Structural improvement of unliganded simian immunodeficiency virus gp120 core by normal-mode-based X-ray crystallographic refinement

**Xiaorui Chen,[a] Mingyang Lu,[b]
Billy K. Poon,[c]‡ Qinghua Wang[b]
and Jianpeng Ma[a,b,c]***

[a]Graduate Program of Structural and
Computational Biology and Molecular
Biophysics, Baylor College of Medicine, One
Baylor Plaza, BCM-125, Houston, TX 77030,
USA, [b]Verna and Marrs McLean Department of
Biochemistry and Molecular Biology,
Baylor College of Medicine, One Baylor Plaza,
BCM-125, Houston, TX 77030, USA, and
[c]Department of Bioengineering, Rice University,
Houston, TX 77005, USA

‡ Current address: Lawrence Berkeley National
Laboratory, 1 Cyclotron Road, Mail Stop
64R0121, Berkeley, CA 94720, USA.

Correspondence e-mail: jpma@bcm.tmc.edu

The envelope protein gp120/gp41 of simian and human immunodeficiency viruses plays a critical role in viral entry into host cells. However, the extraordinarily high structural flexibility and heavy glycosylation of the protein have presented enormous difficulties in the pursuit of high-resolution structural investigation of some of its conformational states. An unliganded and fully glycosylated gp120 core structure was recently determined to 4.0 Å resolution. The rather low data-to-parameter ratio limited refinement efforts in the original structure determination. In this work, refinement of this gp120 core structure was carried out using a normal-mode-based refinement method that has been shown in previous studies to be effective in improving models of a supramolecular complex at 3.42 Å resolution and of a membrane protein at 3.2 Å resolution. By using only the first four nonzero lowest-frequency normal modes to construct the anisotropic thermal parameters, combined with manual adjustments and standard positional refinement using *REFMAC*5, the structural model of the gp120 core was significantly improved in many aspects, including substantial decreases in R factors, better fitting of several flexible regions in electron-density maps, the addition of five new sugar rings at four glycan chains and an excellent correlation of the B-factor distribution with known structural flexibility. These results further underscore the effectiveness of this normal-mode-based method in improving models of protein and nonprotein components in low-resolution X-ray structures.

1. Introduction

Atomic structures of large biomolecular assemblies determined by X-ray crystallography are vitally important to the understanding of their functions. However, many large assemblies contain highly flexible structural components that often undergo anisotropic deformations and consequently result in crystals that only diffract to limited resolution. Conventional methods for X-ray structural refinement are not optimized for dealing with highly flexible structures or structural components, especially at lower resolutions owing to the rather low data-to-parameter ratio.

It has been well documented that a small set of low-frequency normal modes, as collective variables, can effectively approximate the overall anisotropic motion of structures [for reviews, see Ma (2004, 2005) and references therein]. In fact, normal modes have been used to work with X-ray data extensively in the literature for various purposes (Diamond, 1990; Kidera & Go, 1990, 1992; Kidera *et al.*, 1992*a,b*, 1994; Suhre & Sanejouand, 2004; Lindahl *et al.*, 2006; Delarue & Dumas, 2004; Kundu *et al.*, 2002; Kondrashov *et al.*,

2006, 2007; Schroder *et al.*, 2007). However, the successful application of normal modes to anisotropic temperature *B*-factor refinement was severely hindered by the initial energy minimization required in conventional normal-mode analysis, which causes structural models to move out of electron densities. It was only after the development of a new elastic normal-mode analysis that delivers accurate eigenvectors for low-frequency modes without initial energy minimization (Lu *et al.*, 2006; Lu & Ma, 2008) that we eventually succeeded in producing a normal-mode-based anisotropic *B*-factor refinement method to help with structural refinement of large flexible systems at limited resolution (Poon *et al.*, 2007; Chen *et al.*, 2007).

In our previous studies (Poon *et al.*, 2007; Chen *et al.*, 2007), normal-mode-based refinement was found to be effective for modeling anisotropic deformations of biomolecules with a substantially smaller number of parameters than even conventional isotropic refinement. The reduction of independent thermal parameters is achieved by using a small set of low-frequency normal modes, generally fewer than 50, to describe structural deformations collectively and anisotropically. The anisotropic thermal parameters of each atom are calculated from these low-frequency modes and used to replace the original isotropic *B* factors during structural refinement.

Simian and human immunodeficiency viruses (SIV and HIV, respectively) are pathogens that cause autoimmune deficiency syndrome in primates (Wyatt *et al.*, 1998). They both use the envelope glycoprotein gp160 for host-cell invasion, which is cleaved into gp120 and gp41 upon arrival at the surface of an infected cell (Allan *et al.*, 1985; Veronese *et al.*, 1985; Center *et al.*, 2002). The protein gp120 mainly serves to recognize host receptors, while gp41 is primarily involved in membrane fusion. The binding of gp120 to its CD4 receptor triggers a series of conformational changes that facilitate the binding of a co-receptor such as CXCR4 or CCR5 (Dalgleish *et al.*, 1984; Feng *et al.*, 1996; Trkola *et al.*, 1996; Wu *et al.*, 1996). As a protective mechanism for SIV and HIV to evade recognition by the host immune system, gp120 is heavily glycosylated. The glycosylation of gp120 severely limits the ability of the protein to form well ordered crystals. It was only after long, laborious and challenging work that the structure of an unliganded and fully glycosylated SIV gp120 core was solved to a resolution of 4.0 Å (PDB code 2bf1; Chen *et al.*, 2005*a*,*b*). This structure was refined using both *CNS* (Brünger *et al.*, 1998) and *REFMAC*5 (Murshudov *et al.*, 1997) in *CCP*4 (Collaborative Computational Project, Number 4, 1994) and employed TLS refinement for *B*-factor modeling (Schomaker & Trueblood, 1968; Howlin *et al.*, 1993). The final structural model contained sugars at all 13 glycosylation sites and had an R_{cryst} of 38.5% and an R_{free} of 38.8% (Chen *et al.*, 2005*a*,*b*). This structure determination has a very low data-to-parameter ratio: for 3086 non-H atoms in an asymmetric unit there were only 5842 unique reflections in the resolution range 4.0–26.0 Å (Chen *et al.*, 2005*a*,*b*). In the original structure determination, this problem was partially overcome by using heavily restrained refinement for atom positions and one-group TLS refinement for *B* factors. By comparing this unliganded gp120 structure with a previously

determined gp120–CD4 structure (Kwong *et al.*, 1998), it was found that binding of CD4 caused a backbone movement of up to 28 Å in its inner domain (Chen *et al.*, 2005*a*; Fig. 1).

In this study, by using our normal-mode-based refinement method and a new branch normal-mode analysis technique for handling long branched oligosaccharides, the structural model of the gp120 core was improved substantially. The final model has much lower R_{cryst} and R_{free} factors. An improved fit of several flexible regions in the electron density were achieved, especially in the inner domain that is involved in the conformational changes upon CD4 binding (Chen *et al.*, 2005*a*; Kwong *et al.*, 1998). Five new sugar rings at the termini of four glycan chains were added in newly appeared electron density. We expect that this normal-mode-based refinement method will find ample applications in improving models of protein and nonprotein components in low-resolution crystal structures of biomolecules.

2. Methods

2.1. Elastic normal-mode analysis and a new branch normal-mode analysis

The normal-mode vectors were calculated by an extended version of the elastic network model recently developed in our laboratory (Lu *et al.*, 2006). Conventional elastic normal-mode analysis (Atilgan *et al.*, 2001) is vulnerable to a tip effect in the low-frequency eigenvectors. Our method minimizes the tip effect by strengthening local stiffness using additional angular
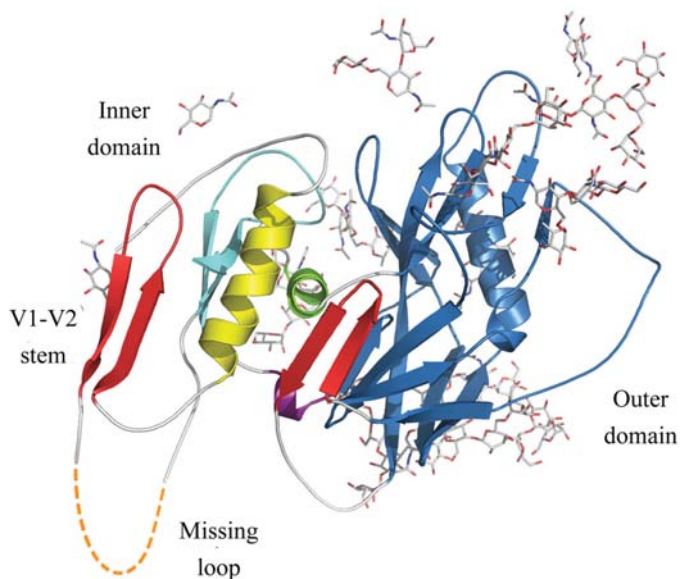

Figure 1
The original structure of the unliganded SIV gp120 core (PDB code 2bf1), with polypeptide chains shown as a ribbon diagram and sugars as stick models, with the same color scheme as in the original publication (Chen *et al.*, 2005*a*). The outer domain is colored blue, while the inner domain is colored according to substructure: yellow for the α1 helix, cyan for the β5–β7–β25 sheet, purple for the outer/inner domain transition, green for the α5 helix and red for the bridging sheet. The V1–V2 stem in the bridging sheet is labelled. The missing loop (residues 220–228) connecting the V1–V2 stem and the β5–β7–β25 sheet is shown as a dashed line in orange.

harmonic terms in the potential function. The modal analysis was then performed in the internal coordinate system.

For a molecule represented by n nodes (see the next paragraph for details of node selection), the potential function can be expressed as

$$V = (\gamma/2) \sum_i \sum_j \sigma_{ij} (|\mathbf{r}_{ij}| - |\mathbf{r}_{ij}^0|)^2 = (\omega/2) \sum_\alpha (\varphi_\alpha - \varphi_\alpha^0)^2.$$

The first part of the potential function is the same as in the conventional elastic normal-mode method, where $|\mathbf{r}_{ij}|$ and $|\mathbf{r}_{ij}^0|$ are the instantaneous and equilibrium distances between the ith and the jth nodes and σ_{ij} is a step function with a cutoff at r_c. The second part of the potential function is composed of additional angular terms, where φ_α and φ_α^0 are the instantaneous and equilibrium values of an internal coordinate angle indexed by α. In our calculations, the internal coordinate angles include all the pseudo-bond angles formed by three consecutive bonded nodes and all the pseudo-dihedral angles formed by four consecutive bonded nodes. These angles, together with the virtual bonds that connect pairs of chains, represent all the degrees of freedom in the internal coordinate system. The weights for the two energy terms are γ and ω, respectively, where $\omega = \xi \min(H_{\alpha\alpha}^0)$. Here, $H_{\alpha\alpha}^0$ is the diagonal element of the Hessian matrix for the first potential term (from the conventional method) calculated in internal coordinates and ξ is an adjustable parameter empirically chosen from 3 to 100.

In this study, we extended our method to macromolecules that contain branched sugar chains. In this derivative method, the C^α atoms of all residues and the geometric centers of all sugars were selected as nodes and the network topology was determined by chemical bond connectivity. As there were branches that were internal to the sugars and branches between the sugars and the protein backbone, the nodes formed a tree-like network structure. For a noncyclic tree network, all nodes were uniquely numbered according to the following rules: (i) amino acids were sequentially numbered from the N-terminus to the C-terminus and (ii) at branch points in the network, nodes in the short branch were numbered first (for branches of equal length, the choice was arbitrary). In this tree structure, each node except for the first node had only one connected node preceding it in the numbered list. Hence, the pseudo-bond and pseudo-dihedral angles for each node were defined based on the two preceding bonded nodes and the three preceding bonded nodes, respectively. The modal analysis in the internal coordinate system followed the procedures given in the literature (Kamiya *et al.*, 2003; Go *et al.*, 1983).

2.2. Positional refinement and manual model adjustment

The theory and details of the normal-mode-based refinement method have been reported in previous publications (Poon *et al.*, 2007; Chen *et al.*, 2007). In addition to normal-mode-based thermal parameters, one TLS group (corresponding to 20 independent refinement parameters) and one TLS scaling factor were generally used (Poon *et al.*, 2007; Chen *et al.*, 2007). The anisotropic *B* factors generated by normal-

Table 1
Refinement statistics of the original model and the normal-mode model of the SIV gp120 core.

	Original model†	Final normal-mode model	TLS on final normal-mode model
R_{cryst} (%)	39.1	34.6	34.7
R_{free} (%)	39.8	35.4	35.7
R.m.s.d. bond lengths (Å)	0.014	0.011	
R.m.s.d. bond angles (°)	2.067	2.055	
Ramachandran plot			
Core (%)	62.1	62.5	
Allowed (%)	29.8	29.7	
Generously allowed (%)	5.9	7.8	
Disallowed (%)	2.2	0.0	

† The original model did not contain the TLS-derived anisotropic *B* factors, which is why its calculated *R* factors from *REFMAC*5 were slightly higher than those reported in the original papers.

mode refinement were used to replace the isotropic *B* factors in the original model, which was then subjected to refinement using *REFMAC*5 v.5.2.0019 with very tight geometric restraints to update the atomic coordinates of the original structural model. The same set of *REFMAC*5 refinement parameters was used throughout the entire normal-mode-based refinement. Similar to the original structural refinement (Chen *et al.*, 2005*a*,*b*), no refinement of individual residual B_{iso} was allowed. Using the new structural model output from *REFMAC*5, a new composite OMIT $2F_o - F_c$ map was calculated to guide manual adjustments in *O*. Atoms were added only when suggested by clear and strong electron densities in the new composite OMIT $2F_o - F_c$ map. The R_{cryst} and R_{free} factors were monitored throughout. For calculation of R_{cryst} and R_{free}, the following equation was used:

$$R = \frac{\sum_{\mathbf{h}} \left| |F_{obs}(\mathbf{h})| - |F_{calc}(\mathbf{h})| \right|}{\sum_{\mathbf{h}} |F_{obs}(\mathbf{h})|}.$$

The same set of free reflections (5%) used in the original structure determination was saved for calculation of the R_{free} factor. The new model converged after five iterations of normal-mode-based refinement, *REFMAC*5 refinement and manual adjustment.

For comparison with the isotropic *B*-factor profiles in the 'original' model (Fig. 4*a*), the anisotropic *B*-factor profile of the normal-mode model was converted to isotropic *B* factors by averaging the diagonal terms of the anisotropic displacement parameters for each atom.

3. Observations

3.1. Normal-mode-based refinement of gp120 core

Structural refinement using different refinement programs tends to yield slightly different values for *R* factors and later versions of the same refinement programs may yield slightly better *R* factors than earlier versions. In order to make a fair comparison of normal-mode-based refinement with other refinement schemes, we generally re-minimized the PDB

structural models using *REFMAC*5 (Murshudov *et al.*, 1997) in *CCP*4 (Collaborative Computational Project, Number 4, 1994) with a set of optimized parameters (Poon *et al.*, 2007; Chen *et al.*, 2007). In this current study, the model obtained from the

PDB (2bf1; Fig. 1; Chen *et al.*, 2005*a*,*b*) was first subjected to further positional refinement using *REFMAC*5 with very tight geometric restraints. However, further positional refinement using *REFMAC*5 resulted in higher R_{free} factors than the
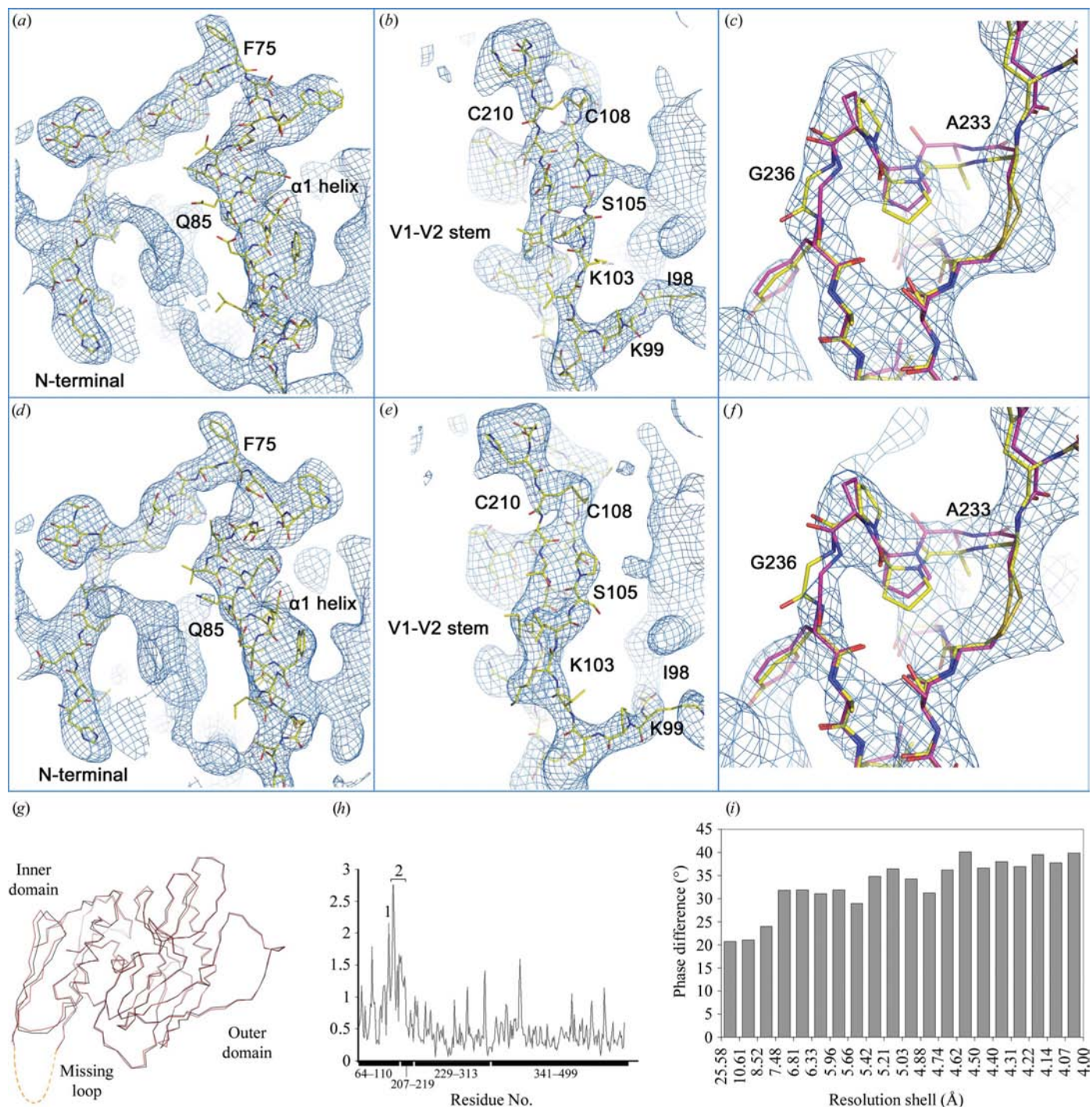

Figure 2
Structural improvement of the normal-mode model and comparison with the original model. In (*a*)–(*f*) the model in the original $2F_o - F_c$ OMIT map (*a*–*c*) is compared with the normal-mode model in the final $2F_o - F_c$ OMIT map (*d*–*f*). (*a*) and (*d*) N-terminus/α1-helix region, (*b*) and (*e*) the V1–V2 stem and (*c*) and (*f*) the region preceding the β5–β7–β25 sheet; In (*c*) and (*f*) C atoms are colored yellow for the original model and purple for the normal-mode model. All maps were contoured at 1.0σ. Images were prepared with *PyMOL*. (*g*) Superposition of the backbones of the original model (red) and the normal-mode model (black). (*h*) The r.m.s.d. between the backbones of the original and normal-mode models as a function of the residue number (*x* axis). The regions with large r.m.s.d. are labeled: the loop preceding the V1–V2 stem (peak 1, residues 97–101), the V1–V2 stem (peak 2, residues 102–213, of which residues 113–208 were deleted in the original construct). (*i*) Phase-angle shifts of the normal-mode model in relation to the original model as a function of resolution.

initial model from the PDB; the refinement was therefore stopped and the initial model was treated as the 'original' model in this article and is used for comparison with the normal-mode-refined model. It is important to note that although one-group TLS was used in the original structure

determination of gp120 (Chen *et al.*, 2005*a,b*), the details of the TLS grouping were unavailable in the PDB. This may explain why the recalculated R values of the original model, at 39.1% for the R_{cryst} factor and 39.8% for the R_{free} factor, were higher than those reported in the original papers (Chen *et al.*,
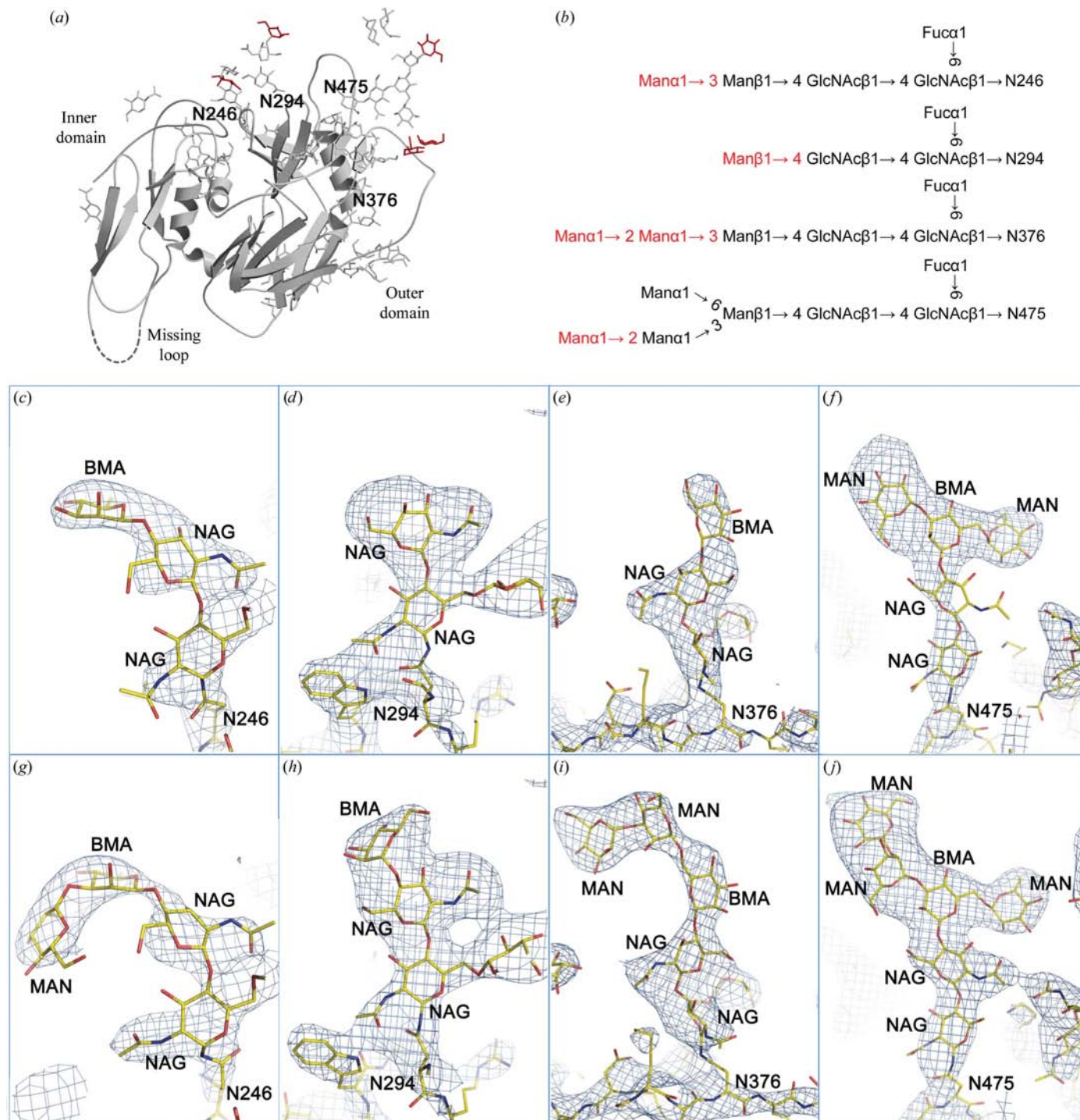

Figure 3
Newly added sugars. (*a*) The overall structure of the unliganded SIV gp120 core, with polypeptide chains shown as ribbon diagrams and sugars as stick models. The newly added sugars are colored red and the residues to which the new sugars anchor are labeled, with the rest of structure in white. (*b*) The specific type, glycosylation site and mode of linkage of glycan chains with newly added sugars (red). (*c*)–(*j*) Newly added sugars. (*c*)–(*f*) show the original model in the original $2F_o - F_c$ OMIT map, while (*g*)–(*j*) show the normal-mode model in the final $2F_o - F_c$ OMIT map. The newly added sugars at four different glycosylation sites are Asn246 (*c, g*), Asn294 (*d, h*), Asn376 (*e, i*) and Asn475 (*f, j*). All maps were contoured at 1.5σ.

2005*a,b*; R_{cryst} of 38.5% and R_{free} of 38.8%). Indeed, when the original model of gp120 was subjected to one round of one-group TLS refinement and subsequent *REFMAC*5 refinement, the *R* factors became 38.8% for R_{cryst} and 39.1% for R_{free}, which were much closer to the published values (Chen *et al.*, 2005*a,b*).

In normal-mode-based refinement of gp120, effort was made to use a minimal set of thermal parameters because of the limited number of unique reflections. We first explored the inclusion of different combinations of normal modes in refinement. By monitoring the resulting *R* factors and the *B*-factor profile from the refinement, we found that the first four nonzero lowest-frequency normal modes, equivalent to $N(N + 1)/2 = 10$ independent thermal parameters, were sufficient to represent the structural deformations of the gp120 core. After the first round of normal-mode-based anisotropic *B*-factor refinement and *REFMAC*5 refinement, before any manual adjustment, the *R* factors were lowered to 38.7% for R_{cryst} and 39.5% for R_{free}. Compared with the original model, this represents decreases of 0.4% in R_{cryst} and 0.3% in R_{free}, presumably owing to the anisotropic treatment of the gp120 structure, as no manual adjustment was involved.

The entire process of normal-mode-based refinement for the gp120 structure took five iterations of normal-mode-based anisotropic *B*-factor refinement combined with heavily restrained positional refinement using *REFMAC*5 and followed by manual structural adjustment in *O* (Jones *et al.*, 1991). The final *R* values converged to 34.6% for R_{cryst} and 35.4% for R_{free}, both of which are about 4.5% lower than the values for the original model. Compared with the *R* factors reported for the TLS-refined model (Chen *et al.*, 2005*a,b*), the final normal-mode model is 3.7% and 3.4% lower in R_{cryst} and R_{free}, respectively.

We then tested whether the TLS method would cause further improvement of the normal-mode-refined final model. By using one TLS group, the refinement yielded a model with 34.7% for R_{cryst} and 35.7% for R_{free}, about 0.1% and 0.3% higher than the normal-mode model (Table 1). Thus, in the case of the gp120 core, subsequent application of TLS refinement to the normal-mode model did result in lower *R* factors than those of the TLS-refined published model, but did not lower the *R* factors below those of the final normal-mode model (see §4 for further discussion).

3.2. Improvements of the normal-mode model

The final normal-mode model had a slightly improved geometry in relation to the original model (Table 1). For instance, in the Ramachandran plot 62.1% and 2.2% of the residues were in the core region and disallowed region, respectively, for the original model compared with 62.5% and 0%, respectively, for our final model. Interestingly, the six residues in the disallowed region in the original model did not universally enter the generously allowed region in the normal-mode model. Rather, there were substantial Ramachandran repartitions. Two residues (Arg381 and Thr439) entered the core region, one (Asp244) became allowed and the remaining

three (Glu217, Thr279 and Asp294) were in the generously allowed region. In terms of root-mean-square deviations (r.m.s.d.s) from the ideal value of bond lengths and bond angles, the original model had r.m.s.d.s of 0.014 Å for bond lengths and 2.067 Å for bond angles, while the final normal-mode model had r.m.s.d.s of 0.011 Å for bond lengths and 2.055 Å for bond angles (Table 1).

It is worth emphasizing that an equivalent or improved geometry for the normal-mode model in comparison to the original model is an important prerequisite for fair comparison of *R* factors between the two models. This is because one can easily lower the *R* factors by simply sacrificing the geometry, especially at low resolution. Thus, the comparison of *R* factors as an indicator of how well the models match the experimental data only becomes meaningful when the two models have a similar geometrical quality.

Probably owing to the intrinsic structural flexibility of the inner domain, as revealed in previous studies (Chen *et al.*, 2005*a,b*; Kwong *et al.*, 1998), electron densities for this region, in particular for the N-terminus, the α1 helix, the V1–V2 stem and the β5–β7–β25 sheet, were weak and broken (Figs. 2*a*, 2*b* and 2*c*). In particular, the backbones of some residues were only covered by weak or broken density, such as Phe75 in the loop connecting the N-terminus and the α1 helix, Ser105 in the V1–V2 stem and Ala233 in the region immediately preceding the β5–β7–β25 sheet. In addition, the side chains of a number of residues in the original model were completely out of electron density, including Gln85 in the α1 helix and the di-sulfide bond between Cys210 and Cys108 in the V1–V2 stem.

The normal-mode-based anisotropic refinement of the gp120 core resulted in an improvement in the $2F_o - F_c$ composite OMIT map, which guided manual adjustments of many individual residues, especially in the flexible regions of the inner domain. The missing or broken densities for some of those abovementioned residues gradually emerged (Figs. 2*d*, 2*e* and 2*f*). In some cases, the new densities indicated a different placement of the backbone (Figs. 2*d*, 2*e* and 2*f*). Examples include Ile98 and Lys99 in the region preceding the V1–V2 stem (Figs. 2*b* and 2*e*), Lys103 in the V1–V2 stem (Figs. 2*b* and 2*e*) and Ala233 and Gly236 in the region immediately preceding the β5–β7–β25 sheet (Figs. 2*c* and 2*f*).

The overall backbone differences between the original and final normal-mode models are depicted in Figs. 2(*g*) and 2(*h*). Not surprisingly, the largest differences between the two models are found at the inner domain, in particular at the loop preceding the V1–V2 stem (peak 1, residues 97–101) and most of the V1–V2 stem (peak 2, residues 102–213; Fig. 2*h*), with an average r.m.s.d. of 1.433 Å for residues 97–213. This r.m.s.d. is much larger than the average r.m.s.d.s of 0.454 Å for all other residues and of 0.521 Å for all residues in the model. The average phase angle shift is 32.8° between the normal-mode model and the original model, with larger shifts at higher resolution (Fig. 2*i*).

3.3. The addition of new sugar rings

In addition to structural adjustments of the protein portion of the model, the improved density map also allowed the

addition of five new sugar rings (Fig. 3a). Owing to the intrinsic structural flexibility of sugar molecules, electron densities for sugars were generally weak and sometimes did not cover all the sugar atoms. By using our normal-mode-based refinement, some of those weak densities became stronger and some new densities appeared that allowed the extension of several glycan chains. The addition of these sugars was based on the scheme of high-mannose glycosylation (Fig. 3b), which was also employed in the original structure determination. The newly built sugars included an α-mannose at the terminus of the glycan chain attached to Asn246 (Figs. 3c and 3g), a β-mannose in the glycan chain at Asn294 (Figs. 3d and 3h), two α-mannoses in the glycan chain at Asn376 (Figs. 3e and 3i) and an α-mannose at Asn475 (Figs. 3f and 3j). The B factors for these new sugar atoms are relatively high, with an average value of 182.3 Å^2. This is consistent with their terminal locations on the glycan chains.

3.4. Structural flexibility reflected in the normal-mode model

Overall, the B-factor distribution of the final normal-mode model of gp120 agrees well with its functionally relevant structural flexibility. As shown in Fig. 4(a), in the normal-mode model the inner domain has a much higher average B factor than the outer domain. In particular, the residues on the V1–V2 stem and on the connecting loop between the α1 helix and the V1–V2 stem have the largest B factors (Fig. 4a). Anisotropic ellipsoids for each C$^\alpha$ atom of the protein component and all atoms in the sugar chains derived from the six anisotropic parameters generated by the normal-mode method are shown in Fig. 4(b). The ellipsoids of the C$^\alpha$ atoms in the inner domain exhibit a higher degree of anisotropy in comparison to the smaller and more spherical ellipsoids in the

outer domain region. Additionally, the glycan chains, especially the terminal sugars, have much larger and more anisotropic ellipsoids, suggesting anisotropic deformation of these sugars on the surface of the gp120 protein, which may protect a protein surface area that is significantly larger than it actually appears to be from recognition by host immune systems.

4. Discussion

Here, we report a normal-mode-based refinement of the SIV gp120 structure that was originally determined to 4.0 Å resolution. One feature of this structure determination is the rather low data-to-parameter ratio: the number of unique reflections is less than twice the number of atoms in the structure. In order to limit the number of parameters in refinement in the original structure determination, the TLS method as implemented in *REFMAC*5 was employed in order to derive a better B-factor model. One TLS group equivalent to 20 independent refinement parameters was used during the refinement (Chen *et al.*, 2005a,b). In the normal-mode-based refinement reported here, we found that the use of the first four lowest-frequency normal modes, which is equivalent to ten independent refinement parameters, in addition to the total of 20 independent refinement parameters for one TLS group and one parameter for the TLS scaling factor (see §2 for more details), was able to produce the best B-factor model. The normal-mode-derived B-factor model revealed much higher structural flexibility in the inner domain than in the outer domain (Fig. 4a). This structural flexibility is also reflected in the elongated shape of the ellipsoids for the inner domain, which are in marked contrast to the more spherical ellipsoids for the outer domain (Fig. 4b). These observations agree well with two lines of evidence from previous studies

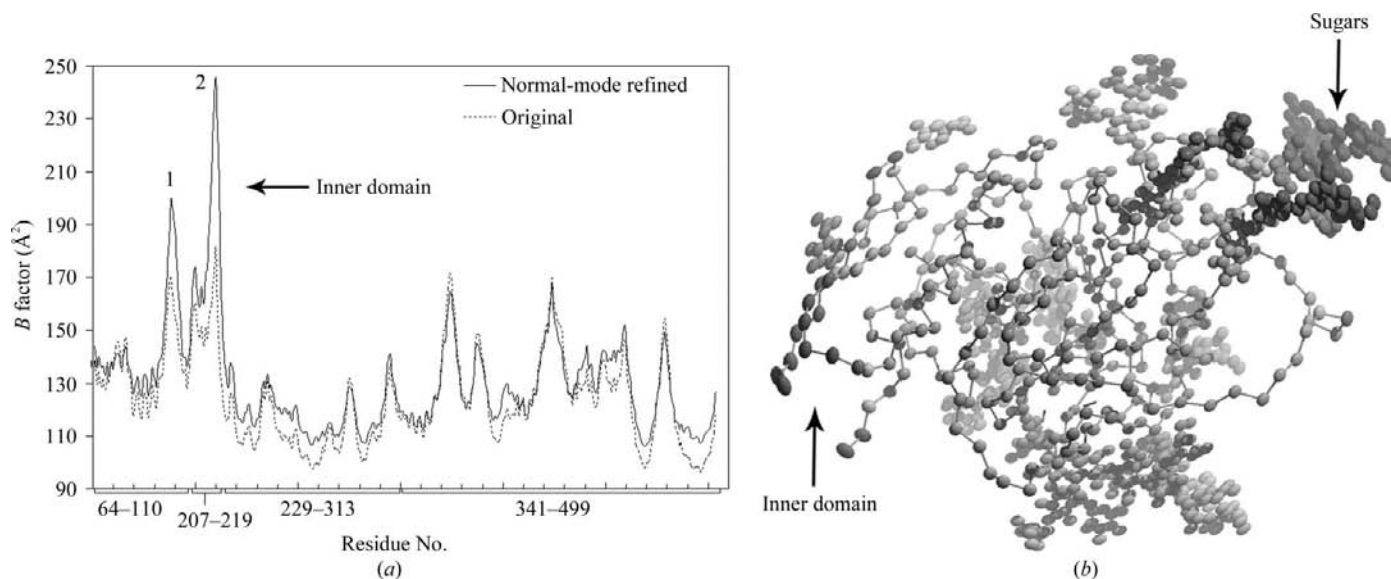**Figure 4**
Comparison of thermal parameters of the original model and normal-mode model. (a) Superposition of the B-factor distributions of the original gp120 core structure (dashed line) and the normal-mode model (solid line). Regions with the highest B factors are labeled. Peak 1, the loop preceding the V1–V2 stem and the β2 sheet of the V1–V2 stem; peak 2, the C-terminal part of the V1–V2 stem. Conversion of the normal-mode anisotropic B factors to isotropic B factors is described in §2. (b) Ellipsoidal representation of the normal-mode model for C$^\alpha$ atoms of the protein portion and all atoms of the glycan chain, made with a probability of 0.1 (Schomaker & Trueblood, 1968).

Chen *et al.* · Simian immunodeficiency virus gp120 core

that suggested substantial flexibility of the inner domain. One was from a comparison of the unliganded SIV gp120 structure with that of a previously solved HIV gp120–CD4 complex, which revealed a relocation as large as 28 Å of the inner domain upon CD4 binding (Chen *et al.*, 2005*a*,*b*; Kwong *et al.*, 1998). The other was from the inability to locate a loop connecting the V1–V2 stem and the $\beta5$–$\beta7$–$\beta25$ sheet in the SIV gp120 core structure (Fig.1). Thus, these results reflect the power of normal-mode-based refinement in modeling structural deformations, especially functionally important ones.

On applying TLS refinement to the final normal-mode model of the gp120 core, we found that TLS did yield lower R factors compared with the published values, but did not bring them below those of final normal-mode model (Table 1). However, it is noteworthy that in our experience the use of normal-mode refinement can sometimes improve the subsequent application of TLS refinement (F. Ni, B. K. Poon, M. Lu, Q. Wang and J. Ma, unpublished data). Thus, in real applications it is always worth exploring whether the combined use of normal-mode and TLS refinements would yield improved refinement.

In this study, we noticed that the composite OMIT map based on normal-mode refinement is quite different from that of the TLS model in some regions, particularly regions with high B factors. These differences were significant enough to allow gradual model improvements based on gradually improved electron-density maps, eventually leading to an overall improvement of the final normal-mode model of the gp120 core. We hope that the better map is a consequence of a more physically relevant representation of atomic motion by normal modes, which provide more accurate phase information than TLS. We compared the phase-angle shifts caused by one round of one-group TLS and normal-mode-based refinement on the original gp120 model (see §3.1). Since the original model had isotropic B factors, the inclusion of anisotropic B factors by TLS or normal-mode methods followed by subsequent positional refinement using *REFMAC*5 offers the best comparison

of anisotropic B-factor models generated by these two methods. The R factors after one round of anisotropic B-factor refinement and *REFMAC*5 refinement were 38.8% for R_{cryst} and 39.1% for R_{free} for TLS refinement, compared with 38.7% for R_{cryst} and 39.5% for R_{free} for normal-mode refinement. Although the two anisotropic B-factor refinement methods yielded comparable R factors, the normal-mode method caused almost twice as large a shift compared with the TLS method in all resolution shells (the average phase-angle shifts were 10.3° and 17.3° for the TLS and normal-mode methods, respectively; Fig. 5). Thus, it is likely that the significantly larger phase-angle shifts caused by normal-mode-based refinement contribute to the substantial improvement of the electron-density maps and the consequent structural models.

In summary, the application of our normal-mode-based method to the structural refinement of SIV gp120 demonstrates the method's potential in improving models of both protein and nonprotein components in X-ray structures and in revealing functionally important structural flexibility, even for rather low-resolution structures. In fact, low- to intermediate-resolution structures particularly need methods like this that can efficiently model B-factor distributions using an exceptionally small number of independent parameters.
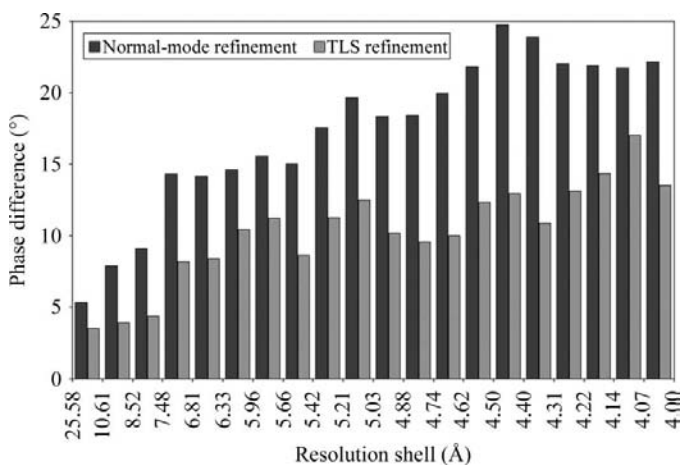
Figure 5
Phase-angle shifts of one-round refinement by the normal-mode method (black bars) and the TLS method (gray bars) in relation to the original model as a function of resolution. The normal-mode method causes significantly larger phase-angle shifts in all resolution shells.

References

Allan, J. S., Coligan, J. E., Barin, F., McLane, M. F., Sodroski, J. G., Rosen, C. A., Haseltine, W. A., Lee, T. H. & Essex, M. (1985). *Science*, **228**, 1091–1094.

Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O. & Bahar, I. (2001). *Biophys. J.* **80**, 505–515.

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.

Center, R. J., Leapman, R. D., Lebowitz, J., Arthur, L. O., Earl, P. L. & Moss, B. (2002). *J. Virol.* **76**, 7863–7867.

Chen, B., Vogan, E. M., Gong, H., Skehel, J. J., Wiley, D. C. & Harrison, S. C. (2005*a*). *Nature (London)*, **433**, 834–841.

Chen, B., Vogan, E. M., Gong, H., Skehel, J. J., Wiley, D. C. & Harrison, S. C. (2005*b*). *Structure*, **13**, 197–211.

Chen, X., Poon, B. K., Dousis, A., Wang, Q. & Ma, J. (2007). *Structure*, **15**, 955–962.

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

Dalgleish, A. G., Beverley, P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F. & Weiss, R. A. (1984). *Nature (London)*, **312**, 763–767.

Delarue, M. & Dumas, P. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 6957–6962.

Diamond, R. (1990). *Acta Cryst.* A**46**, 425–435.

Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. (1996). *Science*, **272**, 872–877.

Go, N., Noguti, T. & Nishikawa, T. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 3696–3700.

Howlin, B., Butler, S. A., Moss, D. S., Harris, G. W. & Driessen, H. P. C. (1993). *J. Appl. Cryst.* **26**, 622–624.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kamiya, K., Sugawara, Y. & Umeyama, H. (2003). *J. Comput. Chem.* **24**, 826–841.

Kidera, A. & Go, N. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 3718–3722.

Kidera, A. & Go, N. (1992). *J. Mol. Biol.* **225**, 457–475.

Kidera, A., Inaka, K., Matsushima, M. & Go, N. (1992*a*). *Biopolymers*, **32**, 315–319.

Kidera, A., Inaka, K., Matsushima, M. & Go, N. (1992*b*). *J. Mol. Biol.* **225**, 477–486.

Kidera, A., Matsushima, M. & Go, N. (1994). *Biophys. Chem.* **50**, 25–31.

Kondrashov, D. A., Cui, Q. & Phillips, G. N. Jr (2006). *Biophys. J.* **91**, 2760–2767.

Kondrashov, D. A., Van Wynsberghe, A. W., Bannen, R. M., Cui, Q. & Phillips, G. N. Jr (2007). *Structure*, **15**, 169–177.

Kundu, S., Melton, J. S., Sorensen, D. C. & Phillips, G. N. Jr (2002). *Biophys. J.* **83**, 723–732.

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

Lindahl, E., Azuara, C., Koehl, P. & Delarue, M. (2006). *Nucleic Acids Res.* **34**, W52–W56.

Lu, M. & Ma, J. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 15358–15363.

Lu, M., Poon, B. & Ma, J. (2006). *J. Chem. Theor. Comput.* **2**, 464–471.

Ma, J. (2004). *Curr. Protein Pept. Sci.* **5**, 119–123.

Ma, J. (2005). *Structure*, **13**, 373–380.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Poon, B. K., Chen, X., Lu, M., Vyas, N. K., Quiocho, F. A., Wang, Q. & Ma, J. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 7869–7874.

Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–76.

Schroder, G. F., Brunger, A. T. & Levitt, M. (2007). *Structure*, **15**, 1630–1641.

Suhre, K. & Sanejouand, Y.-H. (2004). *Acta Cryst.* D**60**, 796–799.

Trkola, A., Dragic, T., Arthos, J., Binley, J. M., Olson, W. C., Allaway, G. P., Cheng-Mayer, C., Robinson, J., Maddon, P. J. & Moore, J. P. (1996). *Nature (London)*, **384**, 184–187.

Veronese, F. D., DeVico, A. L., Copeland, T. D., Oroszlan, S., Gallo, R. C. & Sarngadharan, M. G. (1985). *Science*, **229**, 1402–1405.

Wu, L., Gerard, N. P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A. A., Desjardin, E., Newman, W., Gerard, C. & Sodroski, J. (1996). *Nature (London)*, **384**, 179–183.

Wyatt, R., Kwong, P. D., Desjardins, E., Sweet, R. W., Robinson, J., Hendrickson, W. A. & Sodroski, J. G. (1998). *Nature (London)*, **393**, 705–711.