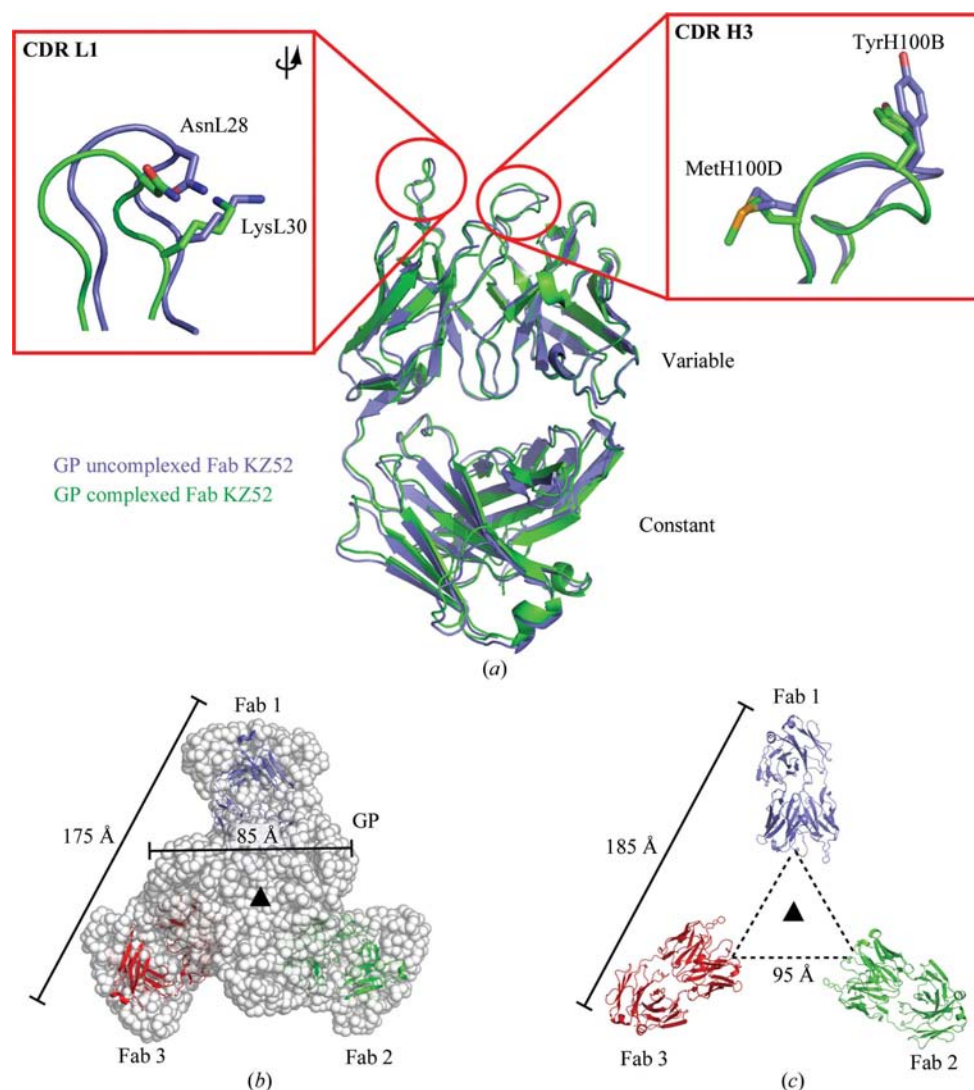

Figure 5
Molecular replacement using an improved search model: Fab KZ52. (a) Comparison of the crystal structures of the uncomplexed (blue; PDB code 3inu) and ZEBOV T42V/T230V GPΔmuc$_{312–463}\Delta$tm-bound Fab KZ52 (green; PDB code 3csy). The ZEBOV T42V/T230V GPΔmuc$_{312–463}\Delta$tm-mediated induced fit of side-chain and main-chain residues in the CDR L1 and CDR H3 regions are shown in the inset boxes. (b) *Ab initio* SAXS reconstruction of the T42V/T230V GPΔmuc$_{312–463}\Delta$tm–KZ52 molecular envelope docked with three Fabs (shown as blue, red and green ribbons). The knob-like structure in the middle of the SAXS molecular envelope and the space between the molecular-replacement arrangements of Fab KZ52, as shown by the dashed outline, are likely to correspond to the GP. (c) Molecular-replacement solution showing the arrangement of three Fab KZ52 molecules on the crystallographic threefold axis. The location of the threefold symmetry axis in both models is shown by the black triangle.

methionine by signal peptidase. The introduction of additional methionines by way of mutagenesis to leucine or isoleucine residues and subsequent expression following established protocols (Barton *et al.*, 2006) allowed the successful incorporation of up to five additional methionines into T42V/T230V GPΔmuc$_{312-463}\Delta$tm. However, the SeMet-incorporated T42V/T230V GPΔmuc$_{312-463}\Delta$tm had either poorer expression or decreased affinity for KZ52, suggesting a perturbation in the overall structure of ZEBOV T42V/T230V GPΔmuc$_{312-463}\Delta$tm. Selenocysteine incorporation was also attempted as T42V/T230V GPΔmuc$_{312-463}\Delta$tm contains several cysteine residues, but selenocysteine was highly toxic to the HEK293T cells. Hence, we took a different approach and incorporated selenomethionine into the KZ52 antibody instead of the glycoprotein. The KZ52 antibody fragment contains only five methionines, resulting in a ratio of one Se atom per 200 residues in the T42V/T230V GPΔmuc$_{312-463}\Delta$tm–KZ52 complex. Although, as expected, the KZ52 selenium signal was insufficient to phase the entire complex on its own, we hoped it could be used in combination with other forms of phasing. Given that only 15 selenium positions would be present per trimeric glycoprotein–antibody complex ($\sim$330 kDa), the anomalous signal was too low for facile detection. Indeed, attempts to locate the selenium positions using the programs *phenix.hyss* (Adams *et al.*, 2002; Grosse-Kunstleve & Adams, 2003), *SnB* (Weeks & Miller, 1999), *SHELXD* (Uson & Sheldrick, 1999), *CNSsolve* (Brünger *et al.*, 1998) and *SOLVE* (Terwilliger & Berendzen, 1999) failed in our hands and a manual analysis of Harker sections from an anomalous difference Patterson map did not reveal a consistent set of significant peaks greater than 3σ. However, to our surprise, T42V/T230V GPΔmuc$_{312-463}\Delta$tm–SeMet Fab KZ52 complex crystals exhibited a dramatic improvement in diffraction over non-SeMet-containing crystals. Data were collected from one of these crystals to 3.4 Å resolution (Fig. 4 and Table 1). SeMet incorporation into Fab KZ52 led to cell shrinkage of 15 Å in the *c* axis compared with the native

crystals. This allowed an increase of $\sim$1400 Å in the buried surface area of crystal contacts that are mediated primarily by the constant domains of the antibody fragment. The additional crystal-contact interactions are likely to explain the improved quality of diffraction of the T42V/T230V GPΔmuc$_{312-463}\Delta$tm–SeMet Fab KZ52 crystals.

3.3.2. Molecular replacement. Search models for molecular replacement could be derived from portions of two available crystal structures of a post-fusion ZEBOV GP2 fragment (Malashkevich *et al.*, 1999; Weissenhorn *et al.*, 1998) or from the antibody fragments bound to GP. Neither of the GP2 structures (in the probable post-fusion six-helix bundle conformation) nor the inner three helices encoding the first heptad-repeat region (HR1) yielded successful MR solutions, an early indication that GP2 was in a different conformation in our crystal. Instead, we attempted MR by screening 300 Fab structures covering a broad range of elbow angles and including several human framework sequences identical to that of KZ52. Although 300 search models were attempted, none yielded successful solutions whether used as intact Fab or broken into individual variable and constant domains.

3.3.3. Improving the molecular-replacement search model: structure of Fab KZ52. At the time, we thought that a more precisely matching Fab search model might improve the signal to noise and yield a successful solution. Hence, we independently crystallized and determined the structure of unbound Fab KZ52 at 2.5 Å resolution in order to use it as a search model (Table 1). We present the structure here. Overall, the framework of the unbound and ZEBOV GPΔmuc$_{312-463}\Delta$tm-bound Fab KZ52 structures do not differ significantly (Fig. 5*a*). The overall root-mean-squared deviation (r.m.s.d.) between all C$^\alpha$ atoms is 1.2 Å. However, residues TyrH100B and AsnL28, which belong to CDRs L1 and H3, respectively, undergo an induced-fit conformational change to improve contacts with the glycoprotein (Fig. 5*a*).

Using the unbound KZ52 structure as a search model, we identified a single Fab KZ52 located on the crystallographic
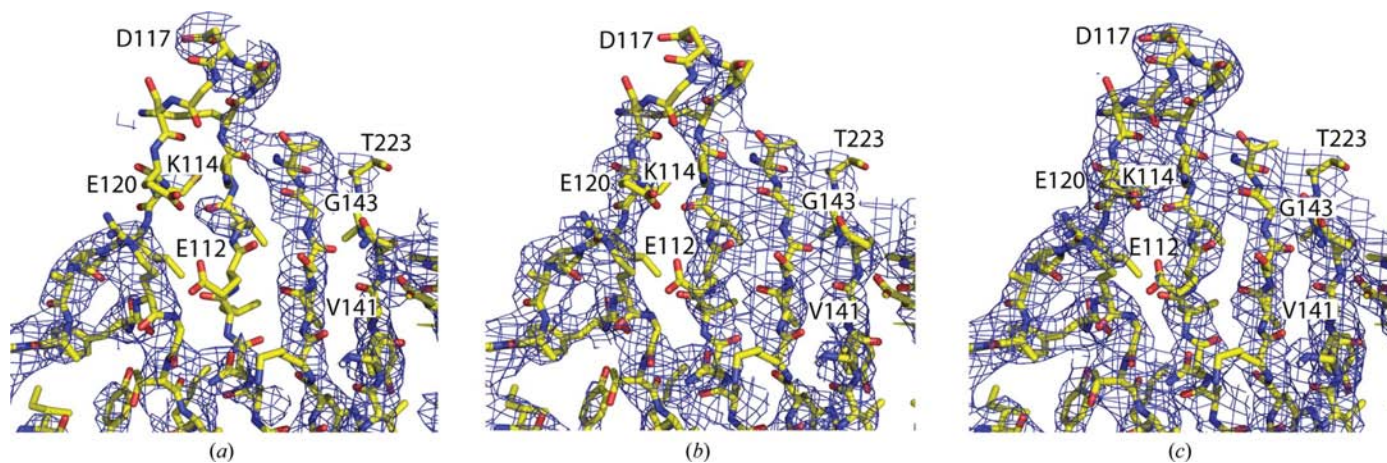

Figure 6
Initial electron-density maps. (*a*) Initial experimental histogram-matched density-modified electron-density map calculated in *SOLOMON* (Abrahams & Leslie, 1996). (*b*) Histogram-matched and NCS-averaged density-modified (1500 cycles) electron-density map calculated in *DM* (Cowtan, 1994). (*c*) Final simulated-annealed σ_A-weighted $2mF_o - DF_c$ electron-density map. The final refined model of ZEBOV GPΔmuc$_{312-463}\Delta$tm–SeMet Fab KZ52 (PDB code 3csy) is superimposed onto each of the electron-density maps. All electron-density maps are contoured at 1σ.

threefold axis using *Phaser* (McCoy *et al.*, 2007; *Z* score = 9.3). Additional solutions were expected as this first Fab plus the expected GP were unable to complete a crystal lattice. However, no additional Fab molecules were identified with the first solution fixed using the program *Phaser*. In addition, no clear solutions could be detected using the programs *CNSsolve* (Brünger *et al.*, 1998), *MOLREP* (Vagin & Teplyakov, 1997) or *EPMR* (Kissinger *et al.*, 1999, 2001). Subsequently, crystallographic symmetry operators were applied to the initial solution to generate a trimeric assembly of Fab KZ52. Molecular replacement of the trimeric Fab assembly using *Phaser* produced a clear solution (*Z* score = 10.4). Each asymmetric unit thus contains one and one-third trimers formed by one full trimeric complex plus a single T42V/T230V GPΔmuc$_{312-463}$$\Delta$tm–SeMet Fab KZ52 monomer unit residing on the threefold crystallographic axis, with a calculated solvent content of 65%.

3.3.4. Small-angle X-ray scattering analysis (SAXS). We used SAXS combined with three-dimensional reconstruction to complement our molecular-replacement efforts. The development of new SAXS modeling algorithms has significantly improved *ab initio* reconstructions of low-resolution macromolecular envelopes (Svergun, 1999; Svergun *et al.*, 1996, 2001; Walther *et al.*, 2000; Chacon *et al.*, 1998, 2000; Takahashi *et al.*, 2003). While these molecular envelopes are fairly low resolution (>10 Å), they still provide useful insights into the overall macromolecular size, architecture and assembly. Three-dimensional reconstructions of ZEBOV T42V/T230V GPΔmuc$_{312-463}$$\Delta$tm–KZ52, as illustrated in Fig. 5(*b*), show a central trimeric knob shape with dimensions of ~90 × 90 × 55 Å, which corresponds to the GPΔmuc$_{312-463}$$\Delta$tm. The base of this GP$\Delta$muc$_{312-463}$$\Delta$tm knob is surrounded by three flatter propeller-like regions which correspond to the bound Fab KZ52. The maximum diameter of the entire complex is ~175 Å. These dimensions were consistent with the MR solution of the trimeric Fab KZ52 assembly (Fig. 5*c*), thus giving us additional confidence in a correct MR solution.

3.3.5. Phase combination. Finding the Fab KZ52 MR solution was paramount to obtaining selenium heavy-atom positions from the SeMet KZ52 by way of anomalous difference Fourier electron-density maps. A total of 20 Se anomalous peaks (>4σ) were identified and manually picked. Superimposition of the peaks onto the previously determined native Fab KZ52 structure shows good agreement with the positions of the methionine sulfur. Phase combination with the selenium anomalous data and phases from the MR-derived trimeric arrangement of KZ52 antibody fragments was sufficient to phase the ~330 kDa trimeric T42V/T230V GPΔmuc$_{312-463}$$\Delta$tm–KZ52 complex. The resulting experimental electron-density maps revealed clear solvent boundaries and continuous main-chain density including clear helical bundles and β-sheet structures (Fig. 6). NCS-averaged density modification (1500 cycles) resulted in a 5% improvement in map correlation coefficients and visual inspection of this electron-density map revealed an extension of continuity in main-chain density, especially in β-sheet areas (Fig. 6).

3.3.6. Retrospective analysis. A retrospective analysis revealed that the model phases from the molecular-replacement solution of Fab KZ52 were the driving force for successful phasing. Comparison of initial experimental electron density (combined phases from the Fab KZ52 model and Se anomalous signal) with electron density calculated from the final refined model phases provides map correlation coefficients of 0.72 for the main chain and 0.44 for side chains. The electron-density map calculated using model phases from the Fab KZ52 molecular-replacement solution alone only produced an ~2% decrease in map correlation coefficients. Visual inspection also failed to identify any major differences in the two electron-density maps. This suggests that incorporation of Se atoms into the antibody fragment was not necessary for phasing, although it fortuitously improved resolution.

Given the importance of molecular replacement in the determination of the T42V/T230V GPΔmuc$_{312-463}$$\Delta$tm–SeMet Fab KZ52 complex, we now analyze our initial molecular-replacement trials to determine the reasons behind their failure and make recommendations on how to tackle future cases. We thought that the initial molecular-replacement search using 300 different antibody fragments from the PDB may have failed because a single Fab only accounts for a small percentage of the total scattering mass of the
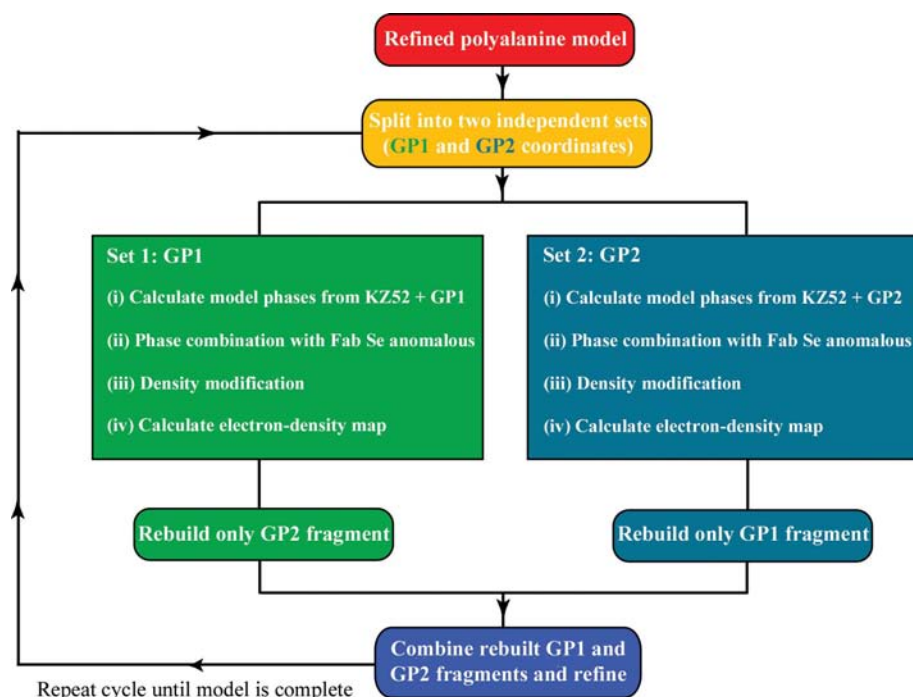

Figure 7
Flowchart of the cross-phasing model-building procedure.

440 kDa high-solvent T42V/T230V GPΔmuc$_{312-463}$Δtm–SeMet Fab KZ52 asymmetric unit and that differences in the elbow angle and sequence of the Fab search models may compromise the rotational and translational searches. Analysis of the SeMet Fab KZ52 bound to ZEBOV GPΔmuc$_{312-463}$Δtm reveals an elbow angle of 148°. This elbow angle is fairly common among IgG1 κ antibodies (mean = 156°, median = 150°; Stanfield *et al.*, 2006) and therefore should be well represented in our Fab search-model database. Analysis of molecular-replacement searches, in particular Fab search models with elbow angles of ∼148° (PDB codes 3hfm, 1yec, 1yee and 1ucb), revealed no clear solutions. In addition, there were no lower scored rotation or translation peaks that corresponded to a correct solution. This suggests that the differences in sequence between the search model Fab and KZ52 may have played a role in deteriorating the signal-to-noise ratio during the rotation and translation functions. As a test, we made a polyalanine model of the Fab with the 148° elbow angle (PDB code 3hfm) and used it as a search model in molecular replacement. We were unable to identify any clear solutions. However, molecular replacement of a polyalanine model of the GPΔmuc$_{312-463}$Δtm-bound SeMet Fab KZ52 resulted in an interpretable solution (*Z* score = 7.5). The overall root-mean-squared deviation between C^α atoms of the polyalanine models of 3hfm and KZ52 is 1.2 Å. This suggests that small differences in the loops and backbones of the search models may be the difference between a failed or successful solution, especially in cases of large macromolecular complexes where each Fab makes up a small percentage of the asymmetric unit. Therefore, in these cases we strongly recommend independent determination of the structure of the Fab to improve the signal-to-noise ratio for the rotation and translation functions when other search models have failed.

3.4. Tools for model building and refinement

Using the density-modified and NCS-averaged phases and *RESOLVE* (Terwilliger, 2000, 2003), we were able to trace ten fragments of around seven residues into each of the four GPΔmuc$_{312-463}$Δtm monomers in the electron-density map. The electron density was of sufficient quality to allow the building of a conservative polyalanine backbone for ∼50% of GPΔmuc$_{312-463}$Δtm. After torsion-angle simulated-annealing refinement, a significant drop from an R_{work} of 47% and an R_{free} of 50% to an R_{work} of 40% and an R_{free} of 43% was obtained. The model bias in our case was minimized by using the actual crystal structure of Fab KZ52 rather than a generic Fab from the PDB. As an additional precautionary measure, main-chain and side-chain building was performed using a cross-building protocol (Fig. 7) similar to the ping-pong method previously described by Hunt & Deisenhofer (2003). Here, the working model of GP was split into two coordinate files, consisting of residues from KZ52 and GP1 and from KZ52 and GP2. The KZ52/GP1 or KZ52/GP2 coordinates were used as a source of external phases and combined with the selenium anomalous phases and density modified. The updated KZ52/GP1 map was solely used to rebuild the GP2 and the KZ52/GP2 map was solely used to trace the GP1 subunit. The use of this cross-building procedure ensured minimal model bias in the areas of rebuilding.

The initial main-chain electron density was tube-like and side-chain density was poor owing to the modest resolution
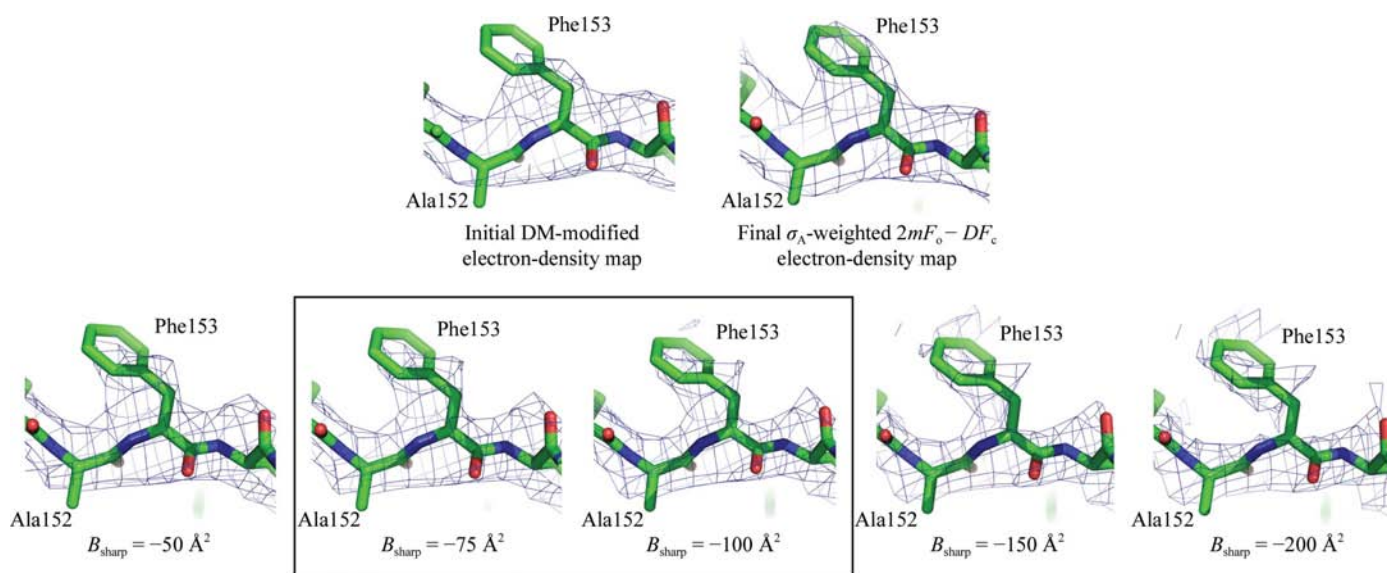

Figure 8
B-value-sharpened electron-density maps. *B*-value sharpening reveals additional side-chain electron-density features. A series of *B*-value-sharpened electron-density maps (B_{sharp} = −50, −75, −100, −150 and −200 Å^2) were generated in *FFT*. For comparison, the initial density-modified (DM) and final $σ_A$-weighted $2mF_o − DF_c$ electron-density maps are shown. All electron-density maps are superimposed with the final refined ZEBOV T42V/T230V GPΔmuc$_{312-463}$Δtm–SeMet Fab KZ52 coordinates (PDB code 3csy). The B_{sharp} = −75 Å^2 and B_{sharp} = −100 Å^2 electron-density maps were used to assist with model building, as these maps have improved side-chain densities and minimal background noise.

and poor initial phases of the T42V/T230V GPΔmuc$_{312–463}$Δtm–SeMet Fab KZ52 complex. The assignment and building of side-chain residues were partially assisted by *B*-value-sharpened electron-density maps ($F_{\text{sharpened map}}$), which involves the use of a negative *B* value, B_{sharp}, in a resolution-dependent weighting applied to a particular electron-density map (F_{map}) (1). *B*-value sharpening is a very useful tool for the enhancement of low-resolution electron-density maps (Bass *et al.*, 2002; DeLaBarre & Brunger, 2003, 2006),

$$F_{\text{sharpened map}} = \exp(-B_{\text{sharp}} \sin^2 \theta / \lambda^2) \times F_{\text{map}}. \qquad (1)$$

Each *B*-value-sharpened map was visually interpreted for signs of improved side-chain electron density (Fig. 8). As noise associated with higher resolution terms is also increased in this type of map, the best *B*-value-sharpened map was carefully chosen and used in combination with an unsharpened density-modified and σ_A-weighted $2mF_o - DF_c$ electron-density map for model rebuilding. We noticed that the $B_{\text{sharp}} = -75$ or -100 Å^2 electron-density maps have some improved features for aromatic residues and minimal noise (Fig. 8). The use of these *B*-value-sharpened electron-density maps helped to define the sequence registry.

Side-chain electron density and/or positions of heavy atoms, such as the Se atom in selenomethionine, are useful in confirming the proper sequence register. However, in our case we were limited by low resolution and the lack of any heavy-atom anchors. The proper sequence register was confirmed by secondary-structure predictions, anchoring N-linked glycosylation sequons to associated electron density and analysis of alternate chain registers. We found that peptide sequences of GP adopted their predicted secondary-structure features with ~73% accuracy. This is consistent with the reported accuracies (~81%) of three-state (helix, strand and coil) secondary-structural prediction algorithms (Cole *et al.*, 2008). Use of these features helped to combine multiple separate fragments

into a smaller subset of larger GP fragments and allowed all fragments to be assigned. At the same time, regions of electron density featuring dimensions and shapes consistent with those previously noted for the chitobiose core (Wormald *et al.*, 2002) were used to locate N-linked oligosaccharides and to confirm the proper register of the GP sequence. This was particularly important in the tracing of residues in the glycan-cap region. Alternate sequence registers were considered in multiple parts of ZEBOV T42V/T230V GPΔmuc$_{312–463}$Δtm, but these models could be eliminated based on inspection of σ_A-weighted $2mF_o - DF_c$ and $mF_o - DF_c$ electron-density maps. In addition, rotamer conformations and Ramachandran plots were constantly calculated throughout model building to scrutinize the stereochemistry of the model. The use of riding H atoms in refinement prevents nonphysical contacts at no cost to refinable parameters and in our case improved stereochemical properties and reduced Ramachandran outliers. The final model contains residues 33–189, 214–278, 299–310 and 502–599 (Fig. 9). No electron density is observed for residues 190–213, 311–312, 464–501 and 600–632. Weak or discontinuous electron density is seen in the loop containing the GP1–GP2 disulfide bridge (residues 49–56) and the outer regions of the GP1 glycan cap (residues 268–278 and 299–310). These regions are modeled as polyalanine fragments.

4. Conclusions and general lessons

The structural determination of the *ebolavirus* glycoprotein required us to overcome technical challenges in expression, crystallization, deglycosylation, phasing and model building. We highlight some of the general lessons and conclusions and also present an overview flowchart of approaches and decision-making branch points in Fig. 10 that we hope will be useful for structural biologists attacking a challenging new viral or mammalian glycoprotein target.

4.1. Construct design and expression

The use of widely available bioinformatics algorithms that predict secondary structure, N- and O-linked glycosylation sites, regions of disorder and transmembrane regions offer an effective guide to designing initial constructs.

Transient transfection of HEK293T cells and Western blot analysis allow rapid screening of large numbers of human or viral protein constructs. The use of a human cell line for expression ensures proper processing and native post-translational modifications.

The use of ten-layer CellSTACKs allows high-level protein production with yields comparable to recombinant protein expressed in stable or baculovirus-infected insect cells.

Figure 9
Structure of ZEBOV T42V/T230V GPΔmuc$_{312–463}$Δtm-SeMet Fab KZ52. The molecular surface of the ZEBOV T42V/T230V GPΔmuc$_{312–463}$Δtm-SeMet Fab KZ52 trimer is viewed from its side. Three lobes of GP1 (shades of blue) form a chalice and three GP2 subunits (white) wrap around GP1, forming a cradle. The three molecules of Fab KZ52 are depicted as yellow ribbons. This figure was adapted from the cover illustration in Lee *et al.* (2008).

4.2. Crystallization

Complexes with antibody fragments that bind conformation-dependent epitopes may stabilize intrinsically flexible multi-subunit proteins and offer new avenues for crystallization.
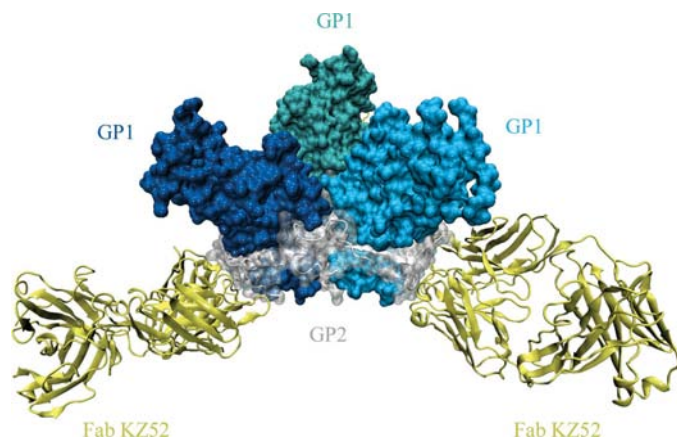
Microfluidic and/or other liquid-handling robotics allow the setup of nanolitre crystallization experiments, thus allowing the screening of multiple constructs and/or complexes with small amounts of protein.

4.3. Deglycosylation and improving diffraction

PNGaseF is effective in deglycosylating N-linked glycans under native-like conditions. However, some N-linked glycan sites are inaccessible to PNGaseF and result in incomplete deglycosylation, which may hinder the formation of tight crystal contacts.

The addition of kifunensine, a mannosidase I inhibitor, to HEK293T cell-culture media allows the expression of glycoproteins with smaller and more homogeneous attached N-linked glycans. Kifunensine provides an effective and easier alternative to expression in insect-cell lines.

Point mutations eliminating the N-X-S/T glycosylation motif is an alternate strategy and mutations of 1–2 glycan sites may be enough to improve diffraction.

Low concentrations of urea ($<3.0\,M$) may be used to perturb the glycoprotein structure to allow PNGaseF better access to more restricted glycosylation sites. This may be combined with the above strategies.

4.4. Phasing

It is possible to incorporate SeMet into recombinant human and viral proteins expressed in HEK293T cells.

Molecular replacement of antibody fragments is sensitive to the elbow angle between the constant and variable domains. Hence, in large macromolecular complexes where the Fab is a small scattering component in the asymmetric unit, even small deviations in the backbone of the search model (~1.0 Å r.m.s.d.) may deteriorate the rotation or translation function signals, leading to failure. When molecular replacement fails, we recommend determining the structure of the unbound Fab in order to improve the accuracy and signal-to-noise ratio of the search model.

Small-angle X-ray scattering combined with three-dimensional reconstruction allows the *de novo* determination of low-resolution molecular envelopes. These structures may be used to provide additional support for questionable molecular-replacement solutions.
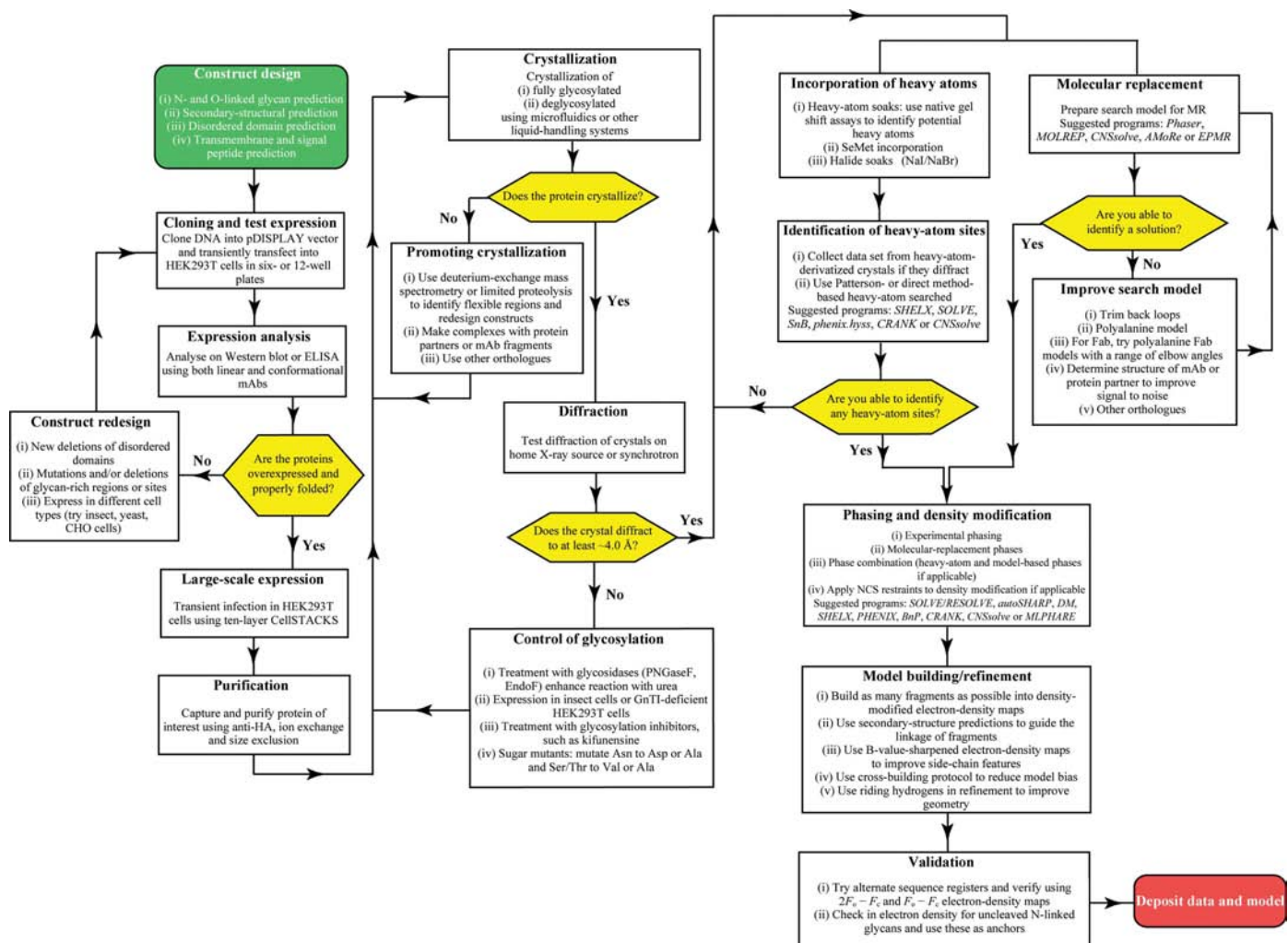

Figure 10
Flowchart of suggested tactics and strategies for tackling a challenging new viral or human glycoprotein.

4.5. Model building and refinement

Cross-building protocols minimize model bias during chain building.

The use of *B*-value-sharpened electron-density maps may reveal new side-chain densities.

In the absence of heavy-atom anchors, such as seleno-methionine residues, alternate sequence registers, secondary-structural predictions and N-linked glycosylation sites may be useful in determining and/or confirming the proper sequence register.

The use of riding H atoms in refinement is effective in preventing clashes and improving geometry while not adding any new parameters to refinement.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.

Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C. & Terwilliger, T. C. (2004). *J. Synchrotron Rad.* **11**, 53–55.

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.

Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.

Afonine, P. D., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl.*, contribution 8.

Agarwal, R. C. (1978). *Acta Cryst.* A**34**, 791–809.

Altmann, F., Staudacher, E., Wilson, I. B. & Marz, L. (1999). *Glycoconj. J.* **16**, 109–123.

Barton, W. A., Tzvetkova-Robev, D., Erdjument-Bromage, H., Tempst, P. & Nikolov, D. B. (2006). *Protein Sci.* **15**, 2008–2013.

Bass, R. B., Strop, P., Barclay, M. & Rees, D. C. (2002). *Science*, **298**, 1582–1587.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bohne-Lang, A. & von der Lieth, C. W. (2005). *Nucleic Acids Res.* **33**, W214–W219.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M.,

Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Brünger, A. T., Krukowski, A. & Erickson, J. W. (1990). *Acta Cryst.* A**46**, 585–593.

Chacon, P., Diaz, J. F., Moran, F. & Andreu, J. M. (2000). *J. Mol. Biol.* **299**, 1289–1302.

Chacon, P., Moran, F., Diaz, J. F., Pantos, E. & Andreu, J. M. (1998). *Biophys. J.* **74**, 2760–2775.

Chang, V. T., Crispin, M., Aricescu, A. R., Harvey, D. J., Nettleship, J. E., Fennelly, J. A., Yu, C., Boles, K. S., Evans, E. J., Stuart, D. I., Dwek, R. A., Jones, E. Y., Owens, R. J. & Davis, S. J. (2007). *Structure*, **15**, 267–273.

Cole, C., Barber, J. D. & Barton, G. J. (2008). *Nucleic Acids Res.* **36**, W197–W201.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Combet, C., Blanchet, C., Geourjon, C. & Deleage, G. (2000). *Trends Biochem. Sci.* **25**, 147–150.

Cowtan, K. (1994). *Int CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.

Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). *Nucleic Acids Res.* **32**, W615–W619.

DeLaBarre, B. & Brunger, A. T. (2003). *Nature Struct. Biol.* **10**, 856–863.

DeLaBarre, B. & Brunger, A. T. (2006). *Acta Cryst.* D**62**, 923–932.

Druar, C., Saini, S. S., Cossitt, M. A., Yu, F., Qiu, X., Geisbert, T. W., Jones, S., Jahrling, P. B., Stewart, D. I. & Wiersma, E. J. (2005). *Immunogenetics*, **57**, 1–9.

Elbein, A. D., Tropea, J. E., Mitchell, M. & Kaushal, G. P. (1990). *J. Biol. Chem.* **265**, 15599–15605.

Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. (2007). *Nature Protoc.* **2**, 953–971.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Feldmann, H., Volchkov, V. E., Volchova, V. A., Stroher, U. & Klenk, H. D. (2001). *J. Gen. Virol.* **82**, 2839–2848.

Fu, J., Gnatt, A. L., Bushnell, D. A., Jensen, G. J., Thompson, N. E., Burgess, R. R., David, P. R. & Kornberg, R. D. (1999). *Cell*, **98**, 799–810.

Grosse-Kunstleve, R. W. & Adams, P. D. (2003). *Acta Cryst.* D**59**, 1966–1973.

Grueninger-Leitch, F., D'Arcy, A., D'Arcy, B. & Chene, C. (1996). *Protein Sci.* **5**, 2617–2622.

Gupta, R., Jung, E. & Brunak, S. (2004). *Prediction of N-Glycosylation Sites in Human Proteins.* http://www.cbs.dtu.dk/services/NetNGlyc/.

Heckman, K. L. & Pease, L. R. (2007). *Nature Protoc.* **2**, 924–932.

Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.

Hsu, T. A., Takahashi, N., Tsukamoto, Y., Kato, K., Shimada, I., Masuda, K., Whiteley, E. M., Fan, J. Q., Lee, Y. C. & Betenbaugh, M. J. (1997). *J. Biol. Chem.* **272**, 9062–9070.

Hunt, J. F. & Deisenhofer, J. (2003). *Acta Cryst.* D**59**, 214–224.

Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T. & MacKinnon, R. (2003). *Nature (London)*, **423**, 33–41.

Julenius, K., Molgaard, A., Gupta, R. & Brunak, S. (2005). *Glycobiology*, **15**, 153–164.

Kall, L., Krogh, A. & Sonnhammer, E. L. (2007). *Nucleic Acids Res.* **35**, W429–W432.

Kingston, R. E., Chen, C. A. & Rose, J. K. (2003). *Curr. Protoc. Mol. Biol.* ch. 9, Unit 9.1.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.

Kissinger, C. R., Gehlhaar, D. K., Smith, B. A. & Bouzida, D. (2001). *Acta Cryst.* D**57**, 1474–1479.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.

Kovari, L. C., Momany, C. & Rossmann, M. G. (1995). *Structure*, **3**, 1291–1293.

Kuhn, J. H., Radoshitzky, S. R., Guth, A. C., Warfield, K. L., Li, W., Vincent, M. J., Towner, J. S., Nichol, S. T., Bavari, S., Choe, H., Aman, M. J. & Farzan, M. (2006). *J. Biol. Chem.* **281**, 15951–15958.

Kwong, P. D., Wyatt, R., Desjardins, E., Robinson, J., Culp, J. S., Hellmig, B. D., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1999). *J. Biol. Chem.* **274**, 4115–4123.

Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998). *Nature (London)*, **393**, 648–659.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lee, J. E., Fusco, M. H., Hessell, A. J., Oswald, W. B., Burton, D. R. & Saphire, E. O. (2008). *Nature (London)*, **454**, 177–183.

Lee, J. E., Fusco, M. H. & Saphire, E. O. (2009). *Nature Protoc.* **4**, 592–604.

Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. & Huber, R. (1995). *Science*, **268**, 533–539.

Luckow, V. A. (1995). *Baculovirus Expression Systems and Biopesticides*, edited by M. L. Shuler, H. A. Wood, R. R. Granados & D. A. Hammer, pp. 51–90. New York: John Wiley & Sons.

Lutteke, T. & von der Lieth, C. W. (2004). *BMC Bioinformatics*, **5**, 69.

Malashkevich, V. N., Schneider, B. J., McNally, M. L., Milhollen, M. A., Pang, J. X. & Kim, P. S. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 2662–2667.

Maley, F., Trimble, R. B., Tarentino, A. L. & Plummer, T. H. Jr (1989). *Anal. Biochem.* **180**, 195–204.

Manicassamy, B., Wang, J., Jiang, H. & Rong, L. (2005). *J. Virol.* **79**, 4793–4805.

Maruyama, T., Parren, P. W., Sanchez, A., Rensink, I., Rodriguez, L. L., Khan, A. S., Peters, C. J. & Burton, D. R. (1999). *J. Infect. Dis.* **179**, Suppl. 1, S235–S239.

Maruyama, T., Rodriguez, L. L., Jahrling, P. B., Sanchez, A., Khan, A. S., Nichol, S. T., Peters, C. J., Parren, P. W. & Burton, D. R. (1999). *J. Virol.* **73**, 6024–6030.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

McRee, D. M. (1993). *Practical Protein Crystallography*. San Diego: Academic Press.

Medina, M. F., Kobinger, G. P., Rux, J., Gasmi, M., Looney, D. J., Bates, P. & Wilson, J. M. (2003). *Mol. Ther.* **8**, 777–789.

Ostermeier, C., Harrenga, A., Ermler, U. & Michel, H. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 10547–10553.

Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.

Prongay, A. J., Smith, T. J., Rossmann, M. G., Ehrlich, L. S., Carter, C. A. & McClure, J. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 9980–9984.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Sanchez, A., Kiley, M. P., Holloway, B. P., McCormick, J. B. & Auperin, D. D. (1993). *Virus Res.* **29**, 215–240.

Sanchez, A., Yang, Z. Y., Xu, L., Nabel, G. J., Crews, T. & Peters, C. J. (1998). *J. Virol.* **72**, 6442–6447.

Saphire, E. O., Montero, M., Menendez, A., van Houten, N. E., Irving, M. B., Pantophlet, R., Zwick, M. B., Parren, P. W., Burton, D. R., Scott, J. K. & Wilson, I. A. (2007). *J. Mol. Biol.* **369**, 696–709.

Shahhosseini, S., Das, D., Qiu, X., Feldmann, H., Jones, S. M. & Suresh, M. R. (2007). *J. Virol. Methods*, **143**, 29–37.

Stanfield, R. L., Zemla, A., Wilson, I. A. & Rupp, B. (2006). *J. Mol. Biol.* **357**, 1566–1574.

Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.

Svergun, D. I. (1999). *Biophys. J.* **76**, 2879–2886.

Svergun, D. I., Petoukhov, M. V. & Koch, M. H. (2001). *Biophys. J.* **80**, 2946–2953.

Svergun, D. I., Volkov, V. V., Kozin, M. B. & Stuhrmann, H. B. (1996). *Acta Cryst.* A**52**, 419–426.

Takada, A., Feldmann, H., Stroeher, U., Bray, M., Watanabe, S., Ito, H., McGregor, M. & Kawaoka, Y. (2003). *J. Virol.* **77**, 1069–1074.

Takahashi, Y., Nishikawa, Y. & Fujisawa, T. (2003). *J. Appl. Cryst.* **36**, 549–552.

Ten Eyck, L. F. (1973). *Acta Cryst.* A**29**, 183–191.

Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.

Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Thygesen, J., Weinstein, S., Franceschi, F. & Yonath, A. (1996). *Structure*, **4**, 513–518.

Uson, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Volchkov, V. E., Feldmann, H., Volchkova, V. A. & Klenk, H. D. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 5762–5767.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.

Walther, D., Cohen, F. E. & Doniach, S. (2000). *J. Appl. Cryst.* **33**, 350–363.

Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. (2004). *Bioinformatics*, **20**, 2138–2139.

Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.

Weissenhorn, W., Carfi, A., Lee, K. H., Skehel, J. J. & Wiley, D. C. (1998). *Mol. Cell*, **2**, 605–616.

Wilson, J. A., Hevey, M., Bakken, R., Guest, S., Bray, M., Schmaljohn, A. L. & Hart, M. K. (2000). *Science*, **287**, 1664–1666.

Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.

Wormald, M. R., Petrescu, A. J., Pao, Y. L., Glithero, A., Elliott, T. & Dwek, R. A. (2002). *Chem. Rev.* **102**, 371–386.

Yang, Z. Y., Duckers, H. J., Sullivan, N. J., Sanchez, A., Nabel, E. G. & Nabel, G. J. (2000). *Nature Med.* **6**, 886–889.