Application of normal-mode refinement to X-ray crystal structures at the lower resolution limit

Fengyun Ni,[a] Billy K. Poon,[a] Qinghua Wang[b] and Jianpeng Ma[a,b]*

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

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

The structural refinement of large complexes at the lower resolution limit is often difficult and inefficient owing to the limited number of reflections and the frequently high-level structural flexibility. A new normal-mode-based X-ray crystallographic refinement method has recently been developed that enables anisotropic B-factor refinement using a drastically smaller number of thermal parameters than even isotropic refinement. Here, the method has been systematically tested on a total of eight systems in the resolution range 3.0–3.9 Å. This series of tests established the most applicable scenarios for the method, the detailed procedures for its application and the degree of structural improvement. The results demonstrated substantial model improvement at the lower resolution limit, especially in cases in which other methods such as the translation–libration–screw (TLS) model were not applicable owing to the poorly converged isotropic B-factor distribution. It is expected that this normal-mode-based method will be a useful tool for structural refinement, in particular at the lower resolution limit, in the field of X-ray crystallography.

1. Introduction

For structural refinement in X-ray crystallography, temperature B factors are used to account for structural deviations from the average positions of atoms. Conventionally, the atomic positional deviation is modeled as a spherical Gaussian and a single scalar constant, the isotropic B factor, is used to model the temperature factor for each atom independently. However, the large number of independent thermal parameters in the isotropic B-factor model, which is generally equal to the number of non-H atoms in the asymmetric unit, imposes a severe burden for lower resolution refinement owing to the low data-to-parameter ratio (risk of overfitting). Moreover, the independence of atomic thermal factors across proteins results in many cases in poor representation of collective molecular motions. This is a particularly severe problem for large and flexible complex structures, the functions of which often involve long-range collective deformations. Finally, many large complexes contain highly flexible structural components that undergo orientation-specific (anisotropic) deformations that would be more accurately described by anisotropic B-factors. However, the highly flexible components in large complexes very frequently compromise the resolution of diffraction, making a full-scale anisotropic refinement that requires six independent thermal parameters for each non-H atom in the asymmetric unit impossible.

Table 1
Comparison of R_{cryst} and R_{free} factors and thermal parameters for the original models, normal-mode refined models and subsequent multi-group TLS refinement for all structures.

PDB code	2a0l	2bbj	2pm9	1suq	2vv5	2aeq	2iny	2v1c
Original R_{cryst}/R_{free} (%)	34.9/37.9	34.8/39.5	23.9/28.7	26.0/31.6	29.3/31.2	28.5/29.1	34.4/43.3	43.1/45.7
NMRef R_{cryst}/R_{free} (No. of parameters) (%)	30.0/35.0 (1203)	29.6/35.8 (1980)	22.4/25.9 (42)	24.0/29.0 (588)	26.1/26.8 (147)	25.0/27.4 (588)	31.4/40.1 (105)	28.3/38.0 (63)
NMRef + TLS R_{cryst}/R_{free} (No. of parameters) (%)	35.1/38.2 (2260)	29.3/34.6 (1500)	21.1/24.3 (600)	25.5/29.9 (200)	25.8/26.8 (420)	23.0/27.0 (900)	32.3/40.8 (200)	29.1/37.9 (600)

One approach for modeling anisotropic thermal motions without a dramatic increase in adjustable parameters is the translation–libration–screw (TLS) model (Schomaker & Trueblood, 1968; Painter & Merritt, 2006). This is an established method with many applications in the literature (Painter & Merritt, 2006). The main concept of TLS refinement is to treat the motion of either the entire protein or parts of it as rigid bodies. Thus, the anisotropic displacement parameters (ADPs) of each atom can be modeled using a drastically reduced number of parameters. Despite its widely observed success, TLS refinement has certain limitations, most of which are related to its rigid-body assumption. For example, the motions between rigid bodies are not interdependent, thus making the TLS model less collective for large-conformational deformations. Another issue of TLS is that the grouping scheme in multi-group refinement can be tricky. The best systematic way of grouping is derived from the distribution of converged isotropic B factors (Painter & Merritt, 2006). Such a feature raises an issue regarding the reliability, or even the possibility, of successful grouping in cases where the isotropic B factors are poorly or not at all converged: a case that is often seen in structural refinement at lower resolutions.

Very recently, a new anisotropic B-factor refinement method, normal-mode-based X-ray crystallographic refinement (NMRef), has been developed (Poon et al., 2007; Chen et al., 2007, 2009). In this approach, the ADPs are reconstructed using a small set of low-frequency normal modes. A major advantage of NMRef is that the modes of motion are collective across the entire system, so that it is expected to be more physically meaningful. Although the concept of NMRef was proposed long ago (Diamond, 1990; Kidera & Go, 1990, 1992; Kidera et al., 1992a,b, 1994), the method was not successfully applied to refinement at moderate diffraction resolution until very recently, when a modified elastic network model for coarse-grained normal-mode calculation was developed (Lu et al., 2006; Lu & Ma, 2008). The new elastic network model delivers reliable eigenvectors for low-frequency modes without requiring initial energy minimization. The new NMRef has been successfully applied to improving the structural refinement of a supramolecular complex (Poon et al., 2007), a membrane-bound ion channel (Chen et al., 2007) and a heavily glycosylated protein (Chen et al., 2009). The application of normal modes to X-ray crystallography other than modeling anisotropic temperature factors has also been reported in the literature (Suhre & Sanejouand, 2004; Lindahl et al., 2006; Delarue & Dumas, 2004; Kundu et al., 2002;

Kondrashov et al., 2006, 2007; Schroder et al., 2007; Hinsen et al., 2005; Ma, 2005).

In this paper, we present a set of cases to demonstrate when and how NMRef is employed to improve structural refinement in X-ray crystallography. The conditions under which NMRef becomes most effective include low resolutions, low data-to-parameter ratios, high structural flexibility as evidenced in B factors or poorly converged B factors. In addition, we have also demonstrated that in some cases the subsequent use of TLS on a normal-mode refined model allowed further improvement of the structures. We expect that this systematic study will provide essential guidance for other users who wish/need to conduct anisotropic B-factor refinement but are restricted by the number of diffraction data.

2. Methods and materials

2.1. Normal-mode-based X-ray crystallographic refinement of anisotropic B factors and positional refinement

The theory of the normal-mode-based refinement method has been reported in previous publications (Poon et al., 2007; Chen et al., 2007). In order to eliminate the differences in R factors from different refinement programs, the original model was first used as input to REFMAC5 to update the R factors, which were used in subsequent comparisons with normal-mode refined structures. In the normal-mode-based refinement, the optimal values of the cutoff distance and stiffness were empirically determined for each individual system as judged by the lowest R_{cryst} and R_{free} values in subsequent normal-mode refinement. A different number of normal modes (N), corresponding to a range of different refinement parameters [equal to $N(N + 1)/2 + 27$], were used in normal-mode refinement to generate anisotropic B factors. The anisotropic B factors then replaced the isotropic B factors in the original model, which was then subjected to REFMAC5 refinement with very tight geometric restraints for positional refinement. The R_{cryst} and R_{free} factors were monitored throughout using the equation

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

in which the same set of free reflections used in the original structure determination was saved for calculation of the R_{free} factor.

For comparison with the isotropic B-factor profile of the original model, 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.

2.2. TLS refinement on normal-mode refined structures

The normal-mode refined structure was submitted to the *TLS Motion Determination* (*TLSMD*) web server. Models with different partitions of each peptide chain were generated. Each model was described by a `tlsin` file containing the parameters fitted for each TLS group and an `xyzin` file containing the atomic coordinates from the input model with modified B_{iso} values. These two files were used in *REFMAC*5 to carry out five cycles of TLS-parameter refinement followed by five cycles of restrained positional refinement. The TLS-parameter refinement and positional refinement were repeated until the lowest R_{free} value was achieved.

3. Results

3.1. Structure of KvAP–33H1 Fv complex at 3.9 Å resolution

The diffraction data for the KvAP–33H1 Fv complex contained 16 234 unique reflections in the resolution range 57.89–3.90 Å. The reported structural model (PDB code 2a0l) refined using *CNS* (Brünger *et al.*, 1998) contained 6740 protein atoms and four potassium ions, with an R_{cryst} and R_{free} of 35.8% and 39.2%, respectively (Lee *et al.*, 2005). This model has an average B factor of 188 Å^2, with the B factors of many atoms exceeding 200 Å^2, indicating a high overall structural flexibility (Fig. 1a, light line). Moreover, the original structure employed a B-factor cutoff of $\sim$240 Å^2, making the B factors of some highly mobile atoms underdetermined.

In order to eliminate the differences in R factors that stemmed from the use of different refinement programs, the normal-mode refinement generally began with a recalculation of the R factors using the original model as input to *REFMAC*5 (Murshudov *et al.*, 1997). The recalculated R_{cryst} and R_{free} factors for the KvAP–33H1 Fv complex were 34.9% and 37.9%, respectively. These values were used as a reference for comparison with the normal-mode-refined model. In order to catch the molecular deformations that are intrinsic to the molecule, the tetrameric biological unit was first generated from symmetry operations. The normal-mode calculation was performed on the tetrameric biological unit with the best combination of cutoff and stiffness values of 13 Å and 30, respectively. Only the portion of eigenvectors corresponding to the asymmetric unit, which was half of the tetramer, was used in the normal-mode anisotropic B-factor refinement (Chen *et al.*, 2007, 2009; Poon *et al.*, 2007) followed by positional refinement using *REFMAC*5 (Murshudov *et al.*, 1997). The refinement yielded lowest R_{cryst} and R_{free} factors of 30.0% and 35.0%, respectively, with the inclusion of 48 modes (1203 parameters; Fig. 1b; Table 1), representing decreases of 4.9% in R_{cryst} and 2.9% in R_{free} (Fig. 9a). It is worth noting that the number of thermal parameters used in the normal-mode-based anisotropic B-factor refinement is more than five times fewer than in the original isotropic B-factor refinement.

In geometry analysis by *MolProbity* (Lovell *et al.*, 2003), the residues in the original model had a distribution of 77.27% in the Ramachandran favored region and 4.66% in the Ramachandran outlier region. After normal-mode refinement, the
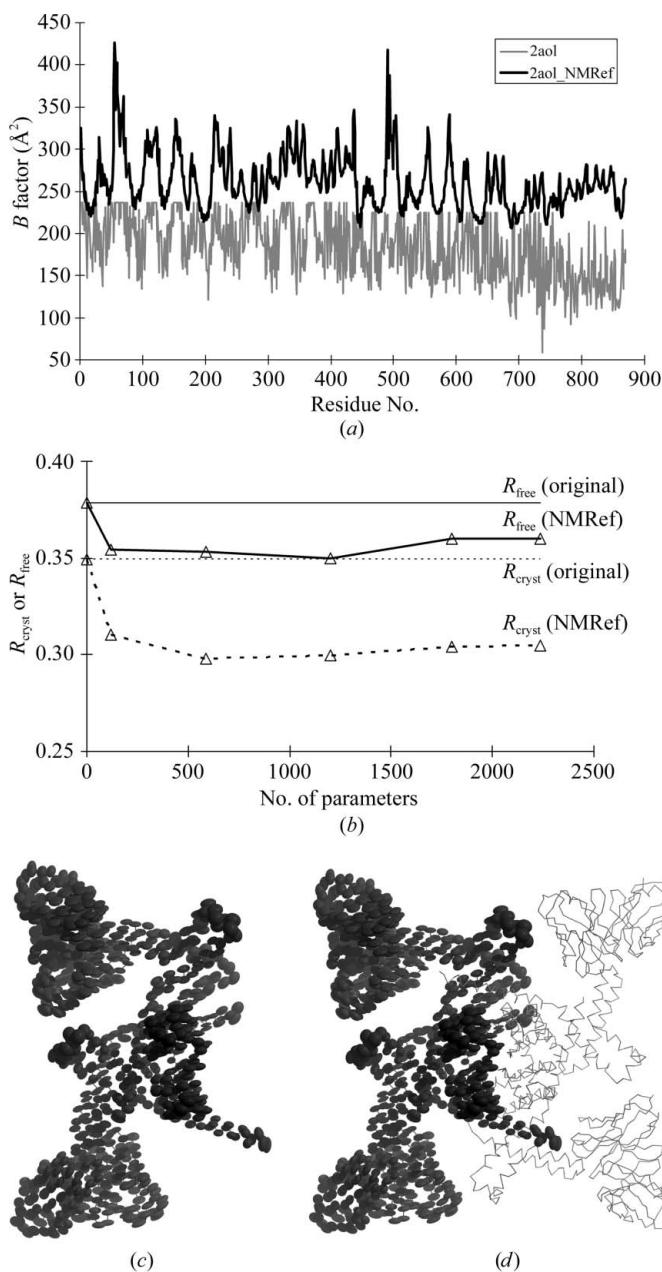
Figure 1
(*a*) B-factor profiles for all C$^\alpha$ atoms in the reported structure of the KvAP–33H1 Fv complex (PDB code 2a0l; light line) and the normal-mode refined structure (dark line). (*b*) The R_{cryst} (dashed line) and R_{free} (solid line) factors with respect to the number of parameters used in the normal-mode refinement. The original values are plotted as a dashed line and a solid line for R_{cryst} and R_{free}, respectively. (*c*) Ellipsoids for the C$^\alpha$ atoms captured from the anisotropic thermal factors derived from normal-mode refinement for the asymmetric unit. (*d*) The biological unit shown together with the ellipsoids for the asymmetric unit. Made with 30% probability.

Table 2
Comparison of residue distributions in the Ramachandran plot for the original models, normal-mode refined models and subsequent multi-group TLS refinement for all structures.

PDB code	Model	Ramachandram favored (%)	Ramachandran outliers (%)
2a0l	Original	77.27	4.66
	NMRef	84.50	3.03
2bbj	Original	90.00	2.30
	NMRef	87.09	2.79
2pm9	Original	74.70	8.38
	NMRef	83.69	6.25
	NMRef + TLS	83.84	5.95
1suq	Original	78.87	4.31
	NMRef	85.54	2.77
2vv5	Original	92.38	2.75
	NMRef	93.84	2.35
	NMRef + TLS	94.01	2.35
2aeq	Original	85.95	3.64
	NMRef	88.43	2.64
	NMRef + TLS	89.92	1.82
2iny	Original	59.21	14.48
	NMRef	65.50	8.87
2v1c	Original	90.79	2.26
	NMRef	77.06	5.33

distribution became 84.50% and 3.03%, respectively, indicating a substantial improvement in geometry (Figs. 9b and 9c; Table 2).

The anisotropic B factors of the normal-mode model were converted to isotropic B factors and are shown in Fig. 1(a) (dark line), which appeared to be a more realistic representation of the thermal motion of a protein at a resolution of 3.9 Å. When the ellipsoid of each C^α atom was mapped to the structure (Figs. 1c and 1d), higher structural flexibility was suggested for the atoms in the peripheral region of the structure.

However, multi-group TLS refinement using the isotropic equivalent B-factor profile from the normal-mode refined structure did not further improve the R_{cryst} or R_{free} factors (Table 1).

3.2. Structure of the CorA Mg^{2+} transporter at 3.9 Å resolution

The reported structure of the CorA Mg^{2+} transporter (PDB code 2bbj) was determined using diffraction data in the resolution range 20.0–3.90 Å with 27 320 unique reflections and the model contained 13 805 non-H protein atoms (Lunin *et al.*, 2006). The published structural model had R_{cryst} and R_{free} factors of 36.3% and 40.6%, respectively. The average B factor for the structural model was 163 Å^2, with an almost flat B-factor distribution (Fig. 2a, light line). Apparently, the flexibility of the protein was not appropriately reflected by the modeled B factors owing to the very low data-to-parameter ratio.

Prior to normal-mode refinement, the R factors for the original PDB structure were recalculated using *REFMAC*5 (Murshudov *et al.*, 1997), yielding R_{cryst} and R_{free} factors of 34.8% and 39.5%, respectively. Since the asymmetric unit of the deposited structure of the CorA Mg^{2+} transporter had a

pentameric organization, which was also its biological form, the normal modes were directly calculated on the pentamer with the optimal combination of cutoff and stiffness values of 13 Å and 3, respectively. The refinement yielded lowest R_{cryst}
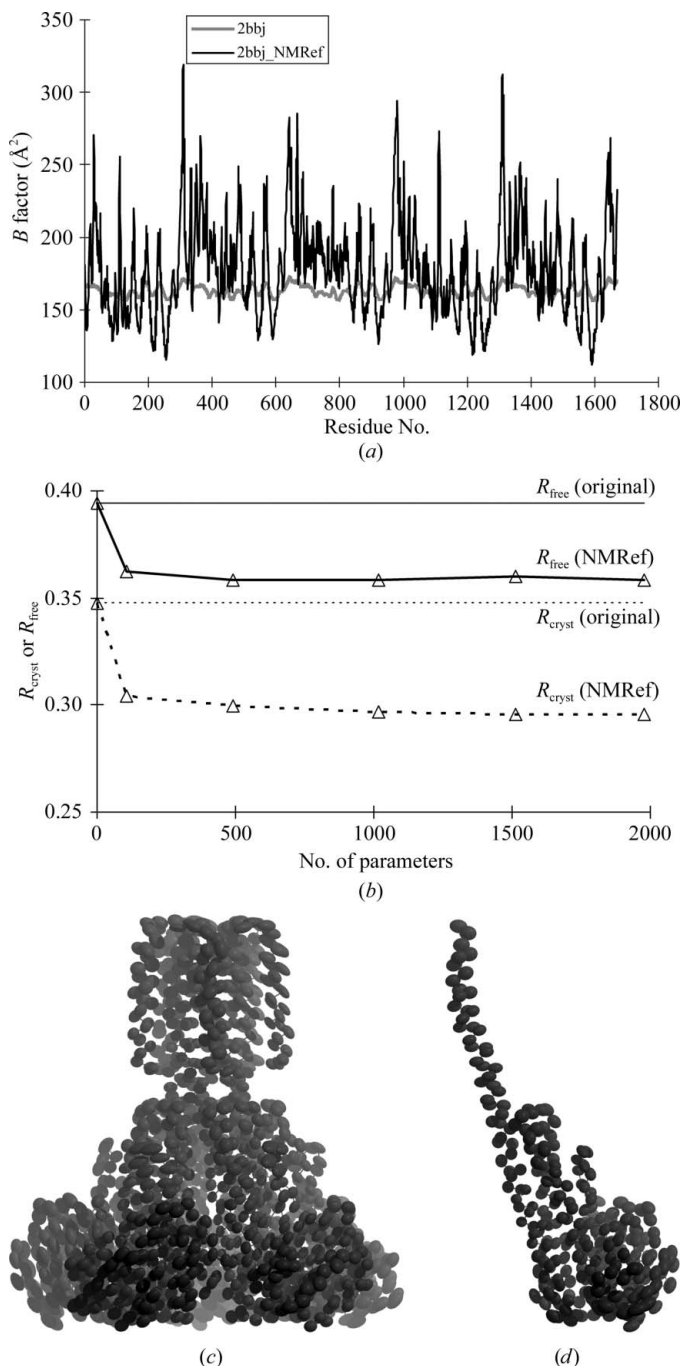
Figure 2
(a) B-factor profiles for all C^α atoms in the original structure of CorA Mg^{2+} transporter (PDB code 2bbj; light line) and the normal-mode refined structure (dark line). (b) The R_{cryst} (dashed line) and R_{free} (solid line) factors with respect to the number of parameters used in the normal-mode refinement. The original values are plotted as a dashed line and a solid line for R_{cryst} and R_{free}, respectively. (c) Ellipsoids for the C^α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. (d) Ellipsoids for a single chain of the homo-pentamer are shown. Made with 30% probability.

and R_{free} factors of 29.6% and 35.8%, respectively (Fig. 2*b*; Table 1). In relation to the original structure, the R_{cryst} and

(a)

(b)

(c)
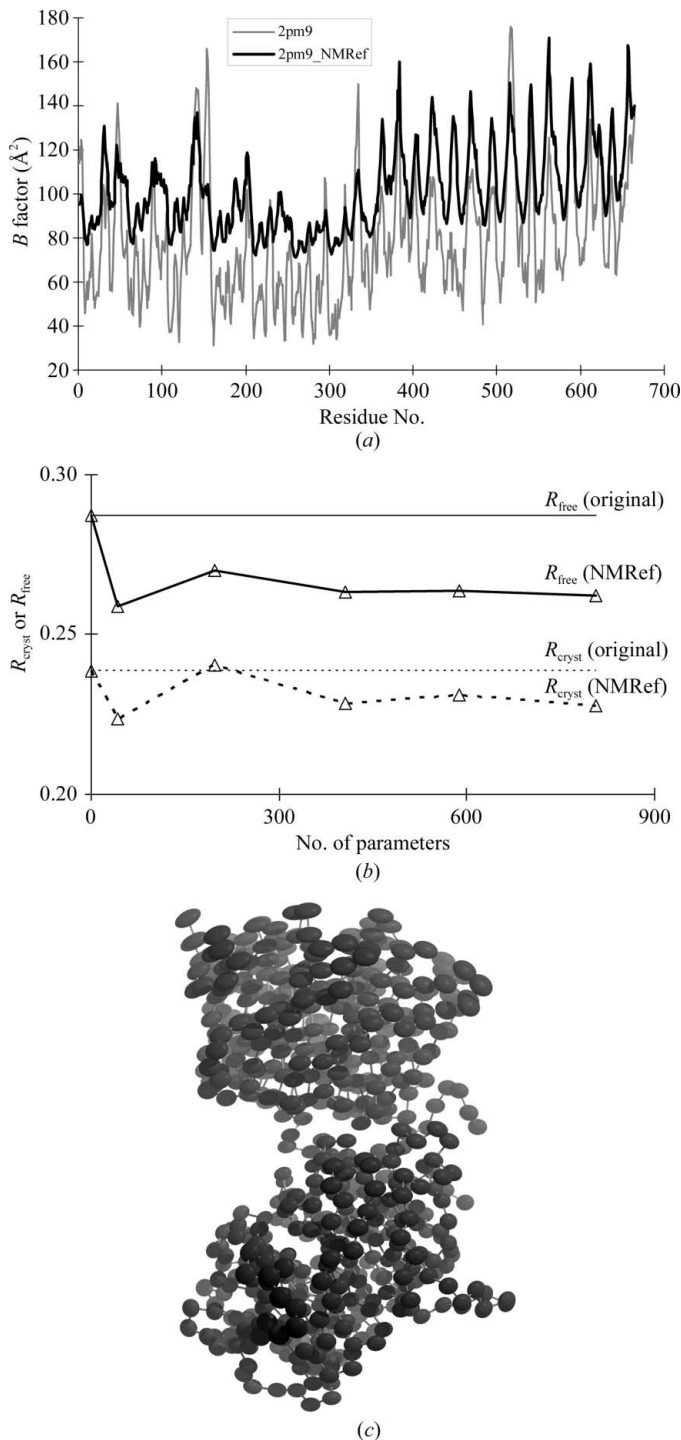
Figure 3
(*a*) *B*-factor profiles for all C$^\alpha$ atoms in the original structure of yeast Sec13/31 vertex element of the COPII vesicular coat (PDB code: 2pm9; light line) and the normal-mode refined structure (dark line). (*b*) The R_{cryst} (dashed line) and R_{free} (solid line) factors with respect to the number of parameters used in the normal-mode refinement. The original values are plotted as a dashed line and a solid line for R_{cryst} and R_{free}, respectively. (*c*) Ellipsoids for the C$^\alpha$ atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. Made with 30% probability.

R_{free} factors were decreased by 5.2% and 3.7%, respectively (Fig. 9*a*). This improvement was achieved using 62 modes, which is equivalent to 1980 thermal parameters or about 1/7 of the thermal parameters used in the original structural refinement.

The residues in the original model had a distribution of 90.00% and 2.30% in the Ramachandran most favored and outlier regions, respectively (Lovell *et al.*, 2003). For the normal-mode refined model, the distribution became 87.09% and 2.79% in the corresponding regions, which is comparable to that of the original model (Figs. 9*b* and 9*c*; Table 2).

The *B*-factor distribution of the normal-mode refined structural model is shown in Fig. 2(*a*) (black line). Mapping it to the structure revealed much smaller *B* factors for the transmembrane helices and higher *B* factors for the peripheral region of the cytoplasmic domain (Figs. 2*c* and 2*d*), consistent with the overall architecture and function of the transporter.

Moreover, multi-group TLS refinement was conducted on the normal-mode refined structure and further reduced the R_{cryst} and R_{free} factors by 0.3% and 1.2%, respectively, using 15 groups per chain (1500 parameters; Table 1).

3.3. Structure of the yeast Sec13/31 vertex element of the COPII vesicular coat at 3.3 Å resolution

The structure of yeast Sec13/31 vertex element (PDB code 2pm9) contained 5170 non-H protein atoms and 28 water molecules (Fath *et al.*, 2007). It was refined against diffraction data in the resolution range 40.0–3.30 Å (18 996 unique reflections) to obtain R_{cryst} and R_{free} factors of 25.1% and 30.5%, respectively. Although the average *B* factor of the model was 79 Å^2, some regions had *B* factors of over 170 Å^2, indicating rather high local flexibility (Fig. 3*a*, light line).

The original model was input to *REFMAC*5, yielding R_{cryst} and R_{free} factors of 23.9% and 28.7%, respectively. The normal-mode calculation was carried out on the asymmetric unit, which was also the biological form, and the best combination of cutoff and stiffness values was 13 Å and 100, respectively. Lowest R_{cryst} and R_{free} factors of 22.4% and 25.9% were obtained (Fig. 3*b*), representing decreases of 1.5% and 2.8% in R_{cryst} and R_{free}, respectively (Fig. 9*a*; Table 1). The improvement was achieved using five normal modes (42 parameters), corresponding to a 100-fold reduction in thermal parameters compared with the original structural model.

The residues in the original model have a distribution of 74.70% and 8.38% in the Ramachandran favored and outlier regions, respectively. The distribution became 83.69% and 6.25% for the residues in the normal-mode refined structure, indicating a substantial improvement in geometry over the original model (Figs. 9*b* and 9*c*; Table 2).

The isotropic *B* factors for each C$^\alpha$ atom, converted from the anisotropic *B* factors of the normal-mode model, are shown in Fig. 3(*a*) (dark line). Mapping the anisotropic *B* factors onto the structure (Fig. 3*c*) revealed highly ordered core residues and much more mobile regions on the surface of the protein complex.

Subsequent multi-group TLS refinement of the normal-mode-refined structure resulted in further improvement. The lowest R_{cryst} and R_{free} factors were 21.1% and 24.3%, respectively, when each chain was divided into 15 groups (600 parameters). This corresponds to decreases of 1.3% and 1.6% in R_{cryst} and R_{free}, respectively. In addition, the geometry was also improved slightly, with 83.84% and 5.95% of residues in the Ramachandran favored and outlier regions, respectively (Figs. 9b and 9c; Table 2).

3.4. Structure of HIV-1 reverse transcriptase (RT) in complex with Jassen-R185545 at 3.0 Å resolution

The original structure of the HIV-1 RT complex (PDB code 1suq) contained 8055 non-H protein atoms and 29 heterogen atoms and was refined against diffraction data in the resolution range 19.99–3.0 Å with 28 520 unique reflections (Das *et al.*, 2004). The final model was refined using *CNS*, with R_{cryst} and R_{free} factors of 26.2% and 32.5%, respectively. It has an average B factor of 84 Å^2. However, the large B factors in some regions (over 100 Å^2) indicated a high degree of local structural flexibility (Fig. 4a, light line).

The original model was input to *REFMAC*5 to recalculate the R factors, yielding R_{cryst} and R_{free} factors of 26.0% and 31.6%, respectively. The normal-mode calculation was directly performed on the asymmetric unit and the optimal combination of cutoff and stiffness values was 13 Å and 30, respectively. With the inclusion of 33 modes (588 parameters), the refinement yielded lowest R_{cryst} and R_{free} factors of 24.0% and 29.0%, respectively (Fig. 4b), representing a decrease of 2.0% in R_{cryst} and 2.6% in R_{free} (Fig. 9a; Table 1).

The residues of the original structural model had a distribution of 78.87% and 4.31% in the Ramachandran favored and outlier regions, respectively. This distribution became 85.54% and 2.77% in corresponding regions for the normal-mode refined structure, indicating a substantial improvement in geometry over the original model (Figs. 9b and 9c; Table 2).

Overall, the isotropic B factors converted from the anisotropic B factors of the normal-mode model agreed well with those of the original structure (Fig. 4a, dark line). However, a number of regions near the surface of the structure (Fig. 4c) appeared to be of much higher mobility, consistent with the overall architecture of the enzyme.

In addition, subsequent multi-group TLS refinement did not further improve the refinement (Table 1).

3.5. The open structure of mechanosensitive channel at 3.45 Å resolution

The crystal structure of mechanosensitive channel MscS in its open form (PDB code 2vv5) was determined to 3.45 Å resolution using 46 433 unique reflections in the resolution range 37–3.45 Å (Wang *et al.*, 2008). The reported model contained 13 692 non-H protein atoms and was refined in *REFMAC*5 with R_{cryst} and R_{free} factors of 29.3% and 31.2%, respectively. In the original structure determination, four TLS groups, one for TM1 and TM2, one for TM3a and TM3b, one for the first cytoplasmic domain and one for the second and

third cytoplasmic domains, were used to facilitate the refinement. This system provided a test case to determine whether normal-mode-based refinement is able to further improve TLS-refined structures.

The original model was first input to *REFMAC*5 to calculate R factors, yielding R_{cryst} and R_{free} factors of 29.3% and 31.2%,
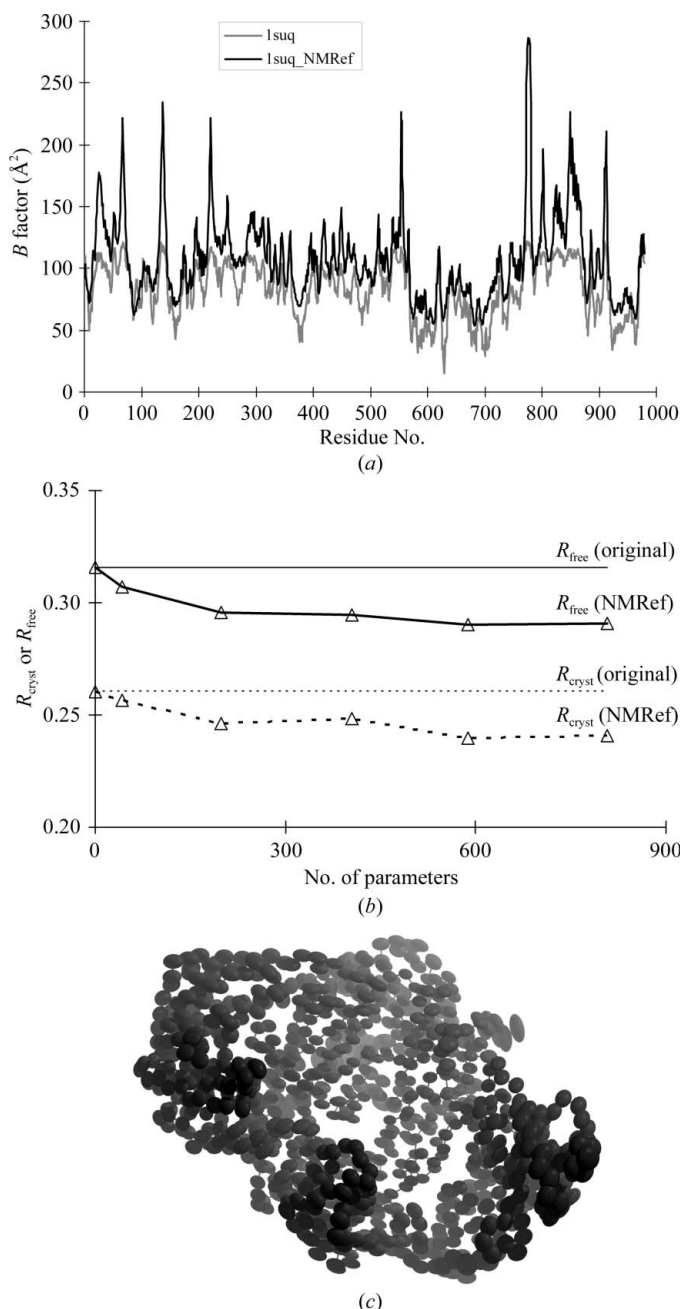
Figure 4
(a) B-factor profiles for all C^α atoms in the original structure of HIV-1 reverse transcriptase (RT) in complex with Jassen-R185545 (PDB code 1suq; light line) and the normal-mode refined structure (dark line). (b) The R_{cryst} (dashed line) and R_{free} (solid line) factors with respect to the number of parameters used in the normal-mode refinement. The original values are plotted as a dashed line and a solid line for R_{cryst} and R_{free}, respectively. (c) Ellipsoids for the C^α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. Made with 50% probability.

respectively: the same as the published values. The normal-mode calculation was directly performed on the heptamer in

(a)

(b)

(c)   *(d)*
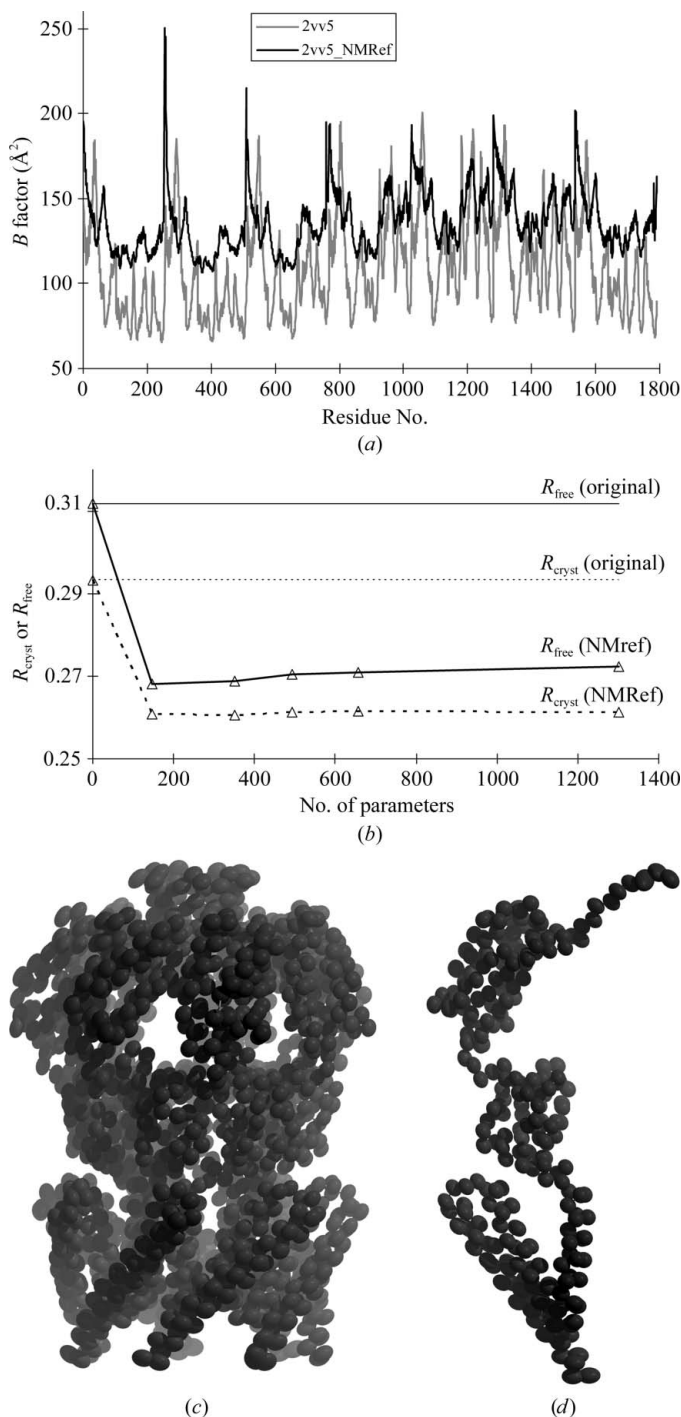
Figure 5
(a) B-factor profiles for all C^α atoms in the original structure of *E. coli* mechanosensitive channel MscS (PDB code 2vv5; light line) and the normal-mode refined structure (dark line). *(b)* The R_{cryst} (dashed line) and R_{free} (solid line) factors with respect to the number of parameters used in the normal-mode refinement. The original values are plotted as a dashed line and a solid line for R_{cryst} and R_{free}, respectively. *(c)* Ellipsoids for the C^α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. *(d)* Ellipsoids on a single chain of the homo-heptamer. Made with 50% probability.

the asymmetric unit and the optimal combination of cutoff and stiffness values was 13 Å and 3, respectively. With the inclusion of 15 modes (147 parameters), the normal-mode refinement yielded lowest R_{cryst} and R_{free} factors of 26.1% and 26.8% (Fig. 5*b*), representing decreases of 3.2% and 4.4% in R_{cryst} and R_{free}, respectively (Fig. 9*a*; Table 1), with a nearly 100-fold reduction in the number of thermal parameters.

The isotropic B-factor profile converted from the anisotropic B-factor profile in the normal-mode refined model is shown in Fig. 5(*a*) (dark line). Overall, the normal-mode model appeared to be more reasonable for the resolution of this structure (Figs. 5*c* and 5*d*).

The residues in the original structure have a distribution of 92.38% and 2.75% in the Ramachandran favored and outlier regions, respectively. For the normal-mode refined model, the distribution became 93.84% and 2.35% in the corresponding regions, which is comparable to that of the original model (Figs. 9*b* and 9*c*; Table 2). However, the number of bad contacts decreased to seven in the normal-mode refined model from 37 in the original model, indicating a slight improvement in geometry.

Subsequent multi-group TLS refinement was carried out based on the normal-mode refined model. When each chain was divided into three groups (totaling 420 parameters), the R_{cryst} factor was further reduced by 0.3%, while the R_{free} factor remained the same (Table 1). The distribution in the Ramachandran plot of the multi-group TLS-refined model did not differ greatly from the normal-mode refined model (Figs. 9*b* and 9*c*; Table 2).
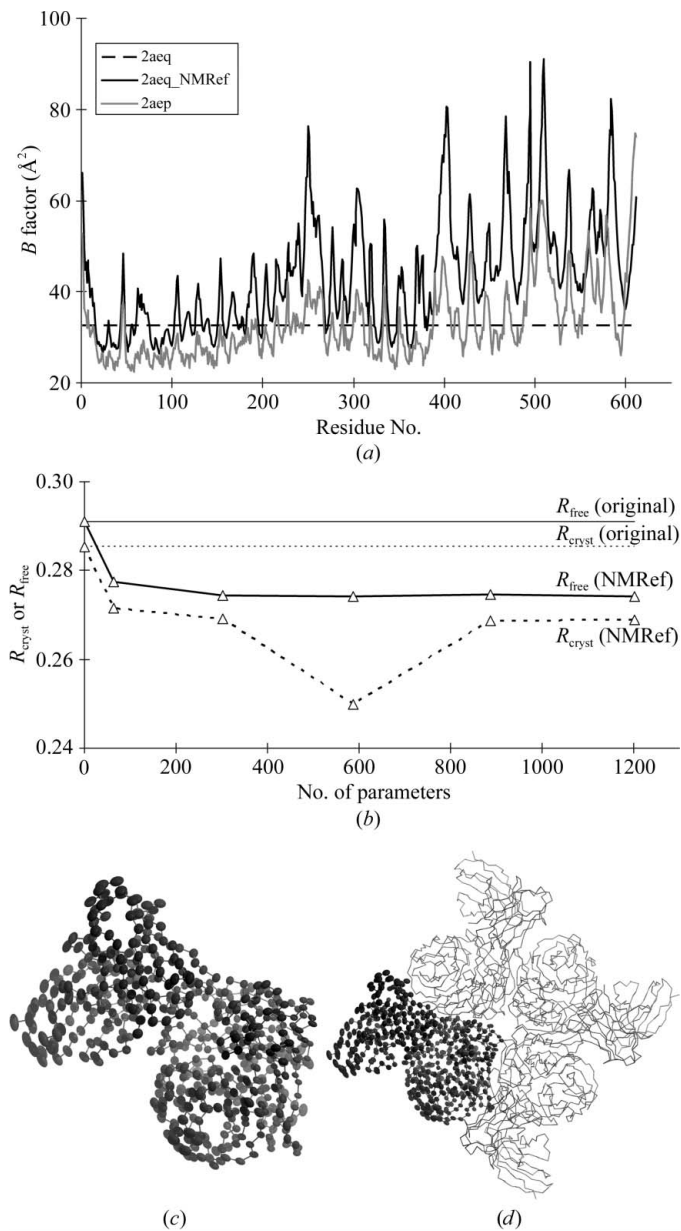
3.6. Structure of influenza virus neuraminidase and its antibody at 3.0 Å resolution

The structure of influenza virus neuraminidase and its antibody was determined to 3.0 Å resolution from diffraction data collected at room temperature (295 K; PDB code 2aeq; Venkatramani *et al.*, 2006). The diffraction data consisted of a total of 23 223 unique reflections, just about four times the total of non-H atoms in the final structural model (4760 protein atoms and 142 heterogen atoms). Probably in order to improve the data-to-parameter ratio, the model was refined using *CNS* (Brünger *et al.*, 1998) with a constant isotropic B factor of 32.64 Å² for all atoms and had R_{cryst} and R_{free} factors of 26.7% and 31.2%, respectively (Fig. 6*a*, light line).

The structural model was first input to *REFMAC*5, yielding R_{cryst} and R_{free} factors of 28.5% and 29.1%, respectively (Table 1). Although neuraminidase is naturally a homo-tetramer, there is only one monomer in the asymmetric unit of this structure (Venkatramani *et al.*, 2006). Thus, normal-mode calculations were performed on the tetramer with the optimal combination of a cutoff distance of 13 Å and a stiffness of 3. The plot of R_{cryst} and R_{free} factors as a function of the number of parameters used in the refinement suggested that the inclusion of 33 modes (totaling 588 parameters) resulted in the lowest R_{cryst} and R_{free} factors (25% and 27.4%, respectively) (Fig. 6*b*). This represents decreases of 3.5% and 1.7% in the R_{cryst} and R_{free} factors, respectively, which were achieved by
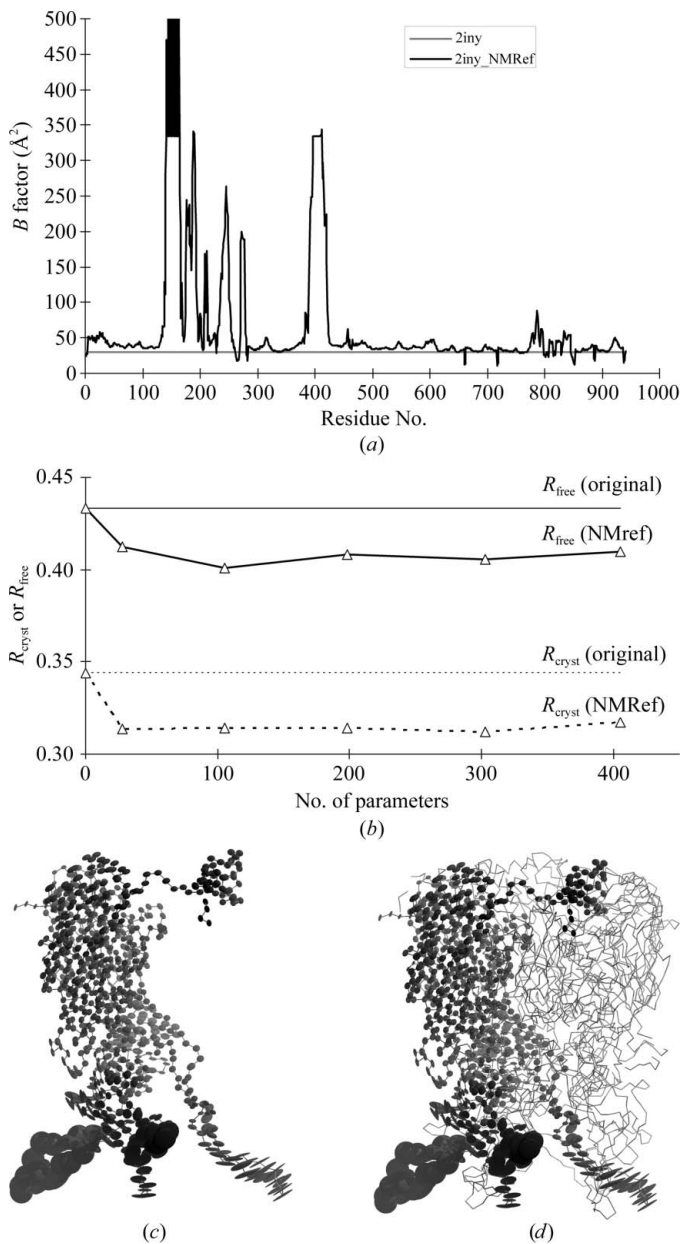
using approximately tenfold fewer thermal parameters than in the original structural refinement (Fig. 9a; Table 1).

In the Ramachandran plot, the original model had 85.95% and 3.64% of the residues in the Ramachandran favored and outlier regions, respectively. The normal-mode refined structural model had a slightly improved geometry, with 88.43% and 2.64% of the residues in these regions, respectively (Figs. 9b and 9c; Table 2).

A second structure of neuraminidase was also reported which was determined from a frozen but otherwise identical crystal at 2.1 Å resolution (PDB code 2aep; Venkatramani et al., 2006). For comparison, Fig. 6(a) shows the isotropic equivalent B factors for C^α atoms in the normal-mode refined structure, the constant B factors in the original structure (PDB code 2aeq) at 3.0 Å resolution and the isotropic B factors in

Figure 6
(a) B-factor profiles for all C^α atoms in the original structure of influenza virus neuraminidase and its antibody (PDB code 2aeq; dashed line), its high-resolution structure (PDB code 2aep; light solid line) and the normal-mode refined structure (dark solid line). (b) The R_{cryst} (dashed line) and R_{free} (solid line) factors with respect to the number of parameters used in the normal-mode refinement. The original values are plotted as a dashed line and a solid line for R_{cryst} and R_{free}, respectively. (c) Ellipsoids for the C^α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. (d) The biological unit is shown together with the ellipsoids for the asymmetric unit. Made with 50% probability.

Figure 7
(a) B-factor profiles for all C^α atoms in the original structure of chicken embryo lethal orphan adenovirus major coat protein (PDB code 2iny; light line) and the normal-mode refined structure (dark line). (b) The R_{cryst} (dashed line) and R_{free} (solid line) factors with respect to the number of parameters used in the normal-mode refinement. The original values are plotted as a dashed line and a solid line for R_{cryst} and R_{free}, respectively. (c) Ellipsoids for the C^α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. (d) The biological unit is shown together with the ellipsoids for the asymmetric unit. Made with 50% probability.

the 2.1 Å resolution structure (2aep). It is clear that the B-factor profile of the normal-mode refined structure matches that of the 2.1 Å structure very well; however, it has higher absolute values, which are consistent with its lower resolution. Thus, it is evident that normal-mode refinement delivers a reasonable B-factor profile for 2aeq that is otherwise not practical in conventional refinement. In addition, the ellipsoids mapped on the structure revealed higher structural flexibility for the outer regions of the tetrameric structure, as expected for protein structures (Fig. 6c and 6d).

Using the B-factor profile from the normal-mode refined structure, automated multi-group TLS refinement was conducted to determine whether further decreases in the R factors can be achieved. Indeed, the R_{cryst} and R_{free} factors were further reduced by 2% and 0.4%, respectively, using 15 groups per chain (equal to 900 parameters; Table 1). Moreover, the normal-mode and TLS model had a distribution of 89.92% and 1.82% in the Ramachandran favored and outlier regions, respectively, indicating a slight further improvement of the geometry.

3.7. Structure of the chicken embryo lethal orphan adenovirus major coat protein hexon at 3.9 Å resolution

The reported model of hexon (PDB code 2iny) had R_{cryst} and R_{free} factors refined in *CNS* (Brünger *et al.*, 1998) of 37.2% and 41.6%, respectively (Xu *et al.*, 2007). The diffraction data in the resolution range 47.06–3.90 Å consisted of 14 124 unique reflections, while the final structural model contained 7523 non-H protein atoms. Thus, the data-to-parameter ratio was rather low. It is probably owing to this reason that the structure was refined by constraining all B factors to 30.0 Å^2 (Fig. 7a, light line). According to the authors, various efforts were made to further refine the structure, but all proved futile.

Prior to normal-mode-based refinement, we recalculated the R factors for the original model using *REFMAC*5, yielding R_{cryst} and R_{free} factors of 34.4% and 43.3%, respectively (Table 1). The normal-mode calculation was performed on the trimeric biological unit and the optimal combination of the cutoff distance and stiffness values was 20 Å and 3, respectively. Only the portion of the eigenvectors corresponding to the monomeric asymmetric unit was included in normal-mode refinement (Fig. 7b). With the inclusion of 12 modes (105 parameters), the refinement yielded lowest R_{cryst} and R_{free} factors of 31.4% and 40.1% (Table 1), representing decreases of 3.0% and 3.2% in R_{cryst} and R_{free}, respectively (Fig. 9a). The improvement was achieved by using 70-fold fewer thermal parameters than in the original structural refinement.

The residues in the original model had a distribution of 59.21% and 14.48% in the Ramachandran favored and outlier regions, respectively. After normal-mode refinement, the distribution became 65.50% and 8.87% in the corresponding regions, indicating a substantial improvement in geometry over the original model (Figs. 9b and 9c; Table 2).

The isotropic B factors converted from the anisotropic B factors in the normal-mode refined model are compared with those of the original model in Fig. 7(a) (dark line). Although the majority of the structure has rather low B factors, there are a number of regions of extremely high mobility (Fig. 7a). As expected, these regions are located far away from other structural components in the native trimeric structure (Figs. 7c and 7d).

However, subsequent multi-group TLS refinement of the normal-model refined structure did not result in further decreases in the R factors (Table 1). Thus, in this case no gain was obtained by further applying TLS refinement.

3.8. Structure of *Deinococcus radiodurans* RecR and RecO (drRecOR) at 3.8 Å resolution

The original structural model of drRecOR (PDB code 2v1c) contained a total of 4778 non-H protein atoms and three Zn atoms and was refined against diffraction data in the resolution range 3.80–71.25 Å (with 6850 unique reflections). Owing to the very low data-to-parameter-ratio, the reported refinement was terminated after a single round of rigid-body refinement and manual rebuilding, with R_{cryst} and R_{free} factors of 45.8% and 44.3%, respectively (Timmins *et al.*, 2007). The final model has a constant B factor of 96 Å^2 (Fig. 8a, light line).

Prior to normal-mode-based refinement, the original model was input to *REFMAC*5, yielding R_{cryst} and R_{free} factors of 43.1% and 45.7%, respectively (Table 1). The normal-mode calculation was performed on the dimeric biologic unit and the best combination of cutoff and stiffness values was 13 Å and 100, respectively. The portion of the eigenvectors corresponding to the monomeric asymmetric unit was included in the normal-mode refinement using a different number of normal modes (Fig. 8b). The lowest R_{cryst} and R_{free} factors were 28.3% and 38.0%, respectively, when the eight lowest normal modes (equivalent to 63 thermal parameters) were used (Table 1). Compared with the original model, this represents decreases of 14.8% and 7.7% in R_{cryst} and R_{free}, respectively (Fig. 9a).

The residues in the original model are distributed with 90.79% and 2.26% in the Ramachandran favored and outlier regions, respectively. The distribution became 77.06% and 5.33%, respectively, for the normal-mode refined structure. The poorer geometry of the normal-mode refined model is probably a consequence of the fact that the original model of drRecOR was constructed based on structures of drRecO and drRecR determined at much higher resolution without further positional refinement. However, the number of bad contacts decreased to 23 in the normal-mode refined structure from 58 in the original model.
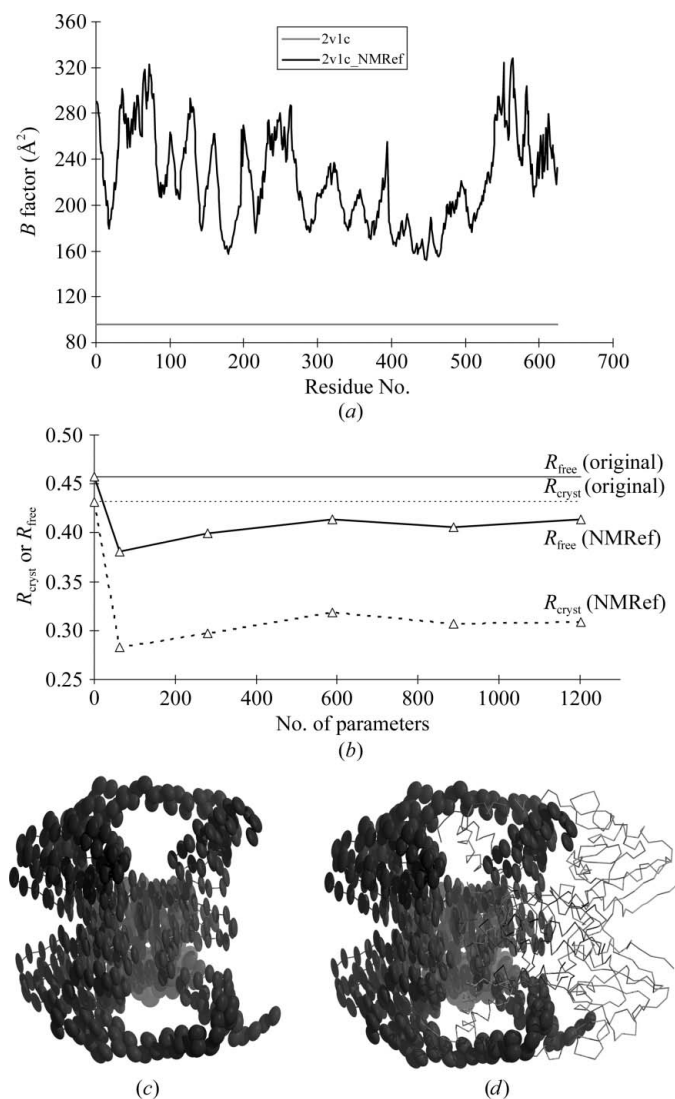
The B factors from the normal-mode refined model (Fig. 8a, dark line) suggest realistic flexibility variations along the structure, with a more ordered region inside and more flexible regions outside (Figs. 8c and 8d).

In addition, subsequent multi-group TLS refinement of the normal-mode refined structure did not produce any remarkable improvement. Although the R_{free} factor decreased to 37.9% with ten TLS groups per chain, the R_{cryst} factor increased to 29.1% (Table 1).

4. Concluding discussion

Structural refinement at limited resolution is generally difficult and very inefficient. Here, we present the application of a new normal-mode-based X-ray crystallographic refinement method (NMRef) to a total of eight low-resolution crystal structures (in the resolution range 3.0–3.9 Å). All these structures have low data-to-parameter ratios (ranging from far below to barely 1.0). The first two systems (PDB codes 2a0l and 2bbj) have very large average B factors, indicating a high overall structural flexibility. Two other systems (PDB codes 2pm9 and 1suq) contain structural regions with high B factors. Possibly owing to the limited data-to-parameter ratio, three

systems (PDB codes 2aeq, 2iny and 2v1c) have a constant B factor for all atoms and one system (PDB code 2bbj) has an

Figure 8
(a) B-factor profiles for all C^α atoms in the original structure of *D. radiodurans* RecR and RecO (PDB code 2v1c; light line) and the normal-mode refined structure (dark line). (b) The R_{cryst} (dashed line) and R_{free} (solid line) factors with respect to the number of parameters used in the normal-mode refinement. The original values are plotted as a dashed line and a solid line for R_{cryst} and R_{free}, respectively. (c) Ellipsoids for the C^α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. (d) The biological unit is shown together with the ellipsoids for the asymmetric unit. Made with 50% probability.
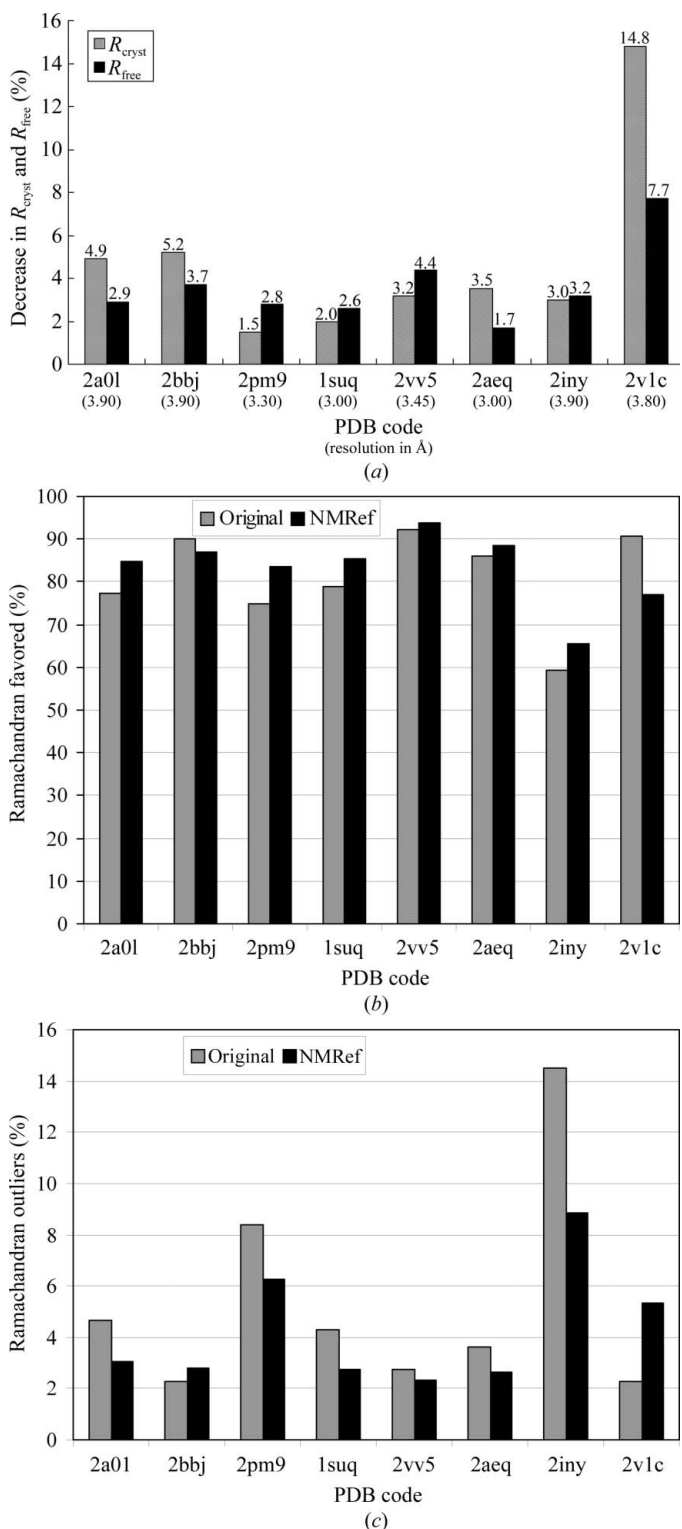
Figure 9
Structural improvements produced by normal-mode refinement over the original models. (a) Decreases in R factors. Grey bars represent decreases in R_{cryst} and black bars those in R_{free}. (b) Percentage of residues distributed in the Ramachandran favored region and (c) percentage of residues distributed in the Ramachandran outlier region, as calculated using *MolProbity* (Lovell *et al.*, 2003). Grey bars are for the original models and black bars are for the normal-mode refined models.

almost constant B factor for all atoms. Of the eight structures, three are membrane proteins and five are soluble supramolecular complexes.

In all cases, the application of NMRef substantially improved the fit between the structural models and the diffraction data as reflected by decreases in the R factors (Fig. 9). The decreases are in the range 1.7–7.7% for R_{free} factors and 1.5–14.8% for R_{cryst} factors, and average at 3.6% for R_{free} and 4.8% for R_{cryst}. In addition, in most cases there is a concomitant improvement in the geometry of the structural model, as reflected by an improvement of the residue distribution in the Ramachandran plot and a reduction in the number of bad contacts. It is worthwhile emphasizing that these substantial improvements were achieved by only one round of normal-mode-based refinement without any manual adjustment, with 5–100-fold fewer thermal parameters than in the original isotropic B-factor refinement. As demonstrated in our previous studies, the use of multiple rounds of normal-mode refinement, combined with diligent manual model adjustment, generally brings about much more substantial improvements (Poon *et al.*, 2007; Chen *et al.*, 2007, 2009).

Since TLS refinement uses a different concept, we tested whether the combined use of TLS with normal-mode refinement will result in further improvement. Indeed, in about half of the cases the application of TLS to normal-mode refined structures further decreased the R factors and it sometimes also improved the geometry. The most significant advantage of NMRef over TLS is seen in the last three systems in Fig. 9. In these systems, the isotropic B factors were set as a constant in the final structures deposited in the PDB, which made the direct application of multi-group TLS refinement impossible. The application of NMRef is not at all affected by the values of the isotropic B factors. However, it is worth pointing out that in some cases TLS refinement may deliver better R-factor improvement than NMRef. This is because the independence of the body motion in TLS may offer a larger flexibility for fitting in cases where structural deformations are less collective.

As clearly indicated from our study on these test systems, it is generally a good idea to test normal-mode refinement using a different number of lowest frequency normal modes. In real applications, we also encourage users to test the combined use of normal-mode refinement with TLS, which in some cases can result in further improvement in structural refinement. Even for cases in which TLS outperforms NMRef, an initial application of NMRef may improve the subsequent application of TLS because the former can deliver a smoother B-factor distribution for better grouping of TLS.

References

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.

Chen, X., Lu, M., Poon, B. K., Wang, Q. & Ma, J. (2009). *Acta Cryst.* D**65**, 339–347.

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

Das, K. *et al.* (2004). *J. Med. Chem.* **47**, 2550–2560.

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

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

Fath, S., Mancias, J. D., Bi, X. & Goldberg, J. (2007). *Cell*, **129**, 1325–1336.

Hinsen, K., Reuter, N., Navaza, J., Stokes, D. L. & Lacapere, J. J. (2005). *Biophys. J.* **88**, 818–827.

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.

Lee, S. Y., Lee, A., Chen, J. & MacKinnon, R. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 15441–15446.

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

Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.

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.

Lunin, V. V., Dobrovetsky, E., Khutoreskaya, G., Zhang, R., Joachimiak, A., Doyle, D. A., Bochkarev, A., Maguire, M. E., Edwards, A. M. & Koth, C. M. (2006). *Nature (London)*, **440**, 833–837.

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

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

Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.

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.

Timmins, J., Leiros, I. & McSweeney, S. (2007). *EMBO J.* **26**, 3260–3271.

Venkatramani, L., Bochkareva, E., Lee, J. T., Gulati, U., Laver, W. G., Bochkarev, A. & Air, G. M. (2006). *J. Mol. Biol.* **356**, 651–663.

Wang, W., Black, S. S., Edwards, M. D., Miller, S., Morrison, E. L., Bartlett, W., Dong, C., Naismith, J. H. & Booth, I. R. (2008). *Science*, **321**, 1179–1183.

Xu, L., Benson, S. D. & Burnett, R. M. (2007). *J. Struct. Biol.* **157**, 424–431.