Considerations for the refinement of low-resolution crystal structures

Byron DeLaBarre[a,b] **and Axel T. Brunger**[a,b]*

[a]Howard Hughes Medical Institute, USA, and [b]Departments of Molecular and Cellular Physiology, Neurology and Neurological Sciences and Stanford Synchrotron Radiation Laboratory, Stanford University, USA

Correspondence e-mail: brunger@stanford.edu

It is often assumed that crystal structures have to be obtained at sufficiently high resolution in order to perform macromolecular refinement. In several recent structures, the threshold of what is considered 'acceptable' has been pushed to lower diffraction resolutions. Here, considerations and modifications to standard refinement protocols are described that were used to solve and refine a particular set of low-resolution structures for the ATPase p97/VCP. It was found that reasonable R_{free} values and good geometry can be achieved upon refinement that includes experimental phase information along with judicious use of restraints at diffraction limits as low as 4.7 Å. At this resolution, the topology and the backbone-chain trace are mostly defined, some side-chain positions can be unambiguously assigned and ligands within known binding sites can be identified. Furthermore, large conformational changes can be discerned when structures in different states are available, information that is not easily obtainable by other means.

1. Introduction

Biological macromolecules often function within larger assemblies or complexes involving tens of thousands of atoms. The structural investigation of these macromolecular complexes is difficult because they typically do not yield crystals that diffract to high resolution. This is especially true in the initial stages of a structural investigation when relatively little is known about the structural details of the complex. The ribosome-structure studies are a case in point: the first crystals were obtained years ago, but only diffracted weakly (Yonath *et al.*, 1980). It took many years and extensive efforts by many laboratories to eventually obtain the first structure of a subunit of a ribosome (Ban *et al.*, 2000).

A commonly used technique is the 'divide-and-conquer' approach to circumvent the inherent problems of studying large macromolecules, whereby structures of small components or fragments of the complex are solved by X-ray crystallography at high resolution and then pieced together using envelopes obtained by cryogenic electron microscopy (cryo-EM). However, such models critically depend on the quality of the cryo-EM density maps and the degree to which individual components can be identified in such maps. It is not uncommon to obtain reasonable fits of protein fragments to density maps that later turn out to be incorrect (Zhang *et al.*, 2000; DeLaBarre & Brunger, 2003). Furthermore, compo-

research papers

Table 1
Effect of resolution on electron-density map interpretation and refinement.

X-ray diffraction limit (Å)	Maps	Refinement
8–5	Domains are visible	Limited to rigid-body refinement
5–4	Possible to trace main chain with prior information. Large stretches of α-helical region will increase the success rate. Small molecules will be visible and gross interpretations can be made in certain cases (*e.g.* ADP *versus* ATP). Some side chains may be visible.	Refinement with the MLHL target function is possible with secondary-structure restraints. Reasonable R/R_{free} values and geometry can be achieved.
4–3.5	Main-chain trace more reliable. β-Sheets can be built with more confidence. Conformational flexibility can be assessed. More side chains visible. When two or more states are crystallized, conformational changes in side-chain positions may be visible.	Standard refinement techniques can be applied.

nents of the complex or protein fragments taken out of the context of the entire complex have the potential to adopt physiologically irrelevant conformations and can even undergo rearrangements at the secondary-structural level. For all these reasons, it is highly desirable to crystallize and to determine the structure of an entire macromolecular complex.

Unfortunately, even after crystallization conditions have been discovered, structural projects can languish for long periods of time or even be abandoned because of limited diffraction. When the macromolecule or complex of interest does not form a well ordered crystalline lattice, atomic resolution observations are not possible. Still, topological and even some detailed structural information is there for the taking (Table 1). Topological properties such as the connection or orientation of domains, the relation of structured domains to the primary sequence and medium- to large-scale conformational changes are within the purview of low-resolution crystallography. Low-resolution models can provide important insights into macromolecular function and, more importantly, provide the basis for further experimentation.

Several structures have been reported over the past decade at low resolution, but these studies were typically restricted to approximate modeling of electron-density maps or were able to benefit from the presence of high non-crystallographic symmetry (NCS). Examples of these types of studies include (this is not meant to be an all-inclusive list): an entire ribosome to 9 Å resolution (Vila-Sanjurjo *et al.*, 2003), studies of the reverse transcriptase from HIV in complex with a target RNA pseudoknot at 4.75 Å resolution (Jaeger *et al.*, 1998), the *Escherichia coli* F1 ATPase at 4.4 Å resolution (Hausrath *et al.*, 2001), the structure of heptameric protective antigen bound to an anthrax toxin receptor at 4.3 Å (14-fold NCS; Lacy *et al.*, 2004), the complex between human CD4 two-domain receptor fragments and MHC class II molecules at 4.3 Å resolution (Wang *et al.*, 2001), the crystal structure of the core of the AP-1 complex at 4 Å resolution (sixfold NCS; Heldwein *et al.*, 2004) and a crystal structure of the dimeric HIV-1 capsid protein at 3.7 Å resolution (Momany *et al.*, 1996).

In this paper, we describe the methodological developments, techniques and experiences that we obtained from the structure solution of the ATPase p97/VCP in three different nucleotide states (DeLaBarre & Brunger, 2003, 2005). We primarily focus on the structure of p97/VCP in complex with ADP·AlF$_x$ that was solved initially at 4.7 Å resolution and later re-refined at 4.4 Å resolution, since it presented significant challenges during the structure-solution and refinement process.

2. Prerequisites

2.1. Structural knowledge of fragments or components at high resolution

The most subjective step in solving a macromolecular structure from diffraction data lies in interpreting the electron-density maps. Methods have been developed to partially automate the process, especially at high resolution, but the eyes of a good crystallographer are still required to complete the task. For crystals that diffract to high resolution, the electron-density maps can be interpreted without reference to other structures. This is not the case for low-resolution crystallography. Discontinuities in the maps and a paucity of side-chain electron density that can be reliably identified are the primary problems. However, low-resolution electron-density maps will contain some recognizable features. Secondary-structural elements such as α-helices can typically be easily placed within low-resolution electron density, β-sheets can be modelled with some difficulty and loop regions are often visible but near-impossible to interpret reliably. So, unless the macromolecule(s) is composed primarily of α-helical domains, *de novo* chain tracing of the entire structure is not a realistic goal. Thus, an important prerequisite is the availability of structures of components or fragments solved at high resolution. Molecular replacement or real-space fitting can then be used to place these components within the electron-density map of the entire structure.

In the case of the structure of p97/VCP in complex with ADP·AlF$_x$, a fragment consisting of two domains, N and D1, had been solved to ~3 Å resolution, comprising ~60% of the full-length protein (Zhang *et al.*, 2000). We obtained crystals of the full-length protein and improved the diffraction from ~8 to 4.7 Å resolution through changes in the crystallization and cryocooling procedures (DeLaBarre & Brunger, 2003). The diffraction data from these crystals were later extended to 4.4 Å by merging data from two collections of slightly

different time exposures (DeLaBarre & Brunger, 2005). Molecular replacement of the N–D1 fragment into the diffraction data of the full-length molecule was straightforward and the resultant maps showed electron density for the missing D2 domain. We were able to recognize helical regions in the electron density of the initially unknown D2 domain, but little else could be interpreted at this point. The sequence identity and similarity between the D1 and D2 domains are 38 and 66%, respectively, suggesting some structural homology between the domains. We used this homology to make tentative assignments for the helical regions visible in the D2-domain electron density.

2.2. Experimental phase information

A critical prerequisite for low-resolution crystallography is good experimental phase information, since molecular replacement with smaller known fragments may not reveal the electron density of the unknown components with sufficient clarity (DeLaBarre & Brunger, 2003). In our experience, experimental phase information from diffraction data to resolution limits as low as 6.5 Å can suffice. Diffraction data from selenomethionine-substituted protein can be particularly useful because the selenium positions provide a guide to where the methionine residues should be positioned during chain tracing. For p97/VCP, 19 methionine residues were distributed evenly throughout the primary sequence. The phases from the molecular-replacement solution were used to compute anomalous difference maps that eventually provided the location of 50 of the 57 Se atoms within the asymmetric unit (three protomers with 19 methionine residues per protomer in the asymmetric unit). In addition to the selenomethionines, a previously unknown zinc ion at the centre of the D1 pore also provided some anomalous signal. In general, the presence of naturally occurring ions could be exploited both for phasing as well as a guide for chain tracing if the coordinating residues are known.

We explored both multiple anomalous dispersion (MAD) and single anomalous dispersion (SAD) phasing combined with density modification. Compared with MAD phasing, the figure-of-merit value for the SAD phase probability distribution was low, the resulting electron-density maps were poor and both the positional and B-factor parameters for the anomalous scatterers were unstable (data not shown). Thus, while SAD is a powerful and efficient way of solving crystal structures at high to medium resolution (Rice *et al.*, 2000), the reliance on density modification makes SAD more challenging at low resolution.

With the substructure of the anomalous scatterers determined, we performed an iterative feedback loop for improving both the p97/VCP atomic model and the parameters of the selenomethionine substructure (Collaborative Computational Project, Number 4, 1994; Terwilliger, 2003; Perrakis *et al.*, 1999). Phase probability distributions of the current p97/VCP atomic model were computed. Next, these model phase probability distributions were used as 'prior' distributions to assist the maximum-likelihood refinement of the selenium

substructure. New experimental phase probability distributions (without the prior distributions) were then computed from the refined phasing model and used to assist the maximum-likelihood refinement of the entire p97/VCP atomic model using the MLHL target function (Pannu *et al.*, 1998; Adams *et al.*, 1999). This loop was repeated to convergence of the standard phasing statistics and electron-density map quality. Standard objective measures of improvement (figure of merit of the selenium substructure, R_{free} of the p97/VCP atomic model) as well as subjective examination of selected regions of electron density indicated significant improvements using this iterative method.

2.3. Conformationally homogeneous crystals

The quality of electron-density maps is not related to the overall diffraction limit of the crystal, but rather to how much of the macromolecule or complex has crystallized with a single unique conformation. Although there are other factors to consider, smaller macromolecules will typically pack with more crystal contacts relative to their surface area, thus locking in specific protein conformations, making the resultant electron-density maps more connected. Conversely, larger macromolecules or complexes will have fewer crystallographic contacts with respect to their surface area and thus allow more possible conformations to be sampled within the crystalline lattice. Thus, crystallization conditions must be found that reduce conformational heterogeneity. For example, in the case of p97/VCP, crystals that diffract to 3.5 Å resolution were obtained in the both the apo and AMP-PNP states (DeLaBarre & Brunger, 2005; Huyton *et al.*, 2003). While the electron-density map for the N and D1 domains showed more atomic detail than the corresponding maps of p97/VCP in the ADP·AlF$_x$ state solved at 4.4 Å resolution, the D2 domain was largely disordered, actually providing less information than obtained for the crystals that diffracted to lower resolution.

Growing crystals in a variety of conditions is a straightforward but very time-consuming way to search for conformational homogeneity. Thus, we turned to reconstructions obtained from single-particle averaging of cryo-EM data to speed the hunt for homogeneity. While the absolute interpretation of cryo-EM reconstructions can be a subjective process, the relative interpretation of a well controlled series of reconstructions is very informative. The parts of the macromolecule that take on variable conformations under one set of conditions will have decreased density in the averaged images. Picking the optimum conditions for crystallization entails translating the cryo-EM sample conditions that yield the most complete reconstruction of the protein. For p97/VCP, we found that incubating the protein with ADP·AlF$_x$ (to create an ATP-hydrolysis transition-state mimic) resulted in cryo-EM reconstructions that showed fuller visible density for all domains (Rouiller *et al.*, 2002). Efforts were focused on improving the crystals of the protein under these conditions. Ultimately, the transition-state analogue ADP·AlF$_x$-bound version of p97/VCP provided the most complete electron-density maps for the wild-type protein. Later, a slight

improvement in diffraction resolution was observed for the ADP-bound state, but this entailed using a selenomethionine derivative of p97/VCP (DeLaBarre & Brunger, 2005).

2.4. Other information

In order to interpret low-resolution electron-density maps, it is necessary to make many subjective judgments. This entails using information from sources other than electron-density maps. If several interpretations of an electron-density map region are possible, there are several ways to evaluate which interpretation is likely to be correct. One can use R_{free} behavior, homology to macromolecules with known structure, secondary-structure prediction of the boundaries of secondary-structural elements and known atomic positions (*e.g.* selenomethionine positions or other anomalous scatterers) to assist in the decision-making process. This is not a process that can be easily automated: the judgment of a skilled crystallographer while inspecting the model and electron-density maps at low resolution is paramount. It may require several trials of building and refining in order to establish that the model is correct.

When there are only marginal differences between multiple possibilities in interpreting electron-density maps, there are several options available. Firstly, the region can be left with the best possible interpretation, something that is certainly highly subjective. Secondly, the region can be left uninterpreted, a conservative but not very informative approach. Thirdly, all possible interpretations can be included in the model. This is probably the least desirable because it adds to the number of free parameters, rendering the structure underdetermined at low to medium resolution.

3. Special considerations for refinement at low resolution

3.1. Use of all diffraction data

In the past, crystallographers typically truncated the observed diffraction data at high resolution using a 2–3 $I/\sigma(I)$ cutoff criterion. This was a reasonable choice when only least-squares methods were used in refinement. With the emergence of maximum-likelihood-based refinement methods (Adams *et al.*, 1997; Pannu *et al.*, 1998), it is possible and desirable to include weaker diffraction data in refinement. Often, there are many significant reflections past the conventional cutoff that would otherwise be omitted from the data analysis. Weak reflections are automatically down-weighted in the likelihood-based target function.

Read and coworkers suggested using the behavior of σ_A as a guide to determine an effective resolution limit (Ling *et al.*, 1998). We applied this approach to set the resolution limit for p97/VCP in complex with ADP (Fig. 1); the suggested resolution limit corresponds to a conventional $I/\sigma(I)$ cutoff of 1.2. For the ADP·AlF$_x$ and AMP-PNP-ligated structures this approach resulted in $I/\sigma(I)$ cutoffs as low as 0.8 (DeLaBarre & Brunger, 2005).

3.2. B-factor sharpening

B-factor sharpening is a very useful tool for the enhancement of low-resolution maps (Bass *et al.*, 2002; DeLaBarre & Brunger, 2003). *B*-factor sharpening entails the use of a negative *B* factor, B_{sharp}, in a resolution-dependent weighting scheme applied to a particular electron-density map,

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

where F_{map} are the structure factors of the particular electron-density map and $F_{sharpened\ map}$ are the structure factors of the sharpened map. Adjusting the B_{sharp} parameter has a qualitative effect on the electron-density map by up-weighting of higher resolution terms. The result of this weighting scheme is increased detail for higher resolution features such as side-chain conformations. The cost of the increased detail is increased noise throughout the electron-density map. Sometimes, the noise can coincide with regions of backbone or side-chain electron density, producing potential artefacts. Thus, it is important to always inspect both the original unweighted and the *B*-sharpened electron-density maps. It should also be noted here that we did not observe much improvement of electron-density maps that were computed with phases derived solely from molecular replacement.

Although *B*-factor sharpening can be viewed as a simple weighting function applied to the observed amplitudes and consequently the electron-density maps by virtue of relative scaling between the atomic model and observed amplitudes, it also has a physical meaning. We found empirically that the B_{sharp} value that produced the most useful electron-density map coincides with the smallest absolute value of B_{sharp} that
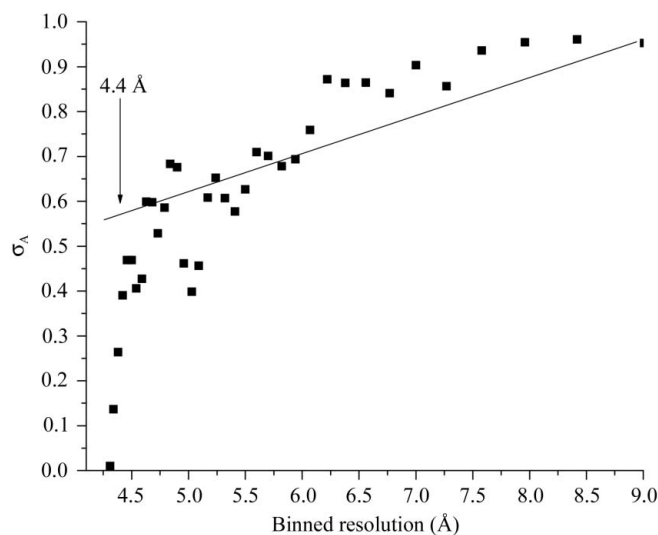

Figure 1
Plot of σ_A *versus* resolution for the p97VCP–ADP·AlF$_x$ diffraction data (DeLaBarre & Brunger, 2005). The cross-validated σ_A value is plotted *versus* resolution. The subset of the diffraction data used for the calculation of the free *R* value was used for computation of σ_A. The continuous line was obtained by linear fit to the diffraction data and the arrow at 4.4 Å indicates the resolution at which σ_A drops sharply from its previous value. This was taken as the effective limit of the diffraction data useful for structure refinement.

results in a Wilson plot that is positive in all resolution bins (Fig. 2b). By choosing such a B_{sharp} value to weigh the observed diffraction data, one is imposing the physical constraint that the average scattering from the crystal is equal to the average scattering from the individual atoms within the crystal (e.g. $\langle F_{hkl}^2 \rangle \simeq \langle f_i^2 \rangle$). In other words, B_{sharp} can be viewed as a pseudo-Wilson scaling of the diffraction data. However, instead of determining the slope and y intercept from the
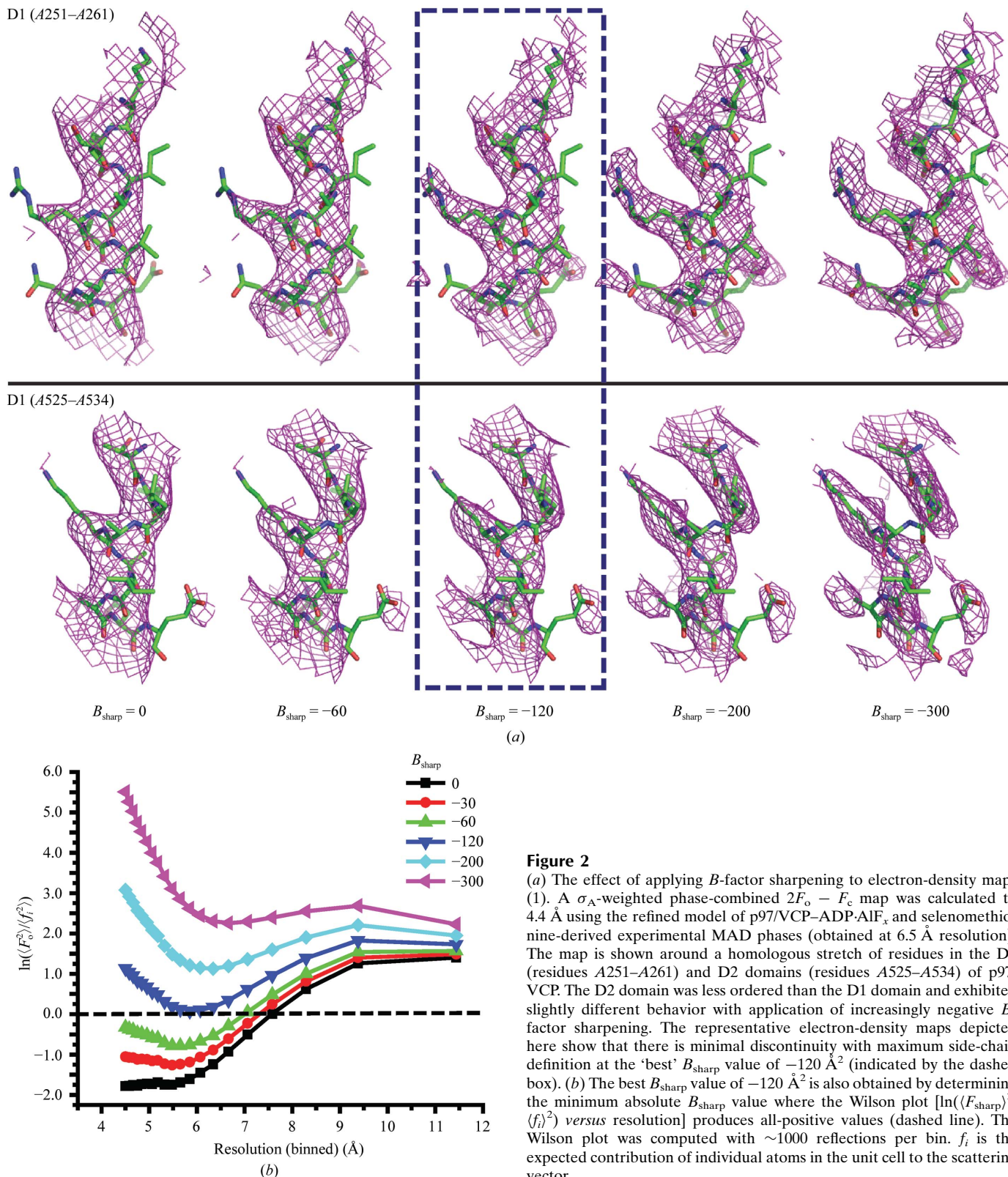

D1 (A251–A261)

D1 (A525–A534)

$B_{sharp} = 0$ $B_{sharp} = -60$ $B_{sharp} = -120$ $B_{sharp} = -200$ $B_{sharp} = -300$

(a)

(b)

Figure 2
(a) The effect of applying B-factor sharpening to electron-density maps (1). A σ_A-weighted phase-combined $2F_o - F_c$ map was calculated to 4.4 Å using the refined model of p97/VCP–ADP·AlF$_x$ and selenomethionine-derived experimental MAD phases (obtained at 6.5 Å resolution). The map is shown around a homologous stretch of residues in the D1 (residues A251–A261) and D2 domains (residues A525–A534) of p97/VCP. The D2 domain was less ordered than the D1 domain and exhibited slightly different behavior with application of increasingly negative B-factor sharpening. The representative electron-density maps depicted here show that there is minimal discontinuity with maximum side-chain definition at the 'best' B_{sharp} value of −120 Å^2 (indicated by the dashed box). (b) The best B_{sharp} value of −120 Å^2 is also obtained by determining the minimum absolute B_{sharp} value where the Wilson plot [$\ln(\langle F_{sharp} \rangle^2/\langle f_i \rangle^2)$ *versus* resolution] produces all-positive values (dashed line). The Wilson plot was computed with ~1000 reflections per bin. f_i is the expected contribution of individual atoms in the unit cell to the scattering vector.

linear (greater than 3 Å) region of the Wilson plot, one is imposing the requirement that the Wilson ratios be positive across the entire resolution range.

3.3. Bulk-solvent model

The correct modeling of the barrier between the bulk solvent in the crystal lattice and the protein itself is an important part of refinement. Iterative building and refinement is a key part of generating improved electron-density maps for interpretation. Therefore, proper bulk-solvent modeling becomes an integral aspect of refining low-resolution structures. The bulk-solvent model and parameters used in *Crystallography and NMR System* (CNS) v.1.1 were established by examining diffraction from proteins of fairly
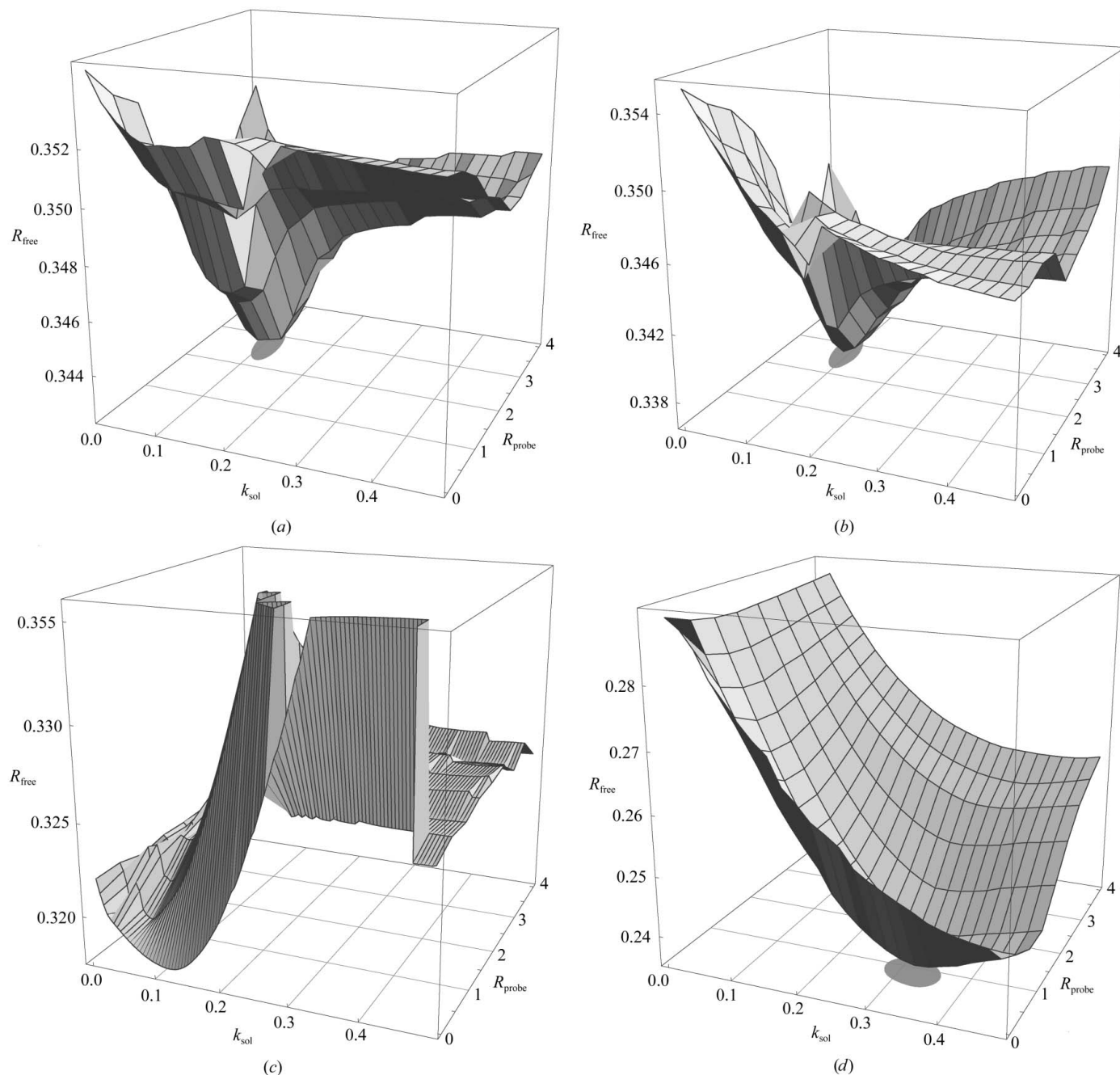

Figure 3
Examples of grid searches for the bulk-solvent parameters k_{sol} and R_{probe} with the shrink radius R_{shrink} set equal to R_{probe}. For each selected pair of parameters, B_{sol} was determined by least-squares fitting of the bulk-solvent model plus macromolecular model structure factors to the observed diffraction data. The resulting R_{free} values are shown as surface plots. The projection of the minimum of the surface onto the k_{sol}, B_{sol} plane is indicated by a gray spot (except for in c). (a) p97/VCP–ADP (PDB code 1yqi) at 4.25 Å resolution. (b) p97/VCP–ADP·AlF$_x$ at 4.4 Å resolution (PDB code 1yq0). (c) VCP–AMP-PNP at 3.5 Å resolution (PDB code 1ypw). This structure is unusual as the distribution is fairly flat and the optimum is reached for $R_{probe} = 0$. This behavior may be a consequence of the very weak and partially disordered electron density in the D2 domain for this crystal form. (d) For comparison with a high-resolution structure, the results are shown for the Sec5–Exo84 complex solved at 1.9 Å resolution (PDB code 1zc3; Jin *et al.*, 2005), producing the expected values of $k_{sol} = 0.3$ e$^-$ Å^{-3} and $R_{probe} = 1$.

high resolution (Jiang & Brünger, 1994). The specific parameters k_{sol}, B_{sol}, R_{probe} and R_{shrink} affect the creation and smoothing of the mask that separates the protein model from the bulk solvent. For the refinement of p97/VCP we found it necessary to adjust the probe and shrink radii, R_{probe} and R_{shrink}, for the computation of the bulk-solvent mask in order to obtain optimum R_{free} values (Figs. 3a, 3b and 3c; DeLaBarre & Brunger, 2003, 2005). At the initial stage of structure refinement, optimization showed that slightly different values for each of the R_{probe} and R_{shrink} parameters could produce slightly lower R_{free} values. However, in the final refinement stages equal values of R_{probe} and R_{shrink} produced optimum results.

The bulk-solvent procedure as implemented in *CNS* v.1.1 sometimes resulted in numerical instabilities for the refinement of the solvent parameters k_{sol} and B_{sol} for structures determined at low to moderate resolution. We therefore modified the procedure by performing a grid search for k_{sol} while determining B_{sol} by least-squares refinement for each selected value of k_{sol}. Others have found a similar solution to this problem involving a grid search of both k_{sol} and B_{sol} (Afonine *et al.*, 2005). In addition, we performed a grid search for R_{probe}, with R_{shrink} set equal to R_{probe}. The new bulk-solvent procedure is robust for structures solved at both high and low resolution (Fig. 3).

3.4. Structure restraints

Restraints or constraints are important for any type of macromolecular structure, but they become particularly important for low-resolution cases. In a simplistic manner, it reduces down to obtaining a useful data-to-parameter ratio. Our crystal form of p97/VCP–ADP·AlF$_x$ (a homohexamer of 806 amino-acid residues with three unique protomers in the asymmetric unit) has ∼16 000 crystallographically unique atoms or ∼44 000 variables (atomic positions and isotropic B factors), but provided only ∼22 000 unique reflections at 4.7 Å resolution to refine against. Clearly, the coordinates of each atom are not free to refine independently because known chemical restraints are imposed on the bond lengths and the angles that may be adopted between atoms. Nevertheless, the refinement of p97/VCP pushed the limits of what has been tried previously.

Non-crystallographic symmetry (NCS) restraints were invaluable for effectively reducing the data-to-parameter ratio. For p97/VCP, there were three independent copies in the ADP and ADP·AlF$_x$ structures and four in the AMP-PNP structure. Various NCS-restraint schemes available in *CNS* were tried with the protein broken into subdomains for the NCS modeling. The behavior of R_{free} upon refinement was used to determine the most appropriate NCS scheme and weights.

Another restraint that was imposed during the refinement of p97/VCP was hydrogen-bonding networks in regions of secondary structures. When α-helices were identified, distance restraints representing the amide hydrogen-bonding network were imposed to further improve the data-to-parameter ratio.

The imposition of expected φ/ψ angles based on analyses of Ramachandran plots was applied manually in the rebuilding stage.

3.5. Atomic B factors

The atomic B factors proved the most problematic to refine. Initially, grouped B factors were used, with side-chain atoms refined separately from main-chain atoms for groups of five residues. This was increased to refinement of split main-chain/side-chain B factors for every residue when slightly higher resolution diffraction data were obtained for the p97/VCP–ADP complex. The overall B-factor model refinement was sometimes unstable (*i.e.* resulting in widely varying B factors approaching 0 Å^2 at one extreme and very large numbers at the other extreme) and dependent on starting conditions. The final B-factor model was obtained by periodically resetting all B factors to the average value and allowing them to refine to values which produced the lowest overall R_{free} values. Unfortunately, there was little correspondence between refined B factors and the presence of electron density for a residue within the map. Establishing and refining the B-factor model for low-resolution structures is an area that needs improvement.

3.6. Multi-crystal averaging

The work by Chen *et al.* (2005) used a refinement strategy similar to that used for p97/VCP but that employed another important tool for low-resolution crystallography: multi-crystal averaging. The structure solution was a remarkable achievement because of the limited resolution and radiation-sensitivity of the crystals, the failure of molecular replacement and the absence of non-crystallographic symmetry. Molecular replacement apparently failed owing to large conformational differences in the liganded form of gp120, so heavy-atom phasing had to be used. A mixed approach of phase combination, B-factor sharpening, density modification, multi-crystal averaging, model building and heavily restrained refinement was used to solve the structure. Thus, multi-crystal averaging of OMIT maps can be a powerful tool for electron-density map improvement and model improvement. One method that might have helped this structure, had it been available, would have been simultaneous refinement directly against the three or four non-isomorphous data sets.

4. Validation

4.1. Reasonable geometry

Standard quality checks are based on structures solved at similar resolution (Laskowski, 1993). However, at this time there are not enough structures solved at resolutions below 3.5 Å to provide a useful pool of examples. Indeed, structures solved and refined from lower resolution diffraction data can at first look terrible to most quality-control checks. After considerable effort, we obtained free R values in the 0.27–0.34 range for p97/VCP crystal structures while ensuring that bond lengths and angles were not severely distorted and the back-

bone exhibited reasonable φ/ψ angles (better than 85% of residues were in the accepted regions of the Ramachandran plot, with almost all of the remaining 15% falling within the generously accepted regions). The use of φ/ψ-angle restraints was purposefully omitted from p97/VCP refinement so that the Ramachandran plot could be used as a guide during refinement. This proved somewhat problematic. Some regions of the protein that appeared to be correctly built sometimes diverged during refinement, so these regions had to be fixed in place in order for the refinement of the overall structure to progress. In retrospect, the explicit inclusion of φ/ψ restraints during refinement might have been a better approach (Held-wein *et al.*, 2004).

4.2. Local real-space *R* value

As part of the iterative process of refinement and rebuilding, the quality of the electron-density maps and the fit of the model to the map should improve continually. Local
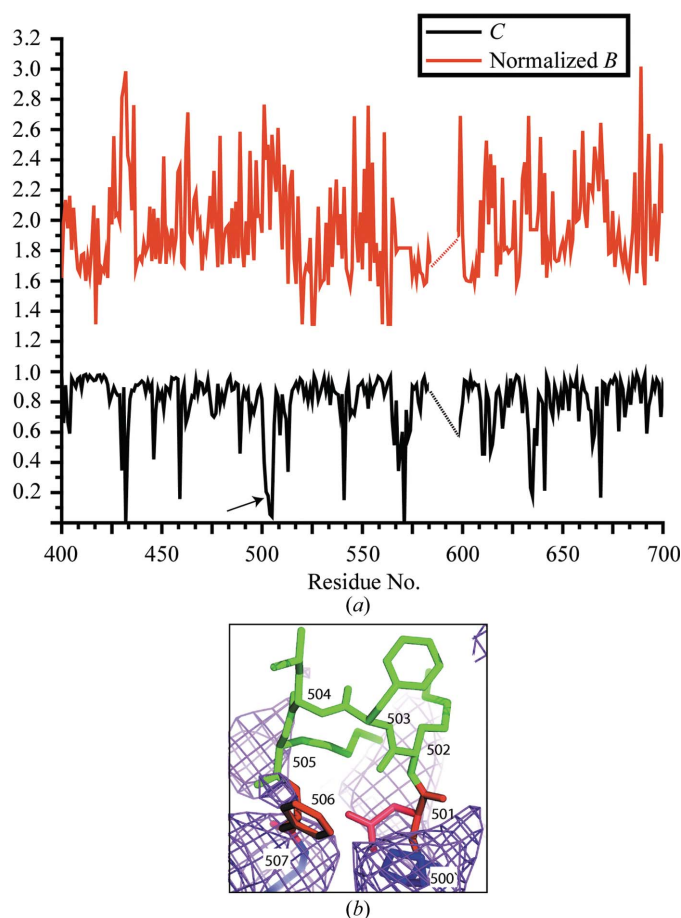

(a)

(b)

Figure 4
A plot of normalized *B* factors and local real-space correlation coefficients *C* calculated for the p97/VCP–ADP·AlF$_x$ structure, residues 400–700 of chain *A*. The normalized *B* factors were offset by 1.5 Å^2 for the sake of clarity. The inset shows residues 501–507 superimposed on an optimally *B*-factor-sharpened $2F_o - F_c$ electron-density map. Residues with *C* less than 0.4 are colored red; all others are colored green. The local real-space correlation plot clearly indicates that the loop is out of density, whereas the *B*-factor plot, although essentially mirroring the behavior of the *C* plot, is quite noisy.

real-space correlation coefficients (*C*) (or alternatively real-space *R* values, defined as RSR = $1 - C$) are useful indicators of regions of the structure that are out of electron density (Fig. 4). It should be noted that the real-space correlation-coefficient calculation in *CNS* v.1.1 assumed a uniform mean electron density throughout the crystal. We modified the protocol such that the average electron density was computed for each residue separately. This improved protocol produced a local correlation coefficient that was more reliable for predicting residues that are not properly fitted or for which there was discernable electron density.

An example is shown in Fig. 4(*b*): the loop between residues 502 and 505 exhibits *C* values that are significantly smaller than the mean and indeed there is no electron density visible. Thus, *C* values can be used as a first screen to identify problematic regions in the model. However, they do not always reveal problems in the fit of the model. For example, *C* can be misleadingly high if the *B* factors of the atoms of the particular residue are very high while the electron density is flat in that particular region. Clearly, such a residue is not well defined. Thus, manual inspection of the fit between the model and the electron-density maps is still important. New approaches, perhaps based on pattern-recognition techniques (Aishima *et al.*, 2005), need to be developed to further assist in this very time-consuming process. Improvements in this area would also lead to more reliable real-space refinement methods at low resolution.

5. Milestones in p97/VCP structure solution

Initial phases were obtained by molecular replacement with the N–D1 p97/VCP structure (Zhang *et al.*, 2000) into a 4.7 Å data set. Self-rotation functions had indicated that a twofold NCS operator was present, so the search was performed with a single protomer from the hexameric p97/VCP molecule. This resulted in a model with *R* and R_{free} values of 51.3 and 52.7%, respectively (Fig. 5*a*). Strict NCS was used initially, but was dropped in favor of NCS restraints applied to the three independent protomers within the asymmetric unit, resulting in *R* values that were below 50% after positional refinement. Next, polyserine helices were placed into the D2-domain electron density in regions that had 'sausage-like' character. The positions of the polyserine helices were refined by rigid-body minimization. These initial helices confirmed the expected structural similarity between the D1 and D2 domains. Using the D1 domain as a guide, the polyserine model for the D2 domain was further extended to produce models for the β-sheets and some information on loop connectivity. Medium-temperature ($\sim$1500 K) torsion-angle simulated annealing was used to refine this model, resulting in *R* and R_{free} values of 36.2 and 40.6%, respectively. MAD phases from the abovementioned selenomethionine-substituted derivative were introduced. The location of Se atoms was used to assign more residues in the model of the D2 domain by inspection of phase-combined electron-density maps and anomalous Fourier difference maps (Fig. 5*b*). The MAD phases enabled successful application of *B*-factor

sharpening to phase-combined electron-density maps (Fig. 5c), yielding maps which enabled a tentative assignment for ~30%
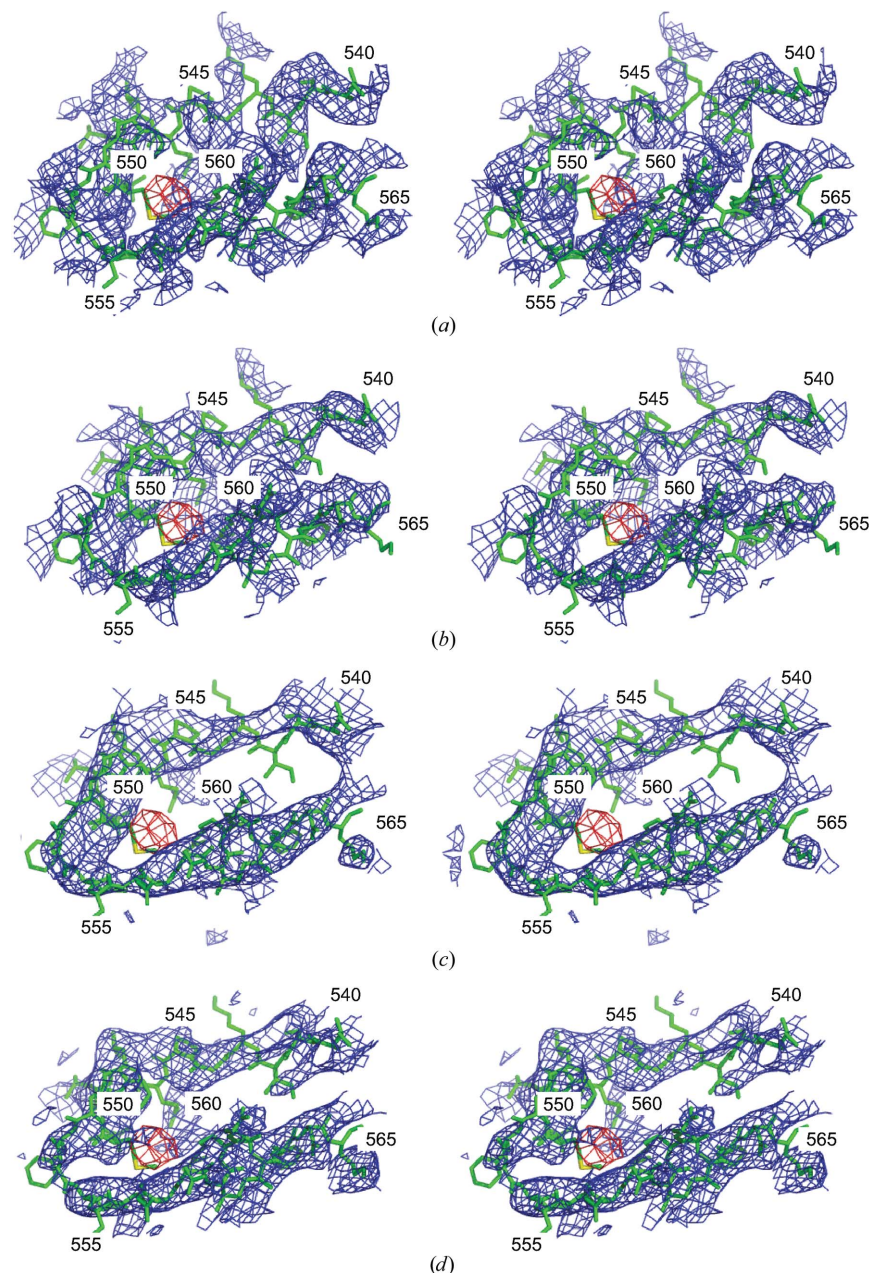

Figure 5
Electron-density maps showing the improvement of phase information as the structure was being completed and experimental phase information introduced. All panels show residues 540–565 from the D2 domain of the *A* chain of VCP/p97–ADP·AlF$_x$ using the coordinates of the final model (green). Also shown in all panels is an anomalous difference map computed from a data set collected at the selenium edge of a selenomethionine-substituted protein crystal, contoured at 4σ (red). Superimposed are σ_A-weighted $2F_o - F_c$ maps calculated to 4.5 Å and displayed at a 1σ contour level (blue). All images are shown in cross-eyed stereo with every fifth residue number indicated. (*a*) F_c and phases for this map were taken from the molecular-replacement solution using the N–D1 model (~60% of the complete model, *i.e.* no information about the D2 domain was used at this point). (*b*) Phase-combined map. Model structure-factor phases for this map were taken from a model based on a minimized N–D1 region with the D2 region built as a discontinuous polyserine trace. The 6.5 Å MAD phases were combined with the model phases. (*c*) Phase-combined map using the final model. This map has neither optimized bulk-solvent correction nor *B*-factor sharpening applied to it. (*d*) As (*c*), but with optimized bulk-solvent model and *B*-factor sharpening.

of the residues in the D2 domain. The phase-combined electron-density maps also allowed identification of the bound nucleotide as ADP·AlF$_x$ in the D2 domain and ADP in the D1 domain. Eventually, the combination of information from the electron-density maps, the selenomethionine positions and iterative model building enabled a complete assignment of residues in the D2 domain. At this point, the improved bulk-solvent method was introduced into refinement. Subsequent rounds of reciprocal-space refinement and manual rebuilding brought the R and R_{free} values down to 28.4 and 30.1%. Finally, the diffraction data were extended from 4.7 to 4.4 Å resolution and a slight asymmetry in both the protomers and their D2-domain-bound nucleotides emerged. Alternating rounds of reciprocal-space refinement and manual rebuilding resulted in the final model with R and R_{free} values of 26.8 and 33.9%, respectively, at 4.4 Å resolution (Fig. 5d).

6. Conclusions

The crystallographer who has become accustomed to seeing macromolecular crystal structures that reveal individual atomic positions, hydrogen-bonding patterns and solvation may well ask the question: why would one bother with the difficulties of solving a structure at the underwhelming limit of 4.7 Å? To answer this, one must accept that the goal of macromolecular crystallography is to obtain biological knowledge. If the right questions are asked, there is rich information even at 4.7 Å resolution. As crystallographers will increasingly study larger and more difficult macromolecular complexes, the intrinsic flexibility of some of these complexes will preclude structure solution at high resolution. Should attempts be made to obtain high-resolution crystals? Absolutely. However, one must not disregard the information that can be obtained from samples that only diffract to low resolution.

The methods discussed in this paper are only the first attempt to interpret low-resolution diffraction data and to refine models against such data. Some of the approaches we have taken could benefit from further improvement, such as *B*-factor sharpening and secondary-structure restraints. We believe that entirely new approaches also need to be developed.

These include, but are not limited to, aids to interpret noisy low-resolution maps, estimators of individual atomic coordinate errors and statistically correct combination of structural information from a variety of sources. These methods will also have applications for structures obtained by cryo-electron microscopy methods.

References

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

Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1999). *Acta Cryst.* D**55**, 181–190.

Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *Acta Cryst.* D**61**, 850–855.

Aishima, J., Russel, D. S., Guibas, L. J., Adams, P. D. & Brunger, A. T. (2005). *Acta Cryst.* D**61**, 1354–1363.

Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). *Science*, **289**, 905–920.

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

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

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

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

DeLaBarre, B. & Brunger, A. T. (2005). *J. Mol. Biol.* **347**, 437–452.

Hausrath, A. C., Capaldi, R. A. & Matthews, B. W. (2001). *J. Biol. Chem.* **276**, 47227–47232.

Heldwein, E. E., Macia, E., Wang, J., Yin, H. L., Kirchhausen, T. & Harrison, S. C. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 14108–14113.

Huyton, T., Pye, V. E., Briggs, L. C., Flynn, T. C., Beuron, F., Kondo, H., Ma, J., Zhang, X. & Freemont, P. S. (2003). *J. Struct. Biol.* **144**, 337–348.

Jaeger, J., Restle, T. & Steitz, T. A. (1998). *EMBO J.* **17**, 4535–4542.

Jiang, J. S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.

Jin, R., Junutula, J. R., Matern, H. T., Ervin, K. E., Scheller, R. H. & Brunger, A. T. (2005). *EMBO J.* **24**, 2064–2074.

Lacy, D. B., Wigelsworth, D. J., Melnyk, R. A., Harrison, S. C. & Collier, R. J. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 13147–13151.

Laskowski, R. (1993). *J. Appl. Cryst.* **26**, 283–291.

Ling, H., Boodhoo, A., Hazes, B., Cummings, M. D., Armstrong, G. D., Brunton, J. L. & Read, R. J. (1998). *Biochemistry*, **37**, 1777–1788.

Momany, C., Kovari, L. C., Prongay, A. J., Keller, W., Gitti, R. K., Lee, B. M., Gorbalenya, A. E., Tong, L., McClure, J., Ehrlich, L. S., Summers, M. F., Carter, C. & Rossmann, M. G. (1996). *Nature Struct. Biol.* **3**, 763–770.

Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Acta Cryst.* D**54**, 1285–1294.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). *Acta Cryst.* D**56**, 1413–1420.

Rouiller, I., DeLaBarre, B., May, A. P., Weis, W. I., Brunger, A. T., Milligan, R. A. & Wilson-Kubalek, E. M. (2002). *Nature Struct. Biol.* **9**, 950–957.

Terwilliger, T. (2003). *Acta Cryst.* D**59**, 1174–1182.

Vila-Sanjurjo, A., Ridgeway, W. K., Seymaner, V., Zhang, W., Santoso, S., Yu, K. & Cate, J. H. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 8682–8687.

Wang, J. H., Meijers, R., Xiong, Y., Liu, J. H., Sakihama, T., Zhang, R., Joachimiak, A. & Reinherz, E. L. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10799–10804.

Yonath, A., Müssig, J., Tesche, B., Lorenz, S., Erdmann, V. A. & Wittmann, H. G. (1980). *Biochem. Int.* **2**, 428–430.

Zhang, X., Shaw, A., Bates, P. A., Newman, R. H., Gowen, B., Orlova, E., Gorman, M. A., Kondo, H., Dokurno, P., Lally, J., Leonard, G., Meyer, H., van Heel, M. & Freemont, P. S. (2000). *Mol. Cell*, **6**, 1473–1484.