Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Error-estimation-guided rebuilding of *de novo* models increases the success rate of *ab initio* phasing

Rojan Shrestha,[a,b] David Simoncini[a] and Kam Y. J. Zhang[a,b]*

[a]Zhang Initiative Research Unit, Advance Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, and [b]Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-0882, Japan

Correspondence e-mail: kamzhang@riken.jp

Received 1 June 2012
Accepted 4 September 2012

Recent advancements in computational methods for protein-structure prediction have made it possible to generate the high-quality *de novo* models required for *ab initio* phasing of crystallographic diffraction data using molecular replacement. Despite those encouraging achievements in *ab initio* phasing using *de novo* models, its success is limited only to those targets for which high-quality *de novo* models can be generated. In order to increase the scope of targets to which *ab initio* phasing with *de novo* models can be successfully applied, it is necessary to reduce the errors in the *de novo* models that are used as templates for molecular replacement. Here, an approach is introduced that can identify and rebuild the residues with larger errors, which subsequently reduces the overall C^{α} root-mean-square deviation (CA-RMSD) from the native protein structure. The error in a predicted model is estimated from the average pairwise geometric distance per residue computed among selected lowest energy coarse-grained models. This score is subsequently employed to guide a rebuilding process that focuses on more error-prone residues in the coarse-grained models. This rebuilding methodology has been tested on ten protein targets that were unsuccessful using previous methods. The average CA-RMSD of the coarse-grained models was improved from 4.93 to 4.06 Å. For those models with CA-RMSD less than 3.0 Å, the average CA-RMSD was improved from 3.38 to 2.60 Å. These rebuilt coarse-grained models were then converted into all-atom models and refined to produce improved *de novo* models for molecular replacement. Seven diffraction data sets were successfully phased using rebuilt *de novo* models, indicating the improved quality of these rebuilt *de novo* models and the effectiveness of the rebuilding process. Software implementing this method, called *MORPHEUS*, can be downloaded from http://www.riken.jp/zhangiru/software.html.

1. Introduction

X-ray crystallography is the principal method for the structure determination of macromolecules, including proteins, to atomic detail. Adequate phases derived from diffraction data using experimental methods are sufficient to solve the structures of even the largest macromolecules, such as eukaryotic ribosome, at atomic resolution (Ben-Shem *et al.*, 2011). In the absence of experimental phases, protein structures can be solved by computational methods such as molecular replacement (MR; Blow & Rossmann, 1961). MR generally requires

template models derived from the structures of homologous proteins.

Recent improvements in computational methods for the prediction of protein structures using only amino-acid sequences, known as *de novo* modelling, have opened a new frontier in structural determination. One of the practical applications of *de novo* modelling has been shown to be the solution of the crystallographic phase problem for new folds (Qian *et al.*, 2007), which can be considered as *ab initio* phasing. Computationally generated models of small proteins have extended the utility of MR in the absence of known starting homologous structures. A successful *de novo* modelling methodology was inspired by the fragment-assembly approach, in which fragments of known structures are combined with the guidance of scoring functions, which consist of all of the major energy terms for protein stability (Bowie & Eisenberg, 1994; Das & Baker, 2008). One of the most successful fragment-assembly methods for protein-structure prediction is *Rosetta* (Rohl *et al.*, 2004), which has been demonstrated to be able to predict the high-quality models necessary for solving the phase problem by MR (Das & Baker, 2009). *Rosetta*'s methodology in the initial stage is to use a coarse-grained model that contains only the main chain and the centroids of side chains in order to search a wider conformational space. In the second stage, it refines the all-atom models derived from coarse-grained models with limited main-chain conformational searches and full side-chain packing and optimization (Das & Baker, 2008). Owing to this two-stage process, the success of all-atom refinement depends highly on the quality of the coarse-grained models generated.

To be successfully used as a template for MR, a predicted model must have the correct fold as present in the target structure. Moreover, a significant portion of the atomic scatterers should spatially match those of the underlying target structure. Much effort has been made to increase the success rate of *ab initio* phasing with *de novo* models. The MR method has generally been executed on models filtered by energy after predicting a large number of models (Das & Baker, 2009). At the other extreme, coarse-grained models or polyalanine models have also been used for MR (Rigden *et al.*, 2008; Das & Baker, 2009). However, the absence of a large portion of the atomic detail appears to be a bottleneck to achieving successful solutions. The number of successful cases in MR trials was significantly increased with models optimized using *Rosetta* all-atom energy (Das & Baker, 2009). Other approaches for increasing the success rate were to spend a large amount of computing power on conformational sampling (Das & Baker, 2009) and to trim the highly flexible loop regions in the predicted models (Rigden *et al.*, 2008). In addition, phasing with intermediate all-atom models during optimization effectively managed the computational time as well as slightly increasing the success rate of MR experiments (Shrestha *et al.*, 2011).

Many factors associated with diffraction data such as resolution, solvent content, noncrystallographic symmetry in the unit cell and others could have significant impact on obtaining successful MR solutions. However, poor correlations have been observed between these factors and the MR success rate using *de novo* models (Das & Baker, 2009; Shrestha *et al.*, 2011). In contrast, improved model quality enabled success in all tested cases (Das & Baker, 2009; Shrestha *et al.*, 2011). Therefore, it is critical to predict highly accurate *de novo* models so that they can be used as templates for successful *ab initio* phasing by MR. There are several areas in which improvements may result in better *de novo* models, such as the improved identification of high-quality fragments, more accurate energy functions and more efficient sampling of conformational space. Recent efforts to improve *de novo* model quality have primarily been focused on conformational sampling (Kim *et al.*, 2009).

The errors in a template model are not uniformly distributed. Removing regions with large errors can produce a template that is closer to the target and increase the chances of success in MR provided that the remaining structure still constitutes a significantly large portion of the scattering matter with respect to the target. Traditionally, many approaches have been employed to increase the success rate of MR from a given template, which is typically from a structural homologue. These include trimming off loops or terminal regions to create a compact core structure, removing side chains to generate a polyalanine model, deleting highly flexible regions identified by high temperature factors in the coordinates and pruning off the side chains of residues that are nonconserved between the template sequence and the target sequence (Stein, 2008). Searching multiple domains or multiple templates simultaneously can also be very powerful in solving difficult cases of MR (McCoy *et al.*, 2007). To take advantage of the ever-increasing number of structures that are being deposited in the Protein Data Bank (Berman *et al.*, 2000), automated pipelines have been created in order to relieve users of the burden of the manual curation of templates for MR, resulting in an increased success rate (Keegan & Winn, 2008; Long *et al.*, 2008). In this paper, the focus is on improving the entire template model for MR.

One way to improve the quality of *de novo* models is to identify loop regions and then focus the conformational sampling on these regions (Canutescu & Dunbrack, 2003; Mandell *et al.*, 2009). The loop region is first identified from the secondary-structure assignment of the predicted models (Kabsch & Sander, 1983). Extensive resampling of these loop regions is then carried out by the cyclic coordinate-descent method inspired by a robotics algorithm. Although it is generally true that loop regions are less accurately predicted, some loops are intrinsically disordered or can adopt multiple conformations. In this scenario, extensive conformational resampling in order to find one energetically most stable conformation may not be fruitful. Moreover, errors in predicted models exist not only in loop regions but also in regular secondary-structure regions.

One important step in improving the *de novo* model quality for *ab initio* phasing could be to initially identify the less accurately predicted regions in the model and then to perform rigorous sampling on these regions. There have been extensive efforts to develop methods that can assess the quality of

computationally predicted models (Kryshtafovych & Fidelis, 2009). These model-quality assessment (MQA) methods have been shown to be very useful in identifying good-quality models and ranking them (Levitt & Gerstein, 1998; Zemla, 2003; Zhang & Skolnick, 2004*b*). Qian *et al.* (2007) used such a strategy to improve the success rate of MR by rebuilding the most conformationally variable regions within an ensemble of structural models. After identifying regions of high conformational variability using a principle similar to PCons (Wallner & Elofsson, 2006), an aggressive sampling was conducted on these regions and the cyclic coordinate-descent method (Canutescu & Dunbrack, 2003) was used to maintain the chain connectivity. The conformational variation has also been exploited as a colony energy for loop prediction (Xiang *et al.*, 2002). When rebuilding was guided by the electron density, the success rate of MR was further improved and many challenging cases could be solved (DiMaio *et al.*, 2011), although an approximate MR solution was required in this case.

Here, we describe a method for improving the quality of *de novo* models. Firstly, local regions in the coarse-grained models with large errors are identified. These errors are estimated by the average pairwise geometric distance per residue computed among selected lowest energy coarse-grained models. Secondly, a rebuilding process that focuses on these more error-prone residues in the coarse-grained models is carried out. Lastly, these rebuilt coarse-grained models are converted into all-atom models and refined with *Rosetta* all-atom energy to produce improved *de novo* models for MR. The score used here for error estimation is similar to many MQA methods. However, instead of a global score for the entire protein model, a per-residue score is calculated in order to identify residues or regions where large errors exist. It was observed in our previous study that computationally predicted *de novo* models could be successfully used for phasing by MR. However, half of the tested targets were not able to succeed in MR trials, primarily owing to a lack of sufficiently accurate input models (Shrestha *et al.*, 2011). These difficult targets were used as test cases for our method to produce improved *de novo* models suitable for phasing by rebuilding local segments that contain large errors. The results showed that the coarse-grained models were refined closer to the native structures and the success rate of phasing with these models after all-atom refinement was significantly increased.

2. Methods

This method is designed to reduce the distance between the coarse-grained models and the native structures. It calculates a geometric distance score for each residue from selected lowest energy coarse-grained models generated by *Rosetta* v.3.2. The geometric distance score for each residue is computed by superposing each model with all of the remaining selected models and calculating a pairwise average root-mean-square deviation on C^{α} atoms (CA-RMSD), which is termed the average pairwise residue distance score (AP-RDS),

$$\text{AP-RDS}(i,j) = \frac{1}{n-1}\sum_{k=1(k\neq j)}^{n-1}[(X_{ij}-X_{ik})^2 + (Y_{ij}-Y_{ik})^2 + (Z_{ij}-Z_{ik})^2]^{1/2}. \quad (1)$$

where i represents the residue number, j represents the model number, k represents all of the other models except model j, n represents the total number of models and X, Y, Z represent the Cartesian coordinates of each C^{α} atom in a residue. The r.m.s.d. from the native for all decoys was computed using the Kabsch algorithm (Kabsch, 1976). Although many different geometric distances, such as GDT (Zemla *et al.*, 1999; Zemla, 2003), MaxSub (Siew *et al.*, 2000), TM-score (Zhang & Skolnick, 2004*b*), Q-score (Ben-David *et al.*, 2009) and percentile-based spread (Pozharski, 2010) can be used, the r.m.s.d. was chosen in this study for its simplicity and generality. This AP-RDS score is computed in a similar way to that implemented in PCons-local (Wallner & Elofsson, 2006), but the score is used not only to identify residue errors but also to estimate the sampling frequency for each residue during rebuilding. The AP-RDS is used to guide conformation sampling during rebuilding so that residues with higher scores are sampled more often than those with lower scores. The coarse-grained models after rebuilding are converted to all-atom models using the fast relax algorithm implemented in *Rosetta* v.3.2 (Tyka *et al.*, 2011). All models after all-atom refinement are tested against the diffraction data for their suitability as templates for solution of the phase problem using the *Phaser* program (McCoy *et al.*, 2007). Our approach is implemented in the C++ programming language and incorporates *Rosetta* and *Phaser* as libraries. It will be referred to as *MORPHEUS* (*MOdel Rebuilding for PHasing with Enhanced sUcceSs*). The number of successful MR solutions determines the continuation or termination of the simulation. Maximum and minimum thresholds for the *Phaser* score are used to control the simulation as in our previous study (Shrestha *et al.*, 2011). Once a few good models have been obtained with *Phaser* scores in the high confidence range, the entire simulation is terminated. In the worst case, MR is run for all generated *de novo* models.

2.1. Data set and coarse-grained model generation

A data set of ten targets that were unsuccessful in MR was chosen from our previous benchmark (Shrestha *et al.*, 2011). Standalone *Rosetta* v.3.2 (Rohl *et al.*, 2004; Tyka *et al.*, 2011; http://www.rosettacommons.org) was executed to generate 300 000 initial coarse-grained models for each target sequence using the RIKEN Integrated Cluster of Clusters (RICC). Fragment libraries for each of these targets were obtained from the *Robetta* server (Chivian *et al.*, 2003). In order to mimic a blind prediction, fragments from the target structure and structures with homologous sequences were excluded from the fragment libraries.

2.2. Determine less accurately predicted regions

1000 lowest energy coarse-grained models were selected from the pool of 300 000 models generated by *Rosetta*. Each

model was superimposed with all other selected models using rigid-body transformation with an optimal translation vector and a rotation matrix that minimizes the sum of the squared distances between two coordinate sets of corresponding atoms (Kabsch, 1976). The AP-RDS was calculated by taking an average of the CA-RMSDs between one model and all other models for each residue position. The correlation between the AP-RDS and the CA-RMSD from the native structure was calculated and used to assess the capability of the AP-RDS to estimate errors in each residue. Furthermore, each coarse-grained model was assigned a score that was the average of the CA-RMSDs between this model and other models covering the entire sequence, which was defined as the average pairwise model distance score (AP-MDS),

$$\text{AP-MDS}(j) = \frac{1}{n-1}\sum_{k=1(k\neq j)}^{n-1}\left[\frac{1}{m}\sum_{i=1}^{m}(X_{ij}-X_{ik})^2+(Y_{ij}-Y_{ik})^2+(Z_{ij}-Z_{ik})^2\right]^{1/2}, \quad (2)$$

where j represents the model number other than model k, i represents the residue number, m represents the total number of residues in the model and n represents the number of models. X, Y, Z represent the Cartesian coordinates of each C^{α} atom in a residue. The AP-MDS can be used to assess the overall quality of coarse-grained models.

2.3. Rebuilding incorrectly predicted residues

A subset of coarse-grained models were selected from the group of 1000 for further model rebuilding. This selection was based on the AP-MDS score of each coarse-grained model. While it is necessary to include a relatively large set of decoys for the calculation of AP-RDS and AP-MDS scores, there is no need to subject all of these models to further rebuilding. Rebuilding only a subset of these models with the lowest AP-MDS scores will enable the inclusion of the majority of high-quality models, with a substantial saving of computational time. In this study, 65% of the models were chosen for subsequent rebuilding. Owing to the long computing time needed to complete the calculation for the entire test set of targets, the choices of selecting the 1000 lowest energy models for AP-RDS and AP-MDS calculation and the subsequent selection of 65% models for rebuilding were not optimized. Alternative choices cannot be exhaustively tested and compared in order to come up with an optimum combination of parameters. These parameters were empirically obtained by testing on the first target and they seemed to work well. Subsequently, they were used for all of the other targets.

During the rebuilding process, the conformation search was biased towards error-prone residues according to the AP-RDS. A total of 5000 rebuilding steps were used for each coarse-grained model and these rebuilding steps were distributed to each residue based on its AP-RDS. This non-uniform sampling was achieved by the roulette-wheel procedure (Supplementary Fig. S1[1]). Instead of including both three-residue and nine-residue fragments as the source for rebuilding, only the three-residue fragment library was used in order to reduce large changes in global conformations. For each coarse-grained model selected, 300 rebuilding trajectories were carried out with random seeds, which was sufficient to explore the conformational space within a reasonable computational time. The *Rosetta* coarse-grained scoring function was used to evaluate the models generated during each rebuilding trajectory. The temperature factor in the Monte Carlo simulated-annealing procedure was modified to make the acceptance rate of high-energy models proportional to the residue error using the equation

$$T_{\text{cur}} = T_{\min} + \frac{(T_{\max}-T_{\min})}{(D_{\max}-D_{\min})}(D_{\text{cur}}-D_{\min}). \quad (3)$$

In the above equation, $D_{\min}$ and $D_{\max}$ are the minimum and maximum AP-RDS of the coarse-grained models. T_{cur} is the current temperature given the value of D_{cur} (the AP-RDS of the current residue). $T_{\min}$ (minimum temperature) and $T_{\max}$ (maximum temperature) were set at 1.0 and 4.0, respectively. Since the local variation (AP-RDS) is used to bias fragment insertion in *ab initio* sampling, the conformation of the downstream residues will be changed upon each new fragment insertion. Regions with large errors will require a large conformational change to correct them and this large conformational change will be likely to cause an energy increase, especially when it affects the downstream well predicted residues, and consequently this change will be rejected. This scenario is mitigated by setting the acceptance threshold (T_{cur}) proportional to the structural variation (D_{cur}) in order to increase the chance of large conformational changes being accepted. Finally, the lowest energy model in each rebuilding trajectory was chosen for side-chain packing and all-atom refinement.

2.4. Side-chain packing and molecular replacement

Each lowest energy model from 300 independent rebuilding trajectories was converted into an all-atom model using the *Rosetta* all-atom conformation-sampling protocol. The *Rosetta* fast relax algorithm (Tyka *et al.*, 2011) was employed to pack the side-chain rotamers and to perform all-atom refinement through energy minimization. *Phaser* (McCoy *et al.*, 2007) was run for each refined all-atom model to assess its quality for phasing. *Phaser* scores were used as a criterion to terminate the entire rebuilding process after a few successful *de novo* models for phasing had been obtained (Shrestha *et al.*, 2011).

3. Results

3.1. The residue divergence score correlates with model accuracy

The correlation between the coarse-grained energy of the models and their accuracy is generally poor when all of the

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: WD5186). Services for accessing this material are described at the back of the journal.

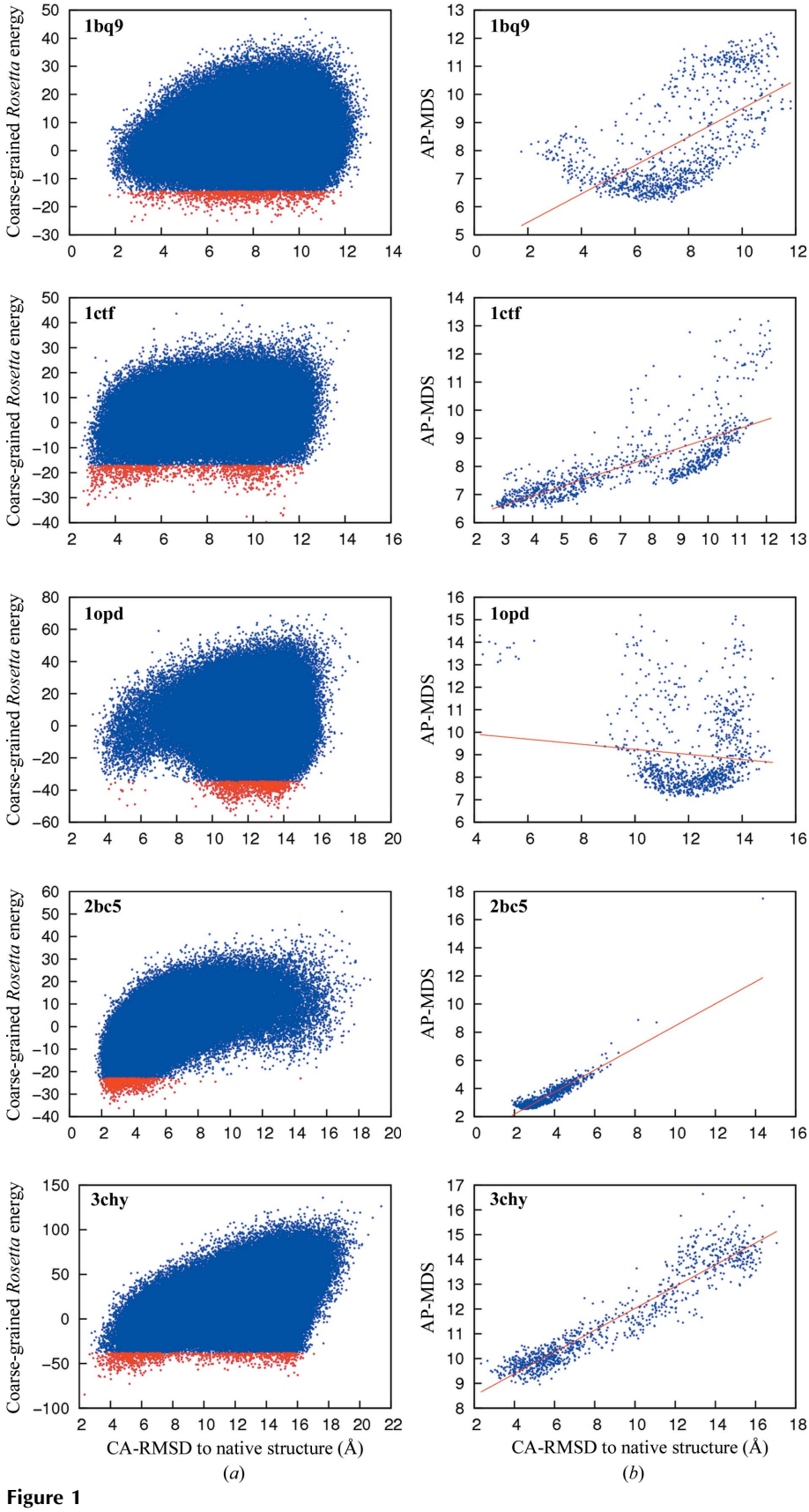

Figure 1
Coarse-grained energy landscape and estimation of prediction accuracy. (*a*) Correlation between *Rosetta* coarse-grained energy and the CA-RMSD from crystal structures. 1000 lowest energy models are represented using red points. (*b*) Correlation between the AP-MDS and prediction accuracy.

models generated are considered, since those models cover a wide range of distances from the native structure and there is a high degree of degeneracy in energy for less accurate models. This can be seen from the energy landscape in the form of a scatter plot of the coarse-grained energies for all models generated *versus* their CA-RMSDs from the native structure (Fig. 1*a*). Multiple local minima were observed in the coarse-grained energy landscape for all targets. The lowest energy models were not always the nearest to the native structure. For molecule 1opd the best predicted models were ∼4.0 Å in CA-RMSD from the native structure and the lowest energy models were very far away from the native structure. However, the distribution of these low-energy coarse-grained models may contain information about their accuracy. Therefore, 1000 protein models with the lowest coarse-grained energy were selected. The AP-MDS for each of these selected models was then calculated and their scatter plots *versus* CA-RMSDs are shown in Fig. 1(*b*). The AP-MDS of those lowest energy models for sequences 2bc5, 3chy, 1ctf and 1bq9 showed significantly good correlation with prediction accuracy. This AP-MDS seems to be a useful measure for the assessment and selection of *de novo* models.

Our goal is to identify residues that are predicted to have large errors. Therefore, the AP-RDS was calculated for each residue using the selected lowest energy coarse-grained models. The CA-RMSD of each residue from the native structure showed a good correlation with the AP-RDS (Fig. 2*a*). The correlation between the AP-RDS and the model accuracy of residues was more than 0.5. Instead of determining an absolute threshold to separate correctly and incorrectly predicted residues, all residues were subjected to rebuilding with the sampling frequency proportional to the estimated residue error based on the AP-RDS. As can be seen in Fig. 2(*a*), the AP-RDS was low for more accurately predicted residues and high for less accurately predicted residues. Therefore, the AP-RDS of each residue can provide an estimate of

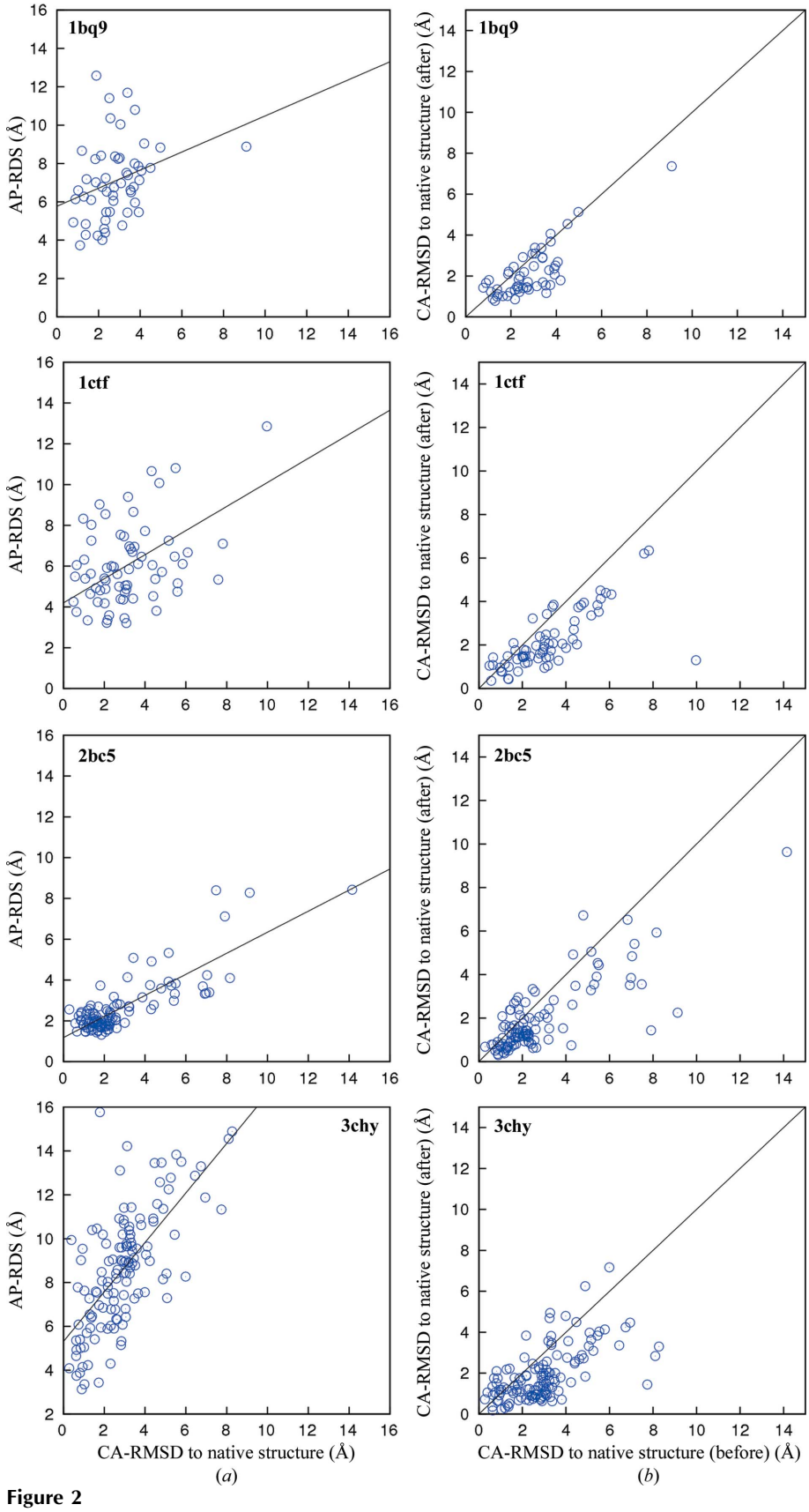

Figure 2
The correlation between the AP-RDS and prediction accuracy for each residue and the model improvement after rebuilding. (*a*) The correlation between the AP-RDS and the CA-RMSD of each residue in selected models. (*b*) The improvement in the CA-RMSD for each residue in coarse-grained models before and after rebuilding using *MORPHEUS*.

the accuracy of that residue in the predicted model.

3.2. Improvement after rebuilding

The aim of the rebuilding procedure is to optimize the coarse-grained models in order to improve the accuracy of all-atom models for successful phasing by MR. The improved coarse-grained models obtained using this rebuilding procedure are only intermediates, since all of the rebuilt models are subjected to all-atom energy optimization. In order to assess the potential improvement of coarse-grained models after rebuilding, the best rebuilt models after 300 independent runs and their corresponding input coarse-grained models were analyzed. A scatter plot of the CA-RMSDs for each model before and after rebuilding is shown in Fig. 3. Rebuilt models showed less geometric deviation from the native structure compared with the initial input models. The CA-RMSD of coarse-grained models was improved on average from 4.93 to 4.06 Å (Fig. 4*a*). The improvements were not limited to loops and termini; buried core regions were also improved significantly. For example, an α-helical region (residues 65–74) in the core of 3chy was improved by 2.5 Å after rebuilding (Figs. 5*a* and 6). An α-helical region (residues 38–44) in the core of 2bc5 was improved by 0.5 Å after rebuilding (Fig. 6). An α-helical region (residues 29–33) in 1be7 was improved by 0.8 Å after rebuilding (Figs. 5*c* and 6). However, improvement cannot be measured in the absence of the native structure.

The rebuilt models can only be productive when their CA-RMSD is below 3.0 Å from the native structure because these are the models that may become suitable templates for MR after *Rosetta* all-atom optimization. Therefore, the quality improvement for those models with a CA-RMSD of <3.0 Å after rebuilding was inspected. The accuracy of these coarse-grained models that are potential candidates for MR was improved from a CA-RMSD of 3.38 Å to 2.60 Å on average (Fig. 4*b*).

The improvement in CA-RMSD for each residue after rebuilding is shown in Fig. 2(*b*) for four targets. One of the best

rebuilt models was selected for each target. A large portion of the residues in these coarse-grained models were accurately rebuilt (Fig. 2*b*). Furthermore, improvement was also observed throughout the entire structure. The accuracy of the N- and C-terminal residues was significantly improved as these regions were sampled more frequently owing to their high AP-RDS scores (Fig. 5). It was also observed that some residues with higher AP-RDS were harder to optimize.

3.3. *Ab initio* phasing with *de novo* models

Of the ten data sets tested with *de novo* models that were unsuccessful in our previous work, seven were successfully phased using the current method. These were difficult targets in the previous experiment since estimation of the input model accuracy necessary for MR showed that more accurate models were required (Shrestha *et al.*, 2011). Using the current method, the accuracy of the input models has been improved for these data sets to enable successful phasing by MR. The CA-RMSD of the rebuilt models to the native structures, the phasing statistics and other relevant information are listed in Table 1. There were seven targets for which models were generated with less than 2.0 Å CA-RMSD from the native. These models produced high TFZ scores after MR. To further evaluate these *Phaser* solutions, an MR validation tool was executed on the models with TFZ values of greater than 5.8, although this procedure cannot be used in the absence of the crystal structure (Shrestha *et al.*, 2011). The CA-RMSD was calculated using rigid-body transformation with an optimal translation vector and a rotation matrix that minimized the sum of the squared distances between two sets of coordinates (Kabsch, 1976). It was also computed by applying crystallographic symmetry operators with all permissible origins of the space group. For *de novo* models to be considered as successful in MR, the difference between the CA-RMSDs computed by these two methods has to be small, *i.e.* ∼1.0 Å. It can be seen in column 8 of Table 1 that those models with small differences (∼1.0 Å) between the CA-RMSD (the first number in column 8) and the SYM-RMSD (the third number in column 8) were validated as successful in MR. This was further confirmed by the low *R* factor and $R_{\rm free}$ generated by *PHENIX AutoBuild* using these MR models (column 9 in Table 1).

The rebuilding procedure not only improved the overall quality of models for targets that were unsuccessful for MR in our previous work but also improved the quality of the models that were successful for MR. When tested on two of the previously successful targets, the CA-RMSD and all-atom r.m.s.d. for 1ig5 were improved from 2.36 to 1.59 Å and from 3.13 to 2.49 Å, respectively. Similarly, the CA-RMSD and all-atom r.m.s.d. for 256b were improved from 2.60 to 1.16 Å and from 2.90 to 1.82 Å, respectively (Table 1).

The successful MR models for these seven targets were further refined using automated model building in order to assess the quality of these *de novo* models and to further validate the MR solutions. The model building and refinement was carried out using the *AutoBuild* protocol implemented in *PHENIX* v.1.3 (Adams *et al.*, 2002) with default parameters. The electron-density maps constructed using phases from the *de novo* models successfully led to complete three-dimensional protein structures for the seven targets with good *R*-factor and $R_{\rm free}$ values. These models that were successful in MR were significantly improved after automated refinement (Fig. 6).

There were three diffraction data sets for which MR solutions could not be found. This could be because the energy landscape is far from ideal. It could also arise from insufficient sampling. In the case of 1opd, the energy landscape has a minimum around 12 Å, resulting in a weak anticorrelation between AP-MDS and CA-RMSD among the selected

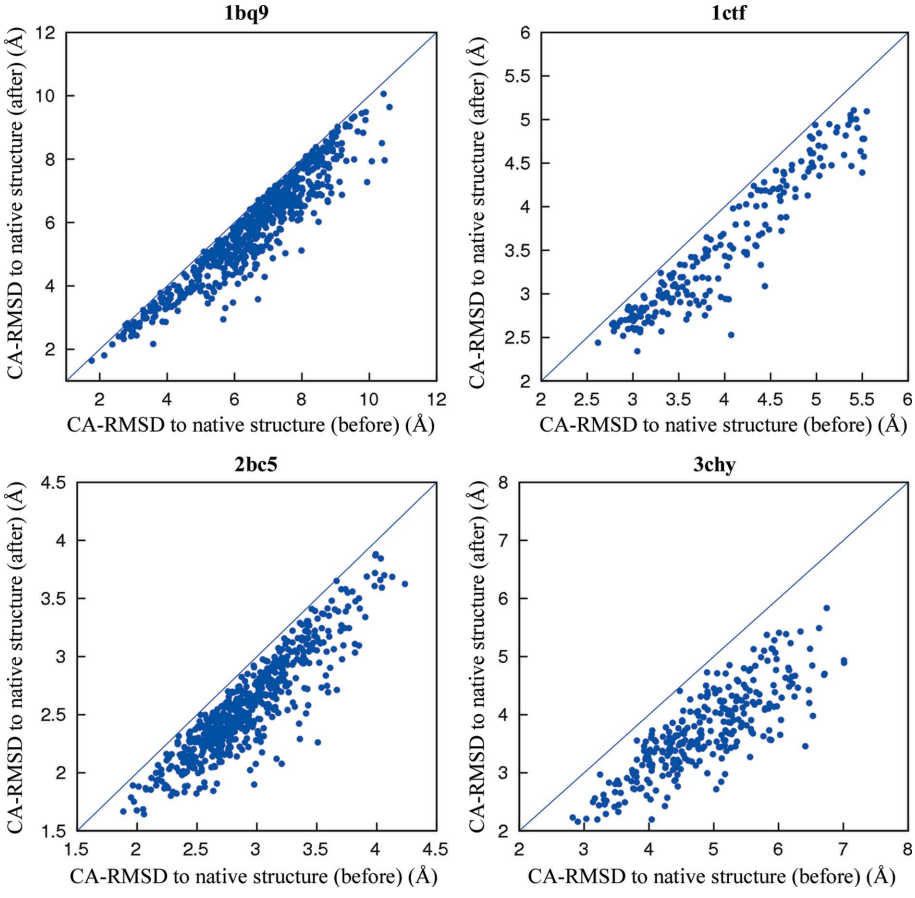

Figure 3
Improvement of the coarse-grained models after rebuilding. A scatter plot of the CA-RMSDs before and after rebuilding for all coarse-grained models is shown. A diagonal line is drawn on each panel. Points below this line represent models with improved CA-RMSD.

low-energy models (Fig. 1). Even for the best rebuilt model, the AP-RDS is distributed over a wide range and there was little improvement (Supplementary Fig. S2). In our study, MR for the ribosomal protein L7/L12 (PDB entry 1ctf) was not successful using the rebuilt *de novo* models, although the average CA-RMSD improvement of potential models for MR after rebuilding was 0.45 Å. *Rosetta* all-atom energy minimization did not yield models with sufficient accuracy to solve the phase problem. The best rebuilt models for histidine-containing proteins from *Escherichia coli* (PDB entries 1opd and 1cm3) were also far away from the native structure, with the largest improvement being from 12 to 8 Å (Supplementary Fig. S3); they were almost impossible to use for phasing in our study.

3.4. Accuracy and computation time

Our method was compared with other similar methods in order to evaluate its success and failure as well as the computation time required. The performance of our method on the ten data sets was compared with those of *Rosetta* 100 CPU day, *Rosetta* large-scale CPU time (Das & Baker, 2009) and *RosettaX* (Shrestha *et al.*, 2011). The results of the comparison are shown in Table 2 and Fig. 7. *MORPHEUS* achieved seven successes out of ten tested cases, which were all failed cases for *RosettaX*. Rosetta 100 CPU-day successfully generated the models required for phasing for two targets. The number of successful cases was increased from two to three using the increased computing power in *Rosetta* large-scale CPU time.

Although *MORPHEUS* showed better results than the previously published methods (Das & Baker, 2009; Shrestha *et al.*, 2011), this comparison might be unfair because of the periodic improvements in *Rosetta* methods and differences in fragment libraries arising from the increased pool of new structures deposited in the PDB. In order to measure the impact of our method on model rebuilding more objectively, it was compared with *Rosetta* v.3.2 using the same fragment library. 300 000 all-atom *de novo* models were generated using *Rosetta* v.3.2 for all targets. 10 000 lowest energy all-atom models were chosen for MR experiments in order to include all near-native models. All of the best models generated by *Rosetta* v.3.2 were included for MR, as can be seen in Table 3.

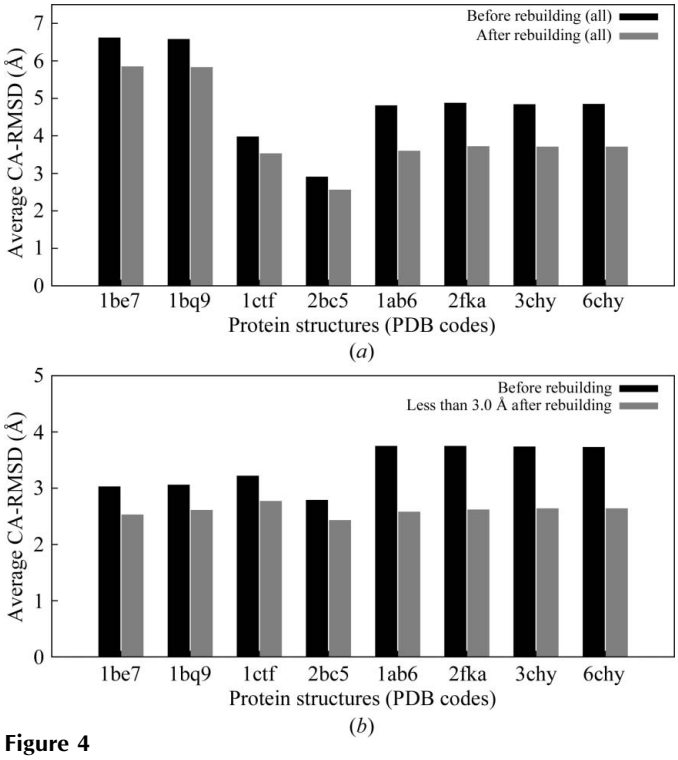

Figure 4
Average improvement in coarse-grained models after rebuilding. The black bar indicates the average CA-RMSD before rebuilding and the grey bar shows the average CA-RMSD after rebuilding. (*a*) shows all models and (*b*) shows those models with a CA-RMSD of less than 3.0 Å from the native structure after rebuilding.

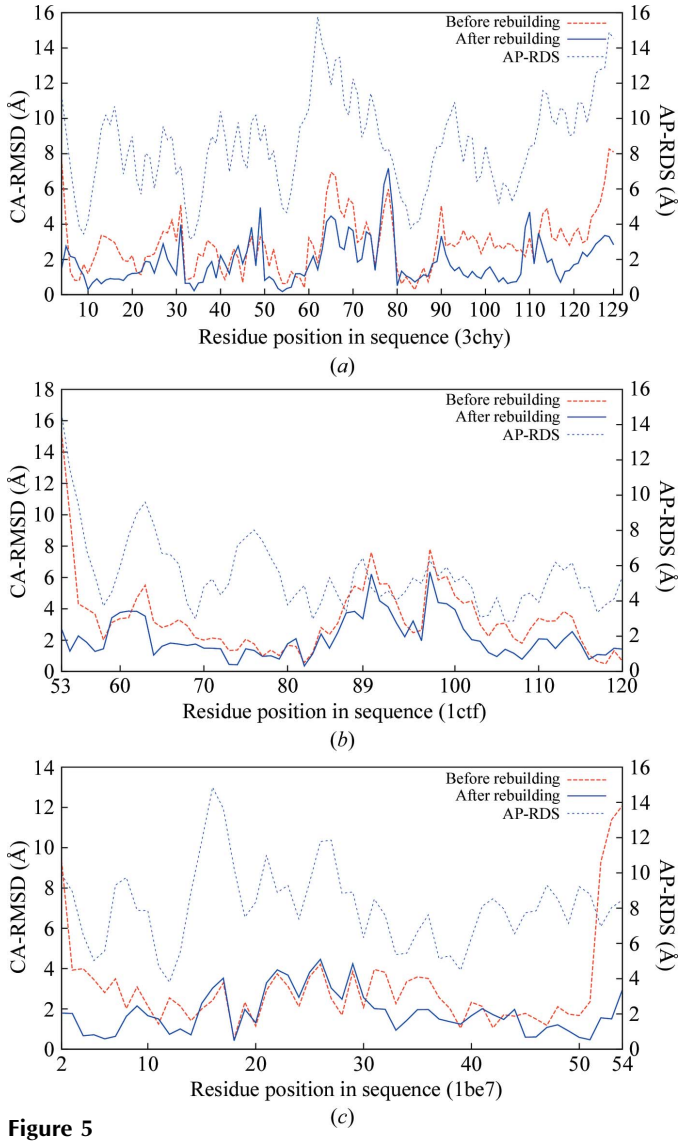

Figure 5
The improvement of coarse-grained models after rebuilding as a function of residue position and their corresponding AP-RDS scores. The CA-RMSD distributions of coarse-grained models before and after the rebuilding process at each residue position are shown for 3chy (*a*), 1ctf (*b*) and 1be7 (*c*). The AP-RDS scores are also shown for each residue for these three proteins.

Table 1
Summary of molecular-replacement and automated model-refinement results with rebuilt models.

Structure factors	Sequence	Space group	No. of copies in asymmetric unit	Sequence length	Resolution (Å)	RFZ, TFZ†	CA-RMSD, r.m.s.d., SYM-RMSD‡ (Å)	R, $R_{\rm free}$
1be7	1bq9	*H*3	1	53	1.67	4.4, 6.5	1.33, 1.63, 2.48	0.18, 0.20
1bq9	1bq9	$P2_12_12_1$	1	53	1.20	3.8, 6.9	1.61, 2.28, 1.77	0.22, 0.22
1ctf	1ctf	$P4_32_12$	1	68	1.70	3.0, 6.1	2.67, 3.20, 27.82	—
1opd	1opd	*P*1	1	85	1.50	4.9, 100.0	9.52, 10.03, 15.93	—
1cm3	1opd	$P2_1$	1	85	1.60	4.6, 3.9	9.46, 10.05, 16.94	—
2bc5	2bc5	$P2_12_12_1$	4	106	2.25	2.5, 9.3	1.26, 2.07, 1.47	0.26, 0.31
1ab6	3chy	$P3_1$	2	128	2.20	4.4, 7.9	1.72, 2.25, 1.81	0.20, 0.26
2fka	3chy	*F*432	1	128	2.00	3.6, 7.0	1.88, 2.60, 1.92	0.24, 0.25
3chy	3chy	$P2_12_12_1$	1	128	1.66	4.0, 8.3	1.80, 2.44, 1.93	0.18, 0.22
6chy	3chy	$P2_12_12$	2	128	2.33	4.0, 9.7	1.96, 2.45, 1.99	0.22, 0.28
1ig5§	1ig5	$P4_32_12$	1	75	1.50	3.7, 6.8	1.59, 2.49, 1.63 (2.36, 3.13, —)	0.21, 0.26
256b§	256b	*P*1	2	106	1.40	9.0, 8.6	1.16, 1.82, 1.18 (2.60, 2.90, —)	0.24, 0.26

† Results after *Phaser* run. RFZ and TFZ are the rotation-function *Z*-score and the translation-function *Z*-score, respectively. ‡ CA-RMSD, r.m.s.d. and SYM-RMSD are the root-mean-square deviations from the native structure calculated using C^{α} atoms, all atoms and all possible crystallographic symmetry operators and origins, respectively. A small difference (∼1 Å) between CA-RMSD and SYM-RMSD indicates successful MR. § Two targets that were successful in MR using models generated by *RosettaX* were subjected to rebuilding by *MORPHEUS*. The models have been improved by rebuilding and the corresponding quality measures from *RosettaX* are shown in parentheses.

Table 2
Comparison of MR experiments using models generated by different methods.

Successful cases are indicated by 1; 0 represents failed cases.

Structure factors	Sequence	Space group	No. of copies in asymmetric unit	Sequence length	Resolution (Å)	*Rosetta*, 100 CPU days†	*Rosetta*, large-scale†	*RosettaX*	*Rosetta* v.3.2	*MORPHEUS*
1be7	1bq9	*H*3	1	53	1.67	0	0	0	1	1
1bq9	1bq9	$P2_12_12_1$	1	53	1.20	0	0	0	1	1
1ctf	1ctf	$P4_32_12$	1	68	1.70	0	0	0	0	0
1opd	1opd	*P*1	1	85	1.50	0	0	0	0	0
1cm3	1opd	$P2_1$	1	85	1.60	0	0	0	0	0
2bc5	2bc5	$P2_12_12_1$	4	106	2.25	1	0	0	0	1
1ab6	3chy	$P3_1$	2	128	2.20	0	1	0	0	1
2fka	3chy	*F*432	1	128	2.00	1	1	0	0	1
3chy	3chy	$P2_12_12_1$	1	128	1.66	0	0	0	0	1
6chy	3chy	$P2_12_12$	2	128	2.33	0	1	0	0	1

† Das & Baker (2009).

Table 3
Comparison of the best models generated and the models that produced the highest MR scores.

In the *Rosetta* v.3.2 run, 10 000 lowest all-atom energy models were selected.

						Rosetta v.3.2				*MORPHEUS*		
Structure factors	Sequence	Space group	No. of copies in asymmetric unit	Sequence length	Resolution (Å)	RFZ, TFZ	CA-RMSD, r.m.s.d. (Å)	Rank by CA-RMSD†	Lowest CA-RMSD, r.m.s.d. (Å)	RFZ, TFZ	CA-RMSD, r.m.s.d. (Å)	Lowest CA-RMSD, r.m.s.d. (Å)
1be7	1bq9	*H*3	1	53	1.67	4.8, 7.1	1.18, 1.70	1	1.18, 1.70	4.4, 6.5	1.33, 1.63	1.29, 2.01
1bq9	1bq9	$P2_12_12_1$	1	53	1.20	4.4, 6.8	1.11, 1.41	1	1.11, 1.41	3.8, 6.9	1.61, 2.28	1.39, 2.09
1ctf	1ctf	$P4_32_12$	1	68	1.70	3.3, 3.7	2.46, 2.96	1	2.46, 2.96	3.0, 6.1	2.67, 3.20	2.31, 3.03
1opd	1opd	*P*1	1	85	1.50	3.7, 100	3.09, 3.99	3	2.89, 3.85	4.9, 100	9.52, 10.03	8.33, 9.18
1cm3	1opd	$P2_1$	1	85	1.60	3.8, 2.5	3.09, 3.97	3	2.96, 3.85	4.6, 3.9	9.46, 10.05	8.38, 9.13
2bc5	2bc5	$P2_12_12_1$	4	106	2.25	—	1.11, 1.78	4	1.09, 1.86	2.5, 9.3	1.26, 2.07	1.04, 1.68
1ab6	3chy	$P3_1$	2	128	2.20	—	2.30, 2.84	1	2.30, 2.84	4.4, 7.9	1.72, 2.25	1.72, 2.25
2fka	3chy	*F*432	1	128	2.00	3.9, 4.1	2.44, 3.12	1	2.44, 3.12	3.6, 7.0	1.88, 2.60	1.85, 2.61
3chy	3chy	$P2_12_12_1$	1	128	1.66	4.2, 4.7	2.37, 3.05	1	2.37, 3.05	4.0, 8.3	1.80, 2.44	1.68, 2.28
6chy	3chy	$P2_12_12$	2	128	2.33	—	2.37, 2.93	1	2.37, 2.93	4.0, 9.7	1.96, 2.45	1.78, 2.27

† The rank of the best models from the simulation that are in the 10 000 selected models. This information is not given for *MORPHEUS* because the models are used for MR as they are generated.

Rosetta v.3.2 produced better quality models for rubredoxin (PDB entries 1bq9 and 1be7) than *MORPHEUS*, as is shown in Table 3. These models easily succeeded in MR with higher confidence in both cases. However, the selection of 1000 lowest energy coarse-grained models did not include all of the best models for 1bq9 (Fig. 1). This led to *MORPHEUS* generating less accurate models than *Rosetta* v.3.2. Aside from these two data sets, the other eight data sets were unable to

achieve successful MR with models generated using *Rosetta* v.3.2. The best predicted *de novo* models could not pass the MR test for cytochrome *c*-b_{562} (PDB entry 2bc5) using both *RosettaX* and *Rosetta* v.3.2. For *MORPHEUS*, slightly less accurate models yielded an MR solution for this target. The reason for this might be that many highly accurate models were examined with the diffraction data set during the simulation. In addition, identical CA-RMSDs could arise from very different structures. The success in MR of a structure with a relatively large CA-RMSD from the native structure could conceivably arise from the errors in the residues being unevenly distributed. Most of the residues were probably more accurately predicted and some residues with large errors might have made the overall CA-RMSD relatively high. The improved prediction accuracy for 1opd and 1cm3 after *MORPHEUS* was not useful since the quality of the best input model was about 9.0 Å in CA-RMSD.

Although both *MORPHEUS* and *Rosetta* v.3.2 are feasible using currently available moderate computing resources, another factor for comparison could be the total elapsed time spent by these methods. The total elapsed time is the accumulation of the time spent in coarse-grained model generation, energy minimization using all-atom models and MR. In addition, the elapsed time for *MORPHEUS* also includes the model-rebuilding time. Both methods used the same computing resource. The average total elapsed time is about equal for the methods, except for two targets (PDB entries 2bc5 and 1be7). The large differences in elapsed time for these two proteins are primarily owing to MR. These are very likely to be the worst-case scenario for *MORPHEUS*. At the other extreme, *MORPHEUS* can obtain an MR solution with fewer models generated compared with *Rosetta* v.3.2. Of the tested cases, the total number of models generated for phasing varied from 1.0×10^4 to 1.8×10^5. *MORPHEUS* also used less elapsed time when an MR solution was found very early in the simulation, such as for the targets 3chy and 6chy. In some cases, the energy-based selection used in *Rosetta* v.3.2 might have missed the best models for *ab initio* phasing although suitable models had already been predicted; *MORPHEUS* does not suffer from this drawback since all generated *de novo* models are used for MR until successful solutions are found.

4. Discussion

4.1. Coarse-grained energy landscape and AP-RDS

The coarse-grained energy function is designed to enable the sampling of a larger conformational space for simplified protein models that contain only the main chain and the centroids of side chains. The objective is to search and find

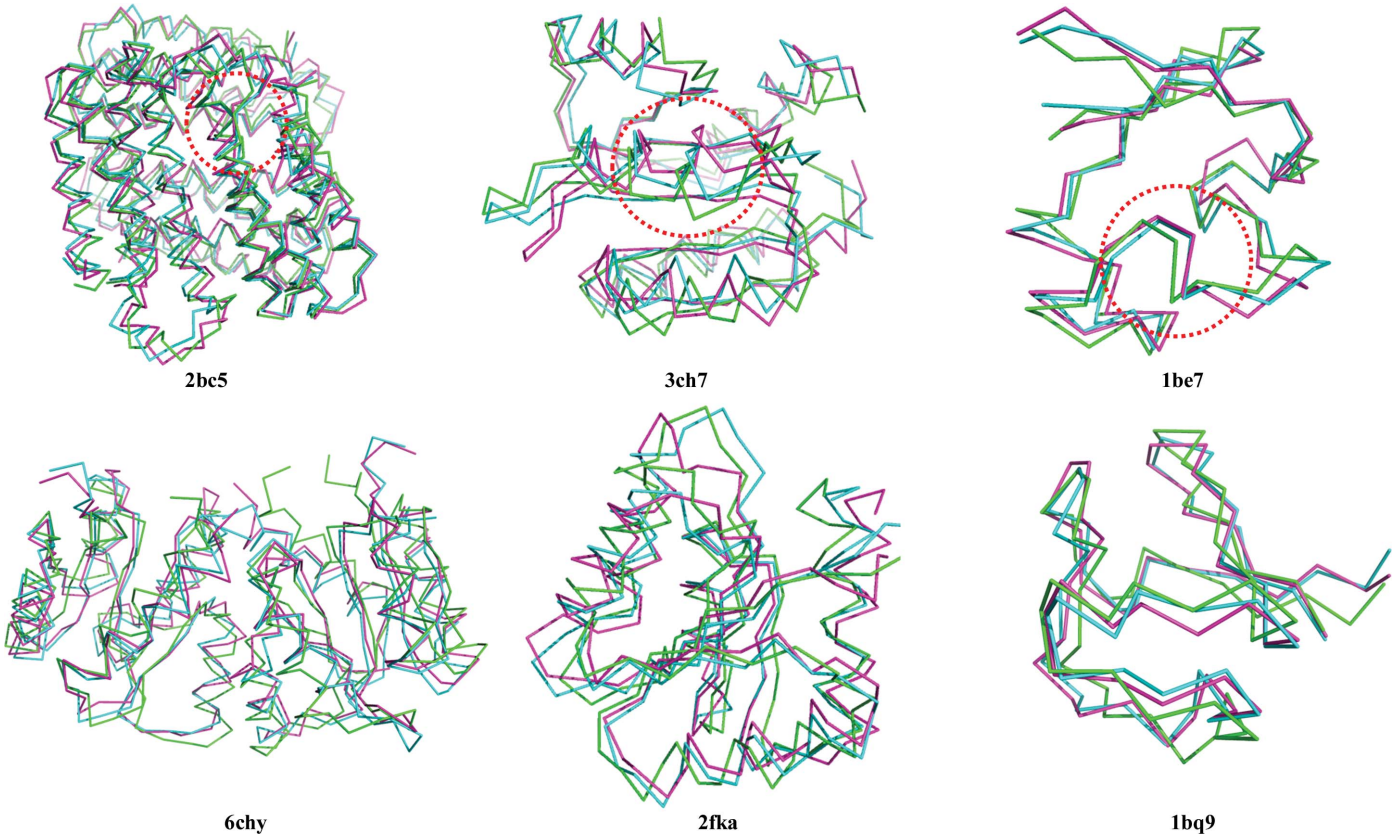

Figure 6
Comparison of models before and after rebuilding with models after automated model building and refinement. Final models were built by the *PHENIX AutoBuild* protocol using successful MR models. Models before rebuilding are coloured green, models after rebuilding used for MR are coloured cyan and models after *PHENIX Autobuild* are coloured magenta. Core regions with significant improvement as mentioned in the text are highlighted by circles with red dashed lines.

the global fold of a target protein by maximizing the burial of hydrophobic side chains and the exposure of hydrophilic side chains. However, missing side-chain atoms in the coarse-grained models make the coarse-grained energy function less accurate. The coarse-grained models generated can be from conformations trapped in multiple minima in a complex energy landscape. It is difficult to discriminate between predicted models using only the coarse-grained energy function. Despite being less accurate than its all-atom counterpart, the coarse-grained energy can be used to generate near-native models (Das & Baker, 2008). It is generally assumed that in a randomly sampled energy landscape there should be more models generated that correspond to lower energy than models whose conformations correspond to higher energy. This principle has provided the foundation for the use of clustering methods to identify native-like protein models (Shortle *et al.*, 1998; Zhang & Skolnick, 2004*a*; Berenger *et al.*, 2011). Our choice of the geometric similarity of low-energy models is similar in principle to the clustering methods that identify native-like models. The AP-MDS will be small if a model has more neighbours because they sample a lower energy level in the energy landscape. The same principle can be used to reason that the AP-RDS will be small if a residue has more neighbours in the generated models because it corresponds to a lower energy conformation. This seems to be the case for our data, as shown in Fig. 1. This concept has also been employed to generate hybrid models with the best residues from selected templates in homology modelling (Wallner & Elofsson, 2006). However, the correlation between the AP-MDS or AP-RDS and model quality depends on the coarse-grained energy landscape as it pertains to the protein target. Although this holds true for most of the targets tested, this correlation breaks down for 1opd (Fig. 1*b*) because the sampled energy landscape has its minimum at about 12–14 Å CA-RMSD from the native structure (Fig. 1*a*). Our model-rebuilding procedure would fail in this case. As the CA-RMSD is calculated by comparing the corresponding atoms between the model and the native structure, it may appear to be very large for the purpose of assessing the suitability of a model for MR since it is the spatial matching of the scatterers rather than the order with which the atoms are connected that is important for MR. The CA-RMSD was used here to measure the quality of a predicted model owing to the critical dependence of the method that was used to generate the model on the connection order of all of the atoms in a protein.

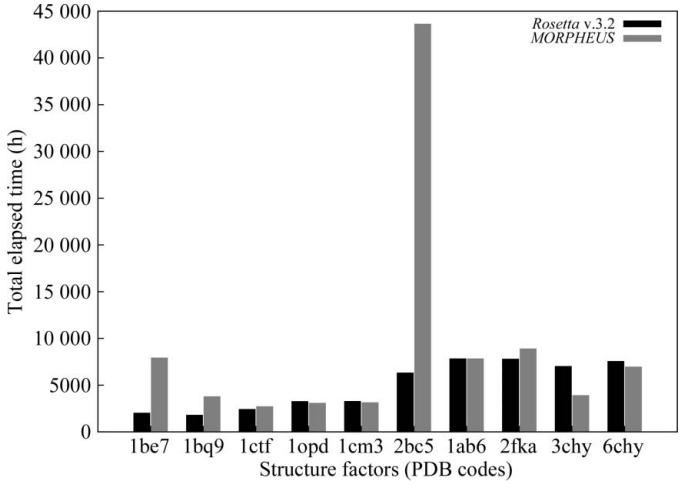

Figure 7
Elapsed computation time spent by *MORPHEUS* and *Rosetta* v.3.2 during simulation.

4.2. Conformational sampling with AP-RDS

The rebuilding procedure concentrates on the effective sampling of potentially incorrect local segments in a coarse-grained model guided by the AP-RDS. The purpose of this sampling strategy is to increase the sampling rate of residues with larger error. No specific threshold is used to discriminate correctly and incorrectly predicted residues. The conformation sampling is carried out at a relative rate proportional to the AP-RDS.

Wrongly predicted segments are non-uniformly distributed in the model. Terminal and loop segments are regions that potentially have the largest structural diversity. The sampling algorithm for loop segments was carried out by fixing two anchor points in a protein without changing the entire conformation (Canutescu & Dunbrack, 2003). The conformational search using the AP-RDS score samples the entire protein structure non-uniformly regardless of the secondary structure. Our results suggest that non-uniform sampling based on the AP-RDS may be an alternative strategy to random sampling or sampling of loop regions only.

Model improvements were observed over the entire region of a protein. The input models were significantly improved at the C- and N-terminal segments, as residues in these segments showed a higher AP-RDS. Regardless of secondary structure, improvements of local segments on the exterior and in the interior of the models were also observed. Despite showing a higher AP-RDS, some local segments in the protein structure were difficult to improve. Possible reasons are that the conformations represented by the short three-residue fragments might not be adequate to sample near-native regions for these residues or that the coarse-grained energy might not be accurate enough to guide the sampling.

4.3. Model rebuilding and molecular replacement

The accuracy of *de novo* models is a critical factor in successful phasing by MR. In practice, loop regions and N- and C-terminal segments are considered to be flexible and harder to predict and these segments are trimmed off in template models prior to the MR experiment. Instead of trimming off these segments in the model, rebuilding these regions can be an alternative strategy for successful MR.

Rebuilding residues with large errors can increase the accuracy of *de novo* models. It is advantageous to rebuild these residues at the coarse-grained model level. The coarse-grained sampling method explores a large conformational space more effectively. In principle, rebuilding can also be performed on all-atom models. However, there is a much

larger conformational space to be sampled and the current search protocol is designed to avoid drastic changes to the global conformation at the all-atom optimization stage.

The existence of higher order noncrystallographic symmetry in the crystal sometimes tends to require more accurate input models for success in MR. This is probably because the successful location of each monomer in the asymmetric unit depends on the solution for the previous monomer and the errors tend to accumulate. For example, cytochrome c-b_{562} (PDB entry 2bc5) is an α-helical bundle and *de novo* modelling generated highly accurate models, but phasing with these models was not successful mainly owing to the presence of four copies of the molecule in the asymmetric unit. The presence of larger loop segments in rubredoxin (PDB entry 1be7) may also increase the difficulty in phasing. The success of *de novo* models for phasing depends not only on the accuracy of backbone atoms but also that of side-chain atoms. The improvement in the main chain alone is insufficient when all-atom optimization cannot lead to the increased model accuracy required for phasing, as in the ribosomal protein L7/L12 (PDB entry 1ctf). The CA-RMSD of coarse-grained models was reduced, but the subsequent all-atom models did not appear to be sufficiently accurate for phasing. Therefore, improvement in all-atom modelling is also important in addition to rebuilding coarse-grained models.

Model quality has been shown to be an important determinant for the success of MR. However, it has paradoxically been observed that two models with very similar r.m.s.d.s to the native structure could have opposite outcomes in MR, as in the case of 2bc5 in Table 3. This might be owing to the use of the r.m.s.d. as a single measure of the structural differences between two models. The r.m.s.d. is degenerate and cannot distinguish various scenarios of structural differences between two models. The quadratic nature of the r.m.s.d. gives a higher weight to the region that differs the most, whereas for MR the matched regions between the template and target give rise to signal while the mismatched regions generate noise. It is conceivable that alternative measures such as GDT (Zemla *et al.*, 1999; Zemla, 2003), MaxSub (Siew *et al.*, 2000), TM-score (Zhang & Skolnick, 2004*b*), Q-score (Ben-David *et al.*, 2009) or percentile-based spread (Pozharski, 2010) might be used as a better predictor of success in MR for a given model.

The conformational variation in an ensemble of predicted models has previously been exploited not only for model-quality assessment but also for model rebuilding. Although conformational variation has been employed as a 'colony energy', which is used as a post-filtering measure for loop prediction (Xiang *et al.*, 2002), it is used here as both a global (AP-MDS) and a local (AP-RDS) measure not only for the estimation of errors but also to guide the sampling and rebuilding of the entire region in a model. The local structural variation of models has been used to identify regions that are most likely to be in error and to aggressively sample and refine them to improve model quality (Qian *et al.*, 2007). There are several differences between our method and that of Qian and coworkers. Firstly, *MORPHEUS* uses coarse-grained models to estimate errors and rebuilds coarse-grained models before subjecting them to all-atom refinement, whereas Qian and coworkers use all-atom models to identify error-prone regions and the rebuilding also uses all-atom models. Secondly, *MORPHEUS* uses local variation to estimate errors and then uses this variation to guide the sampling proportional to the estimated errors. There is no threshold needed to identify a particular region for rebuilding and the sampling is non-uniform. In contrast, Qian and coworkers use local variation to identify regions that are most likely to contain errors and then aggressively sample these regions uniformly regardless of the actual amount of variation within and among regions. Thirdly, *MORPHEUS* rebuilds the entire model with sampling proportional to the structural variation. There is no chain break created during the rebuilding process. A modified acceptance criterion proportional to the structural variation in the form of a temperature factor is introduced to enable large conformational changes that cause an increase in energy to be accepted during the Monte Carlo sampling. Qian and coworkers fix the C- and N-terminal ends adjacent to the region to be rebuilt and then use the cyclic coordinate descent to close the chain break.

5. Conclusions

This study explores the rebuilding of error-prone residues in a coarse-grained model in order to generate more accurate all-atom models that could be used as templates for phasing by MR. To evaluate our method, a set of targets that were unsuccessful in our previous study were tested. The CA-RMSD of potential coarse-grained models for MR (less than 3.0 Å CA-RMSD from the native structure) was reduced from 3.4 Å to 2.6 Å on average. The large errors present in the N- and C-terminal segments were more significantly reduced. Since terminal segments are very difficult to predict accurately in *de novo* modelling, the rebuilding methodology may be a method for improving the accuracy of terminal segments. Moreover, the local errors in the protein models were reduced regardless of secondary structure. Model improvements are found not only in the termini but also in the core regions.

Ab initio phasing with rebuilt coarse-grained models after all-atom optimization increased the success rate of molecular replacement. 70% of the tested cases succeeded in molecular replacement, primarily owing to the improved model quality. Moreover, the phases obtained after successful molecular replacement were sufficient to generate high-quality electron-density maps for automated model building and refinement.

We wish to thank the Advanced Center for Computing and Communication, RIKEN, Japan for the computing resources of the RIKEN Integrated Cluster of Clusters (RICC) system. We are grateful to Dr David Baker and Dr Randy Read for making the *Rosetta* and *Phaser* source codes available. We thank members of our laboratory for stimulating discussions. We acknowledge the Initiative Research Unit program of RIKEN, Japan for funding.

References

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

Ben-David, M., Noivirt-Brik, O., Paz, A., Prilusky, J., Sussman, J. L. & Levy, Y. (2009). *Proteins*, **77**, Suppl. 9, 50–65.

Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G. & Yusupov, M. (2011). *Science*, **334**, 1524–1529.

Berenger, F., Zhou, Y., Shrestha, R. & Zhang, K. Y. J. (2011). *Bioinformatics*, **27**, 939–945.

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

Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.

Bowie, J. U. & Eisenberg, D. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 4436–4440.

Canutescu, A. A. & Dunbrack, R. L. (2003). *Protein Sci.* **12**, 963–972.

Chivian, D., Kim, D. E., Malmstrom, L., Bradley, P., Robertson, T., Murphy, P., Strauss, C. E., Bonneau, R., Rohl, C. A. & Baker, D. (2003). *Proteins*, **53**, Suppl. 6, 524–533.

Das, R. & Baker, D. (2008). *Annu. Rev. Biochem.* **77**, 363–382.

Das, R. & Baker, D. (2009). *Acta Cryst.* D**65**, 169–175.

DiMaio, F., Terwilliger, T. C., Read, R. J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H. L., Das, D., Vorobiev, S. M., Iwaï, H., Pokkuluri, P. R. & Baker, D. (2011). *Nature (London)*, **473**, 540–543.

Kabsch, W. (1976). *Acta Cryst.* A**32**, 922–923.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* D**64**, 119–124.

Kim, D. E., Blum, B., Bradley, P. & Baker, D. (2009). *J. Mol. Biol.* **393**, 249–260.

Kryshtafovych, A. & Fidelis, K. (2009). *Drug Discov. Today*, **14**, 386–393.

Levitt, M. & Gerstein, M. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 5913–5920.

Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.

Mandell, D. J., Coutsias, E. A. & Kortemme, T. (2009). *Nature Methods*, **6**, 551–552.

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

Pozharski, E. (2010). *Acta Cryst.* D**66**, 970–978.

Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007). *Nature (London)*, **450**, 259–264.

Rigden, D. J., Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* D**64**, 1288–1291.

Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. (2004). *Methods Enzymol.* **383**, 66–93.

Shortle, D., Simons, K. T. & Baker, D. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 11158–11162.

Shrestha, R., Berenger, F. & Zhang, K. Y. J. (2011). *Acta Cryst.* D**67**, 804–812.

Siew, N., Elofsson, A., Rychlewski, L. & Fischer, D. (2000). *Bioinformatics*, **16**, 776–785.

Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.

Tyka, M. D., Keedy, D. A., André, I., Dimaio, F., Song, Y., Richardson, D. C., Richardson, J. S. & Baker, D. (2011). *J. Mol. Biol.* **405**, 607–618.

Wallner, B. & Elofsson, A. (2006). *Protein Sci.* **15**, 900–913.

Xiang, Z., Soto, C. S. & Honig, B. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 7432–7437.

Zemla, A. (2003). *Nucleic Acids Res.* **31**, 3370–3374.

Zemla, A., Venclovas, C., Moult, J. & Fidelis, K. (1999). *Proteins*, **37**, Suppl. 3, 22–29.

Zhang, Y. & Skolnick, J. (2004*a*). *J. Comput. Chem.* **25**, 865–871.

Zhang, Y. & Skolnick, J. (2004*b*). *Proteins*, **57**, 702–710.