STRUCTURAL BIOLOGY
ISSN 2059-7983

MR-REX: molecular replacement by cooperative conformational search and occupancy optimization on low-accuracy protein models

Jouko J. Virtanen and Yang Zhang*

Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA.
*Correspondence e-mail: zhng@umich.edu

Received 17 March 2017
Accepted 10 April 2018

Edited by R. J. Read, University of Cambridge, England

Keywords: phase problem; molecular replacement; replica-exchange Monte Carlo simulation; occupancy optimization; *MR-REX*.

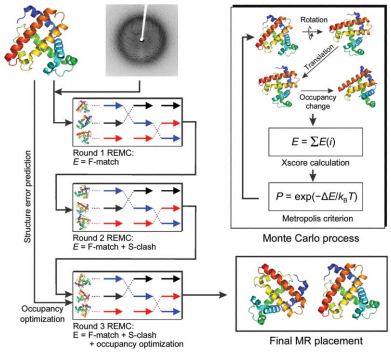
Supporting information: this article has supporting information at journals.iucr.org/d

Molecular replacement (MR) has commonly been employed to derive the phase information in protein crystal X-ray diffraction, but its success rate decreases rapidly when the search model is dissimilar to the target. *MR-REX* has been developed to perform an MR search by replica-exchange Monte Carlo simulations, which enables cooperative rotation and translation searches and simultaneous clash and occupancy optimization. *MR-REX* was tested on a set of 1303 protein structures of different accuracies and successfully placed 699 structures at positions that have an r.m.s.d. of below 2 Å to the target position, which is 10% higher than was obtained by *Phaser*. However, cases studies show that many of the models for which *Phaser* failed and *MR-REX* succeeded can be solved by *Phaser* by pruning them and using nondefault parameters. The factors effecting success and the parts of the methodology which lead to success are studied. The results demonstrate a new avenue for molecular replacement which outperforms (and has results that are complementary to) the state-of-the-art MR methods, in particular for distantly homologous proteins.

1. Introduction

Molecular replacement (MR) is a technique that is employed to determine the phase information in X-ray diffraction by replacing the target protein with a protein that has a known homologous structure (Rossmann & Blow, 1962; Crowther, 1972; Drenth, 2007). Advanced MR methods can now use predicted protein structures (Wang *et al.*, 2016), *ab initio* models (Bibby *et al.*, 2012), small secondary-structure elements (Rodríguez *et al.*, 2009) and even single atoms (McCoy *et al.*, 2017). Nearly two thirds of the X-ray structures deposited in the PDB have been solved by MR (Long *et al.*, 2008). Nevertheless, the failure rate of automated MR techniques is high when there are no closely homologous protein structures (typically with a sequence identity of <30%; Schwarzenbacher *et al.*, 2004). Even with closely homologous proteins, the correct positioning of the search model in MR becomes nontrivial in cases where only low-resolution data are available (Baker *et al.*, 1995; Giorgetti *et al.*, 2005), the unit cell is densely packed (Chang & Lewis, 1997; Glykos & Kokkinidis, 2000, 2001), the protein is elongated (Chang & Lewis, 1997) or the space group is of high symmetry (Baker *et al.*, 1995; Tong, 1996).

In traditional MR approaches, such as those used by *Phaser* (McCoy, 2007; McCoy *et al.*, 2007), *MOLREP* (Vagin & Teplyakov, 1997, 2000, 2010), *AMoRe* (Navaza, 1987, 1990, 1993, 1994, 2001; Castellano *et al.*, 1992), *CNS* (Grosse-Kunstleve & Adams, 2001) and *COMO* (Jogl *et al.*, 2001), the placement of the probe model consists of two general steps. Firstly, the model is oriented (by rotation) using a systematic



grid-based search. Secondly, the model is positioned (by translation) for a given orientation using another systematic grid-based search. When there are multiple proteins in the asymmetric unit, the proteins are typically placed one by one. While this strategy works well for easy cases, in which the structure of the homologous model is similar to that of the target, it may be difficult to identify candidates for the correct orientation of the model in harder cases without first approximating the position of the model, because methods that split the rotational and translational searches need to make approximations that break down at some point (Brünger, 1990, 1993, 1997). For example, Patterson function-based approaches make the approximation that the intra-molecular and intermolecular Patterson functions are separable, which can fail in cases where this assumption is invalid, such as when the unit cell is densely packed or the protein is elongated (Evans & McCoy, 2008). Maximum-likelihood-based methods (McCoy, 2007) can also fail in such cases because the likelihood of a model given only the orientation cannot be determined accurately when the model is not near the native. As a structure moves further from the native it becomes more difficult to choose the correct orientation for the translational search. In hard cases, the correct orientation of the model may not be a local minimum of the function that is used to assess the agreement between the calculated diffraction pattern and the experimental data (Jogl *et al.*, 2001; Kissinger *et al.*, 1999; Tong, 1996). While it is possible to perform a six-dimensional grid search such as that performed by *SOMoRe* (Jamrog *et al.*, 2003) or *MPI_FSEARCH* (Liu *et al.*, 2003), it becomes difficult to perform a 6*n*-dimensional grid search for *n* protein components in the asymmetric unit when *n* is greater than 1.

Here, we present *MR-REX* (Molecular Replacement by Replica-Exchange Simulation), which uses replica-exchange Monte Carlo (REMC) simulations (Swendsen & Wang, 1986) to integrate the different conformational search components into a unified process to improve the efficiency of MR. REMC is an advanced Monte Carlo simulation method designed to improve the speed of the canonical Metropolis Monte Carlo simulation approach (Metropolis *et al.*, 1953); the latter tends to become trapped in local energy minima when the energy landscape of the system is rugged. In REMC, multiple replicas of simulations are performed in parallel and at different temperatures; the high-temperature replicas may help the low-temperature replicas to jump across the energy barriers by periodically swapping the temperatures of different replicas following the Metropolis criterion. The *MR-REX* protocol enables the consideration of a clash score during the MR

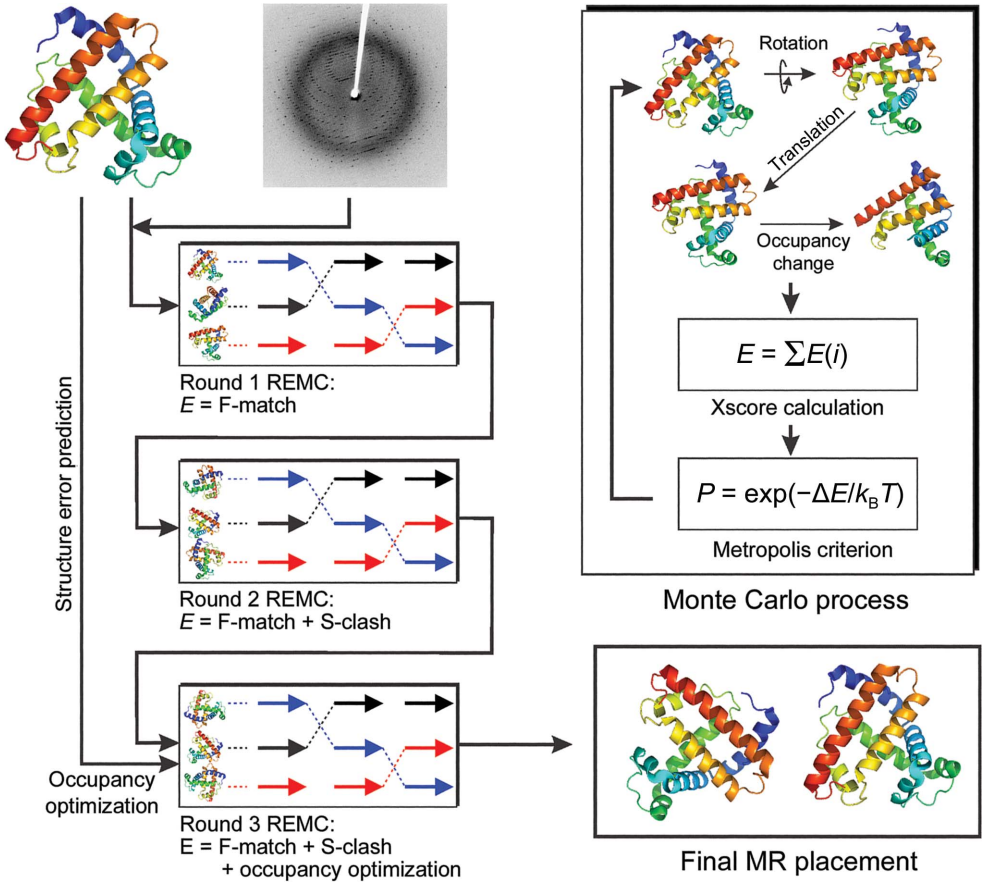


Figure 1
Flowchart for *MR-REX*. The pipeline consists of three consecutive replica-exchange Monte Carlo simulations designed for diffraction data match, clash removal and occupancy optimization, respectively. The inset at the upper left highlights the common Monte Carlo process, with movements containing rotation, translation and occupancy changes, followed by diffraction calculation, *B*-factor correction and Metropolis movement acceptance.

search to increase the success rate. Since inaccuracies in the protein model are more detrimental to the success of MR than deletions (Sammito *et al.*, 2014; Bibby *et al.*, 2012; Wang *et al.*, 2016, 2017; Shrestha & Zhang, 2015), *MR-REX* takes advantage of this fact by predicting possible inaccurate segments of the structural model and optimizing the occupancies of these segments during MR. A similar approach of pruning inaccurate atoms has also been applied by other programs, including *CHAINSAW* (Stein, 2008) and *phenix.sculptor* (Bunkóczi & Read, 2011), to improve the efficiency of MR. The *MR-REX* method is carefully examined using large-scale benchmark data sets with structure models with different distances from the native. It has been demonstrated that the method is able to find correct MR solutions for some difficult cases where the other state-of-the-art phasing method could not. The source code and online server for *MR-REX* have been made freely available at https://zhanglab.ccmb.med.umich.edu/MR-REX/.

2. Methods

MR-REX generates molecular-replacement models through three steps of consecutive REMC simulations for optimized diffraction data match, steric clash removal and the follow-up occupancy optimization, respectively. A flowchart of the pipeline is depicted in Fig. 1, with the detailed procedure described below. The algorithm is benchmarked mainly with *Phaser* (McCoy *et al.*, 2007), one of the most accurate programs in molecular replacement; *PHENIX* 1.11 was used to run *Phaser* and *phenix.autobuild* with default settings.

2.1. Input and output data

The input to *MR-REX* is a set consisting of structure-factor data, a structure file in PDB format and a parameter file. The parameter file needs to contain the space group, unit-cell information and the number of models in the asymmetric unit. There are a number of other options in the parameter file, which are described in Supplementary Table S1. Most of these options have reasonable defaults, which do not have to be modified by the user. The output contains a set of candidate MR solutions and a log file specifying the scores of the candidate MR solutions.

2.2. Scoring function and weight optimization

The scoring function that guides the *MR-REX* search is a linear combination of four terms that quantify how well the calculated ($F_{\rm calc}$) and experimental ($F_{\rm obs}$) diffraction data match and three terms that evaluate the intermolecular clashes:

$$\begin{aligned}\mathrm{XScore} = (R_Z + w_1 D_Z + w_2 P_Z + w_3 \mathrm{ML}_Z) \\ + (w_4 E_Z^{\rm SS} + w_5 E_Z^{\rm CS} + w_6 E_Z^{\rm CC}).\end{aligned} \quad (1)$$

For each term X_Z the subscript Z denotes the renormalization of the raw score X by its standard deviation, *i.e.*

$$X_Z = \frac{(X - \overline{X})}{\left[\overline{(X - \overline{X})^2}\right]^{1/2}},$$

which helps to combine the energy terms of different scales. The R factor is defined as

$$R = \frac{\sum_{hkl} \left| s|F_{\rm calc}(hkl)| - |F_{\rm obs}(hkl)| \right|}{\sum_{hkl} F_{\rm obs}(hkl)}, \quad (2)$$

where $s = \sum_{hkl} |F_{\rm obs}(hkl)| / \sum_{hkl} |F_{\rm calc}(hkl)|$. The standard deviation of the calculated and observed diffraction data is defined as

$$D = \sum_{hkl} [c|F_{\rm calc}(hkl)|^2 - |F_{\rm obs}(hkl)|^2]^2, \quad (3)$$

where c is a scale factor to balance the two terms. P is defined to maximize the Pearson correlation coefficient between $F_{\rm calc}$ and $F_{\rm obs}$:

$$P = 1 - \frac{\sum_{hkl} [|F_{\rm obs}(hkl)| - |\overline{F_{\rm obs}}|][|F_{\rm calc}(hkl)| - |\overline{F_{\rm calc}}|]}{\left\{\sum_{hkl} [|F_{\rm obs}(hkl)| - |\overline{F_{\rm obs}}|]^2\right\}^{1/2} \left\{\sum_{hkl} [|F_{\rm calc}(hkl)| - |\overline{F_{\rm calc}}|]^2\right\}^{1/2}}. \quad (4)$$

ML in (1) is the maximum-likelihood score defined in Read (2001*a*).

The last three terms in (1) represent the penalties on the clashes between surface atoms ($E^{\rm SS}$), the clashes between core and surface atoms ($E^{\rm CS}$) and the clashes between core atoms ($E^{\rm CC}$) of the protein copies, where the penalty energies take a unified form of $E = r_{\rm vdw}^2 - r^2$ if the distance of two target atoms $r < r_{\rm vdw}$ or $E = 0$ otherwise. The van der Waals distance cutoff is set as $r_{\rm vdw}$ = 3 Å. The weight on clashes varies depending upon whether they involve surface or core atoms. The idea is for the surface clashes to a have low weight since these clashes might indicate a minor error in the model or in the placement of the model, while clashes involving core atoms usually indicate that either the model is far from the native and/or that the MR solution is incorrect, and therefore a stronger weight is assigned. To increase the speed of simulations, only C^α atoms are used in the calculation of the clash score. A C^α atom is defined to be a core atom if its accessible surface area is below 20%, where the radius of each C^α atom for the purpose of calculating accessible surface area is set to 4 Å. The accessible surface area is calculated using the LCPO method (Weiser *et al.*, 1999). In the first round of simulation the clash scores and the ML score are turned off to speed up the simulation process, but in the second round both parts of the match and clash scores from (1) are used (Fig. 1). Later, the ML score is turned on.

To determine the weighting parameters, we created 1387 structural decoys from 40 nonredundant proteins using *3DRobot* (Deng *et al.*, 2016; see below). Here, the decoys refer to computationally generated structural models of the protein of varying quality (*i.e.* r.m.s.d. and TM-score) relative to the

native. For each decoy model *MR-REX* creates 300 candidate MR solutions, and for each set of 300 candidate MR solutions that with the lowest XScore is selected to calculate the average crystallographic r.m.s.d. (cRMSD) of the 1387 decoys. To calculate the cRMSD, the protein model is first superimposed on the native structure such that the electron-density correlation is maximized; this is considered to be the best placement of the search model. The *fitmap* command of *Chimera* (Pettersen *et al.*, 2004) was used to place the protein model into the electron-density map of the native protein. Next, the C^{α} r.m.s.d. without superimposition is calculated between the MR candidate solution and the best placement of the search model for all pairs of symmetry mates, which is termed the cRMSD. All possible alternate origins are considered, and the lowest cRMSD is reported. Here, solutions with a cRMSD of >8 Å are considered to be equally bad and their cRMSDs are all set to 8 Å; this cutoff can prevent some spurious large reductions from decreasing the average cRMSD values. Next, a quick Monte Carlo (MC) simulation is performed to search through the parameter space with movement involving random changes to one randomly picked weight. After each movement, the candidate MR solutions are re-ranked based on (1) using the new weighting parameters and the average cRMSD values are calculated based on the newly selected MR solutions. The MC iteration continues until there is no further improvement in the cRMSD within each addition of 100 MC cycles. The initial weight parameter (w_i) is taken as a random value between 0 and σ_i/σ_D, with σ_i being the standard deviation of the ith scoring term and σ_D being that of D in (1). Multiple runs with different initial parameters are found to converge to the same optimized weights: $w_1 = 0.646$, $w_2 = 0.503$, $w_3 = 18.09$, $w_4 = 0.027$, $w_5 = 0.302$ and $w_6 = 2.876$. The ML score has by far the highest weight and dominates the other terms.

We note that although the *MR-REX* search is driven by the XScore, the XScore or the *Z*-score alone are not good indicators of the success of MR solution. However, we found that the greatest sign of success is that there is a tight cluster of solutions with Xscores far lower than other placements, a finding analogous to the approach used for protein structure prediction (Zhang & Skolnick, 2004*a*).

2.3. Molecular-replacement search by replica-exchange Monte Carlo sampling

2.3.1. Setting the replica temperatures. The temperature of the nth replica in the REMC simulations is given by

$$T_n = T_{\min}\left(\frac{T_{\max}}{T_{\min}}\right)^{(n-1)/(n_{\rm rep}-1)}, \qquad (5)$$

where $n_{\rm rep} = 300$ is the total number of replicas. Since different proteins have different sizes and energy scales, we set the temperature ranges according to the initial XScore of the systems, *i.e.* the highest temperature ($T_{\max}$) equals 0.1 times the initial score of the worst-scoring replica and the lowest temperature ($T_{\min}$) is 0.005 times the initial score of the best-scoring replica; this temperature set can help to ensure that the simulations of different protein systems have an approximately constant acceptance rate of the replica-swap movements.

The REMC simulation consists of up to 9000 cycles, where each cycle runs 20 local movements. At the end of each cycle, global swap movements are attempted iteratively between pairs of adjacent replicas which are accepted/rejected on the standard Metropolis criterion (Metropolis *et al.*, 1953). Supplementary Fig. S1 shows a typical example of the XScore *versus* the REMC cycles for one of the models of the hexamerization domain of *N*-ethylmaleimide-sensitive fusion protein (PDB entry 1d2n), which shows that the $n_{\rm rep} = 300$ is high enough to maintain sufficient overlaps of the simulations, which is essential to give a high acceptance rate for the replica swaps.

2.3.2. Rescaling MC movements. Before the start of the MC simulations, the center of mass of the protein is translated to the origin. The placement of the protein is then specified by rotations around and translations along the x, y and z axes. There are a total of six degrees of freedom per protein, where the rotational degrees of freedom are effectively applied first and the protein is then effectively translated. If there are multiple copies in the asymmetric unit, the proteins can be placed one by one or all at once. The default is to place them all at once. In all cases in this paper there is only one copy of the protein in the asymmetric unit. The sizes of the movements are randomly picked according to a Gaussian distribution, the standard deviation of which is updated for each replica and degree of freedom once every 50 steps, *i.e.*

$$\sigma_{i,j} = \frac{\sum_k (s_{ijk}^2 |d_{ijk}| p_{ijk})/q_{ijk}}{\sum_k (|d_{ijk}| p_{ijk})/q_{ijk}}, \qquad (6)$$

where $\sigma_{i,j}$ is the standard deviation of the movement size distribution of the ith degree of freedom of the jth replica, s_{ijk} is the size of the kth movement of the ith degree of freedom of the jth replica, d_{ijk} is the change in XScore resulting from the movement, p_{ijk} is the probability of accepting the movement and q_{ijk} is the probability of making the movement. At high temperatures p_{ijk} tends to be high even for large movements (large s_{ijk}), and therefore $\sigma_{i,j}$ will tend to be larger at high temperatures than at low temperatures. Thus, (6) ensures that low-temperature replicas make smaller movements while high-temperature replicas make larger movements, to keep a reasonable acceptance rate of movements in different replicas. In addition, the movement sizes of the low-temperature replicas tend to decrease throughout the simulation as convergence is reached to fine-tune the conformational search near low-energy basins, as p_{ijk} decreases for large movements during the simulation.

The standard deviation of the size distributions of the initial translation and rotation movements is set as

$$\begin{cases} \sigma_{i,j}^{\rm tran} = a_i \sigma_{\min}^{\rm tran} + a_i(\sigma_{\max}^{\rm tran} - \sigma_{\min}^{\rm tran})(j-1)/(n_{\rm rep}-1) \\ \sigma_{i,j}^{\rm rota} = \sigma_{\min}^{\rm tora} + (\sigma_{\max}^{\rm rota} - \sigma_{\min}^{\rm rota})(j-1)/(n_{\rm rep}-1) \end{cases}, \qquad (7)$$

where a_i are the unit-cell vector lengths, $\sigma_{\rm min}^{\rm tran} = 0.01$, $\sigma_{\rm max}^{\rm tran} = 0.5$, $\sigma_{\rm min}^{\rm rota} = 0.02$ and $\sigma_{\rm max}^{\rm rota} = 0.12$. After a translation move, the structure factor is evaluated by

$$F(\mathbf{H}, \Delta x) = \sum_s F_s(\mathbf{H}) \exp(\mathbf{R}_s \Delta x \mathbf{H}), \quad (8)$$

where $F(\mathbf{H}, \Delta x)$ is the structure factor of the unit cell when the protein has been translated by Δx, $\mathbf{H}$ is the Miller index vector, F_s is the structure factor of the sth symmetry mate and $\mathbf{R}_s$ is the associated rotation matrix. (8) greatly speeds up movements involving translations because recalculating the structure factor owing to the orientation is not necessary. When a rotation is performed, the new structure factor is calculated using a fast rotation method (Castellano *et al.*, 1992). Since there is a difference in the effectiveness of each movement type per unit time, the ith degree of freedom of the jth replica is sampled with a probability equal to

$$p_{i,j} = \frac{\sum_k |d_{ijk}| p_{ijk} \big/ \sum_k \Delta t_{ijk}}{\sum_n \sum_k |d_{njk}| p_{njk} \big/ \sum_k \Delta t_{njk}}, \quad (9)$$

where Δt_{ijk} is the time taken to calculate Δd_{ijk} and the associated scattering amplitudes, and n runs over all degrees of freedom. Translations tend to be sampled more than rotations, since they are faster to compute than rotations.

2.3.3. Search termination and model selection. After every 50 REMC cycles, the standard deviations of the XScores of all replicas and the 30 lowest-temperature replicas are calculated. If the standard deviation of the XScores of the 30 lowest-temperature replicas is lower than 10% of the standard deviation for all 300 replicas, the cRMSDs of the lowest-temperature replica to the other five lowest-temperature replicas are calculated. If all of the calculated cRMSDs are below 0.5 Å the simulations are considered to have converged, the REMC search is terminated early to save CPU time and all 300 potential solutions are output. In case an early termination does not occur, the simulation will run for a predetermined time, which is set in the parameter file. *MR-REX* converged early in 10% of cases. The average time to early convergence is one and half hours. The maximum run time of each *MR-REX* round used in this paper is 2 h, although a benchmark test showed that the average performance of *MR-REX* is reduced by only 5% when it is run for only 1 h. The speed of *MR-REX* is mainly determined by the number of reflections used and the space group of the crystal, but not by the size of the protein since *MR-REX* uses a method similar to that of Castellano *et al.* (1992) to calculate the scattering in an efficient manner. The placement with the minimum XScore is selected as the final MR model by *MR-REX*.

Overall, *MR-REX* implements an iterative Monte Carlo simulation procedure, as shown in Fig. 1, which enables a convenient incorporation of the steric clash score and occupancy optimization. To make the simulations more efficient, the REMC simulations are performed in multiple rounds, where a new feature of the program is turned on each round. Some features of the program are slow and may not be needed for successful MR, so they are not used at first and are then turned on in later rounds. For example, in the first round of *MR-REX* simulations the clash score is turned off and all occupancies are equal to 1, in the second round the clash score is turned on, and in the third round occupancy optimization (see below) is performed, which aims to dynamically remove some of the structurally variable or incorrectly modeled regions through Monte Carlo simulations.

2.4. Occupancy optimization

Since incorrect portions of the protein model affect the success rate of MR more than deletions, *MR-REX* implements an occupancy-optimization procedure to delete incorrect portions of the protein model. Although it cannot be known with certainty which portions of the protein model are incorrect, it is possible to make an educated guess as to which regions are inaccurate. For instance, the termini, unstructured (or intrinsically disordered) regions and surface residues are more likely to be inaccurate than other regions. Here, the categorization of different structure regions is determined by *DSSP* (Frishman & Argos, 1995).

To make a quantitative estimation of the local structure quality, we take the same set of 1387 decoy models from the training data set, which are superimposed on their native structures using the TM-score (Zhang & Skolnick, 2004*b*). The probability of the deviation of each residue from the native is calculated along with its distance from the nearest terminus in sequence space ($d_{\rm tail}$), its solvent-accessible surface area (SA) and its secondary-structure type (SS), *i.e.*

$$p(d|d_{\rm tail}, {\rm SA}, {\rm SS}) = \frac{n(d|d_{\rm tail}, {\rm SA}, {\rm SS})}{N(d_{\rm tail}, {\rm SA}, {\rm SS})}, \quad (10)$$

where $n(d|d_{\rm tail}, {\rm SA}, {\rm SS})$ is the number of residues that have a deviation d from the native for a given bin of ($d_{\rm tail}$, SA, SS) and $N(d_{\rm tail}, {\rm SA}, {\rm SS})$ is the total number of residues in the bin.

Before the REMC simulation, the probability that the residue deviates from the native by $d > 2$ Å is calculated according to (10). A residue is considered to be potentially inaccurate if the estimated probability that the residue deviates from the native by >2 Å is greater than 40%. Neighboring inaccurate residues are grouped together into segments and the ten longest segments are subjected to random occupancy changes during the REMC simulations. If a segment is chosen at random the occupancies of the atoms in that segment will either be changed from 1 to 0 or from 0 to 1, with the acceptance of the changes guided by the standard Metropolis criterion (Metropolis *et al.*, 1953). The benchmark results show that up to six occupancy-optimization iterations are needed to achieve the best results.

Estimates of local error can also be used by *Phaser* by either setting *B* factors according to predicted errors (Bunkóczi *et al.*, 2015) or removing regions that are predicted to be inaccurate.

2.5. Structure-factor calculation for X-ray diffraction

To facilitate the calculation of the structure factor of the entire unit cell for protein placements, we first move the center of mass of the protein to the origin of the Cartesian system and

orient the structure such that the principal axes of the protein line up with the unit-cell vectors. The X-ray diffraction is calculated for space group $P1$ on a finely spaced grid of non-integer value Miller index vectors, $\mathbf{H}$, using

$$F_{P_1}(\mathbf{H}) = \sum_j^{n_{\rm atom}} g_j(\mathbf{H}) \exp\left[2\pi i \mathbf{H}\mathbf{X}_j - \frac{B_j(\mathbf{H}\mathbf{C}_f)^2}{12}\right], \quad (11)$$

where $n_{\rm atom}$ is the number of atoms, $\mathbf{X}_j$ is the location of the jth atom in fractional coordinates, B_j is the B factor and $\mathbf{C}_f$ is the matrix converting Cartesian coordinates to fractional coordinates. $g_j(\mathbf{H})$ is the structure factor of the jth atom with the excluded volume taken into account using

$$g_j(\mathbf{H}) = f_j(\mathbf{H}) - \rho f_{{\rm exc},j}(\mathbf{H}), \quad (12)$$

where $f_j(\mathbf{H})$ is calculated according to Rez *et al.* (1994), ρ is the electron density of the buffer (set to 0.334 e^- Å^{-3}) and $f_{{\rm exc},j}(\mathbf{H})$ is the scattering of a Gaussian sphere with a volume equal to the excluded volume of the atom, *i.e.*

$$f_{{\rm exc},j}(\mathbf{H}) = \pi^{2/3} r_{{\rm exc},j}^3 \exp[-\pi^2(\mathbf{H}\mathbf{C}_f)^2 r_{{\rm exc},j}^2], \quad (13)$$

where $r_{{\rm exc},j}$ is the radius of atom j corresponding to the excluded volume. $F_{P_1}(\mathbf{H})$ is effectively rotated and translated using the procedure in Castellano *et al.* (1992) in order to identify the sum of the contributions of the individual symmetry mates, which equal the total structure factor. When performing occupancy optimization, a separate $F_{P_1}(\mathbf{H})$ is calculated for each segment and the scatterings for the appropriate segments are summed.

It was found that the calculated X-ray diffraction at low and high resolution has been systematically overestimated, most likely owing to imperfections in how the buffer is accounted for and the lack of side chains. A correction factor for the calculated X-ray intensity as a function of q is estimated in the following way. X-ray diffraction reflection intensities are calculated for a set of 1078 nonredundant proteins randomly collected from the PDB using *PISCES* (Wang & Dunbrack, 2003). These 1078 proteins are from the same list of proteins obtained by *PISCES* from which the test and training proteins were taken. The proteins in the test set described in §3.1 are excluded from this list. The average scattering intensity is calculated in every 0.1 Å^{-1} q bin and compared with the experimental data. For every protein, the ratio of the average calculated intensity in every 0.1 Å^{-1} q bin to the experimental intensity is calculated; thus, the ratio of the calculated diffraction intensity to the experimental diffraction intensity is calculated as a function of q. The results from the protein set are combined and the average ratio of the calculated diffraction intensity to the experimental diffraction intensity is calculated as a function of q. These data are used to provide a correction factor that reduces the simulated X-ray diffraction error. This helps to compensate for errors in properly accounting for the effect of the buffer and the lack of side chains.

2.6. *B*-factor setting and calculation

The B factor is initially set as $B_{{\rm init},j} = 20$ Å^2 in *MR-REX* and the structure factor is then calculated by

$$F_{\rm calc}(\mathbf{H}) = \sum_s^{n_{\rm sym}} \sum_j^{n_{\rm atom}} g_h(\mathbf{H}) \exp\left[2\pi i \mathbf{H}(\mathbf{R}_s\mathbf{X}_j + \mathbf{T}_s) - \frac{B_{{\rm init},j} q^2}{48\pi^2}\right], \quad (14)$$

which can be different from the observed diffraction

$$F_{\rm obs}(\mathbf{H}) = \sum_s^{n_{\rm sym}} \sum_j^{n_{\rm atom}} g_h(\mathbf{H}) \exp\left[2\pi i \mathbf{H}(\mathbf{R}_s\mathbf{X}_j + \mathbf{T}_s) - \frac{B_{{\rm exp},j} q^2}{48\pi^2}\right], \quad (15)$$

where $B_{{\rm exp},j}$ is the actual B factor of the jth atom. The difference between $B_{{\rm init},j}$ and $B_{{\rm exp},j}$ is thus

$$\delta B = B_{{\rm exp},j} - B_{{\rm init},j} = -\frac{48\pi^2}{q^2} \ln\left[\frac{|F_{\rm obs}(\mathbf{H})|}{|F_{\rm calc}(\mathbf{H})|}\right], \quad (16)$$

which can be estimated by linear regression of the natural log of the experimental amplitudes to the calculated amplitudes. The value of $B_{{\rm init},j} + \delta B$ is used as the final B factor in the diffraction calculation.

3. Results

3.1. Data sets

To train and test *MR-REX*, we collected a set of 78 non-redundant proteins with a pairwise sequence-identity cutoff of 30% from the PDB using *PISCES*. All of the testing proteins are at most distantly related to the training proteins. These proteins contain one chain and have a resolution higher than 3 Å, R factors lower than 0.3 and sequence lengths of 40–300 residues. From the list of proteins found using *PISCES*, 40 proteins were randomly selected to train the *MR-REX* program (see §2) and another 38 randomly selected proteins were used for testing.

The structure decoys for the proteins were created by *3DRobot* (Deng *et al.*, 2016), a program designed for the creation of nonhomogenous and well packed protein conformations from the native structure based on *I-TASSER* structure-assembly simulations (Yang *et al.*, 2015). Here, multiple levels of threading templates, including those with very low sequence identity, are used in *3DRobot*/*I-TASSER* to generate protein-like conformations of varying quality. As the decoys are built for the query sequences, the *3DRobot* decoys have the same side chains as the query, but the side chains are removed before use in MR. In order to have structures of diverse quality, the *3DRobot* decoys are split into 40 bins in TM-score space from 0.59 to 0.99, where one model is randomly selected from each bin for each protein; this results in 1303 (1387) nonredundant structure models that constitute our test (training) model set. As some TM-score bins have no decoys, the number of final decoy models (1303 or 1387) is slightly lower than 38 × 40 or 40 × 40. The decoy structures for both testing and training proteins can be downloaded at https://zhanglab.ccmb.med.umich.edu/MR-REX/DataSets.tar.gz.

Table 1
Summary of MR results by *MR-REX* and *Phaser* on 38 test proteins with decoy models generated by 3*DRobot* and *LOMETS*.

For each protein multiple decoy models are attempted, where the decoy with the lowest TM-score (for the 3*DRobot* decoys) or with the lowest sequence identity (for the *LOMETS* templates) which is successfully placed with a cRMSD of <2 Å is reported in the last four columns. The data in parentheses are obtained when a successful case is defined based on *phenix.autobuild*, *i.e.* an R_{free} of <0.4 and the model quality is improved. M and CC in the fold class column indicate membrane and coiled coil, respectively.

Protein PDB code	Length (amino acids)	Resolution (Å)	Solvent content	Fold class	Space group	Worst TM-score		Worst sequence identity (%)	
						Phaser	*MR-REX*	*Phaser*	*MR-REX*
1d2n	246	1.75	52	α/β	$P6$	0.83 (0.83)	0.82 (0.86)	70 (70)	27 (70)
1okc	292	2.20	55	M	$P2_12_12$	0.76 (0.89)	0.76 (0.90)	52 (52)	52 (52)
1r0u	142	1.75	55	β	$P3_221$	0.84 (0.84)	0.84 (0.84)	100 (100)	100 (100)
1su0	136	2.30	42	$\alpha+\beta$	$I422$	0.79 (0.87)	0.76 (0.83)	22 (22)	22 (46)
1v05	96	1.43	35	β	$P6_122$	0.77 (0.77)	0.77 (0.83)	31 (31)	31 (31)
1vpq	260	2.20	44	α/β	$C222_1$	0.82 (0.87)	0.80 (0.91)	33 (100)	33 (33)
1x8y	74	2.20	71	CC	$P6_522$	0.81 (0.59)	0.77 (0.59)	64 (64)	64 (64)
2bou	137	1.90	61	β	$P2_12_12$	0.76 (0.86)	0.87 (0.87)	99 (99)	99 (99)
2il5	162	2.30	61	$\alpha+\beta$	$P6_2$	0.74 (0.85)	0.67 (0.85)	12 (17)	12 (17)
2p17	249	1.52	32	β	$P2_12_12_1$	0.77 (0.88)	0.79 (0.88)	35 (35)	35 (35)
2rbk	261	1.00	46	α/β	$P2_12_12_1$	0.79 (0.91)	0.90 (0.90)	21 (30)	30 (100)
2yq9	213	1.90	43	α/β	$P2_12_12_1$	0.74 (0.84)	0.77 (0.82)	48 (79)	48 (79)
3b7c	121	1.70	56	$\alpha+\beta$	$P6_522$	0.83 (0.84)	0.75 (0.83)	15 (15)	15 (15)
3bw6	137	2.50	66	α/β	$P3_221$	0.79 (0.79)	0.72 (0.72)	40 (40)	40 (40)
3chv	279	1.45	40	α/β	$C121$	0.74 (0.86)	0.77 (0.93)	61 (97)	24 (14)
3fzx	212	2.20	59	β	$P6_422$	0.83 (0.83)	0.81 (0.87)	41 (41)	41 (40)
3hyn	186	1.20	34	α/β	$P2_12_12_1$	0.86 (0.88)	0.81 (0.83)	96 (96)	96 (96)
3k93	223	2.15	60	α/β	$P321$	0.74 (0.79)	0.74 (0.81)	98 (98)	98 (98)
3mt0	281	1.58	40	α/β	$P12_11$	0.81 (0.81)	0.84 (0.84)	30 (44)	44 (44)
3n2q	282	2.55	55	α/β	$P3_121$	0.85 (0.91)	0.86 (0.91)	40 (40)	40 (40)
3onj	97	1.92	35	α	$P2_12_12_1$	0.80 (0.83)	0.74 (0.88)	100 (100)	100 (100)
3pu6	138	2.60	44	α/β	$P2_12_12_1$	0.76 (0.88)	0.68 (0.84)	98 (98)	98 (98)
3pyw	178	1.80	65	α	$P4_12_12$	0.70 (0.70)	0.72 (0.72)	100 (100)	100 (100)
3q6b	155	1.50	39	α/β	$C121$	0.62 (0.59)	0.62 (0.59)	18 (18)	18 (18)
3vqf	85	1.20	43	α/β	$C121$	0.70 (0.70)	0.68 (0.77)	27 (27)	22 (22)
3vwc	146	1.50	50	β	$C121$	0.76 (0.76)	0.65 (0.80)	22 (22)	12 (22)
3zdb	246	1.47	59	α/β	$P2_12_12$	0.79 (0.79)	0.76 (0.81)	34 (34)	34 (34)
4a3z	136	1.55	45	β	$P4_32_12$	0.80 (0.80)	0.79 (0.80)	98 (98)	22 (98)
4dcd	184	1.69	43	β	$I222$	0.78 (0.85)	0.71 (0.85)	44 (44)	44 (44)
4is7	140	2.75	70	α	$P6_522$	0.80 (0.80)	0.78 (0.81)	59 (59)	18 (59)
4kr1	219	2.50	52	α/β	$P6_322$	0.80 (0.80)	0.73 (0.82)	40 (40)	40 (54)
4l8g	157	1.52	38	α/β	$P6_36$	0.75 (0.81)	0.75 (0.81)	25 (37)	25 (37)
4lvp	124	2.32	55	α	$P6_422$	0.77 (0.88)	0.69 (0.82)	100 (100)	100 (100)
4m6t	177	2.50	74	$\alpha+\beta$	$H32$	0.70 (0.81)	0.74 (0.74)	100 (100)	100 (100)
4mdn	94	1.91	68	β	$P4_12_12$	0.64 (0.64)	0.65 (0.64)	67 (67)	67 (67)
4mjf	225	1.99	50	α	$C121$	0.79 (0.86)	0.74 (0.88)	52 (97)	52 (97)
4nbr	268	1.35	41	α/β	$I222$	0.88 (0.88)	0.84 (0.91)	22 (22)	22 (22)
4oq4	186	1.49	59	α/β	$P3_121$	0.80 (0.80)	0.79 (0.84)	22 (22)	22 (24)
Average	183	1.88	60			0.77 (0.82)	0.76 (0.81)	54 (59)	49 (58)

3.2. A comparison of failure points of MR by *MR-REX* and *Phaser* on 3*DRobot* decoys

Not all of the structural decoys can be correctly placed in the unit cell. For each protein up to 40 decoy structures from 3*DRobot* with nearly continuous accuracy are attempted, where the decoy with the worst structure for which MR is successful is recorded. Table 1 lists the worst model according to the TM-score (Zhang & Skolnick, 2004*b*; Xu & Zhang, 2010) for which *MR-REX* and *Phaser* succeeded for each protein, where a success is defined as the cRMSD between the MR placement and the best placement of the search model being below <2 Å. Using this criterion, the average point of failure is TM-score = 0.76 for *MR-REX* and TM-score = 0.77 for *Phaser*. The failure-point TM-score comparison can be visualized in Fig. 2. Here, we note that TM-score is not a perfect measure of model quality, especially for multi-domain proteins, because the model may capture both domains accurately but have an incorrect orientation between the two domains, which will result in a low TM-score value. Nevertheless, we found that for all of the 14 multi-domain proteins in our benchmark data set most of the decoy structures have a similar domain orientation relative to the native structure (see Supplementary Fig. S2 as an illustrative example).

We also tried to assess the failure points of the algorithms using the criterion of structure determination, where MR is considered to be successful if R_{free} < 0.4 and the TM-score of the final model is higher than that of the initial input model when running the *phenix.autobuild* program (Adams *et al.*,

2010; Afonine *et al.*, 2012; Terwilliger, 2004; Terwilliger *et al.*, 2008) based on the given MR solution. This criterion is similar to what has been used in previous MR studies (Giorgetti *et al.*, 2005; Wang *et al.*, 2016). The last iteration of *MR-REX*, which used the ML score during the REMC search, was not used in this analysis. The comparison of *MR-REX* and *Phaser* is tabulated in parentheses in Table 1, where the average failure points of *MR-REX* and *Phaser* are TM-scores of 0.81 and 0.82, respectively. The difference in the worst TM-score is not statistically significant, with the *p*-value of Student's t-test being 0.21.

Here, the TM-scores of the testing models at the failure points according to a *phenix.autobuild* $R_{\rm free}$ of <0.4 are significantly higher than those according to a cRMSD of <2 Å for both *MR-REX* and *Phaser*, indicating that the first criterion is generally stricter than the second. There are, however, two cases (PDB entries 1x8y and 3q6b) for which the failure-point TM-scores by *MR-REX* according to $R_{\rm free}$ < 0.4 are obviously lower than those according to cRMSD < 2 Å. PDB entry 1x8y from the human lamin coil 2B is a small protein consisting of a single α-helix (Supplementary Fig. S3). Translating PDB entry 1x8y by a single helix turn leaves the electron density, which is what matters in X-ray crystallography, largely unchanged but results in a cRMSD that is over 2 Å. PDB entry 3q6b is the BamA POTRA4-5 protein from *Escherichia coli*, the topology of which has an approximate 180° rotational axis which leaves the electron density approximately the same but results in a large cRMSD (Supplementary Fig. S3*b*). Overall, although $R_{\rm free}$ < 0.4 seems to be a stricter and more practical criterion to assess the success of MR than cRMSD < 2 Å, the threshold depends on the power and efficiency of the follow-up structure-refinement method. It can be expected that even if *phenix.autobuild* is not able to solve a protein structure with a low cRMSD, the case may be solved by other methods. On the contrary, the definition of cRMSD represents the structural closeness of the MR model to the best placement of the probe model, which is independent of the follow-up structure-determination programs. Therefore, we will use both criteria in our benchmark experiments.

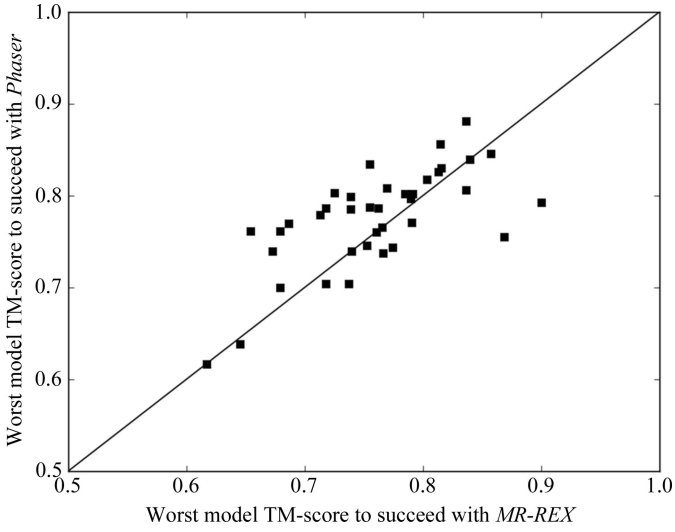


Figure 2
The TM-score of the worst initial structure model to succeed with *MR-REX versus* that with *Phaser*. For each target, up to 40 decoys created by 3*DRobot* are attempted, where the worst decoys with the highest TM-score with a successful MR solution are shown in the plot.

3.3. Aggregate results for MR on 1303 decoy structures

To further examine the performance of the methods in detail, in Fig. 3 we present the cRMSDs of *MR-REX versus Phaser* for all 1303 structure decoys. The data were split into three different groups based on the accuracy of the 3*DRobot* structure decoys. Fig. 3(*a*) displays the results for the 285 high-solution decoys with an r.m.s.d. to the native of below 2 Å. In this r.m.s.d. range the vast majority of these cases succeeded using both methods, but the number of failures with *MR-REX* were slightly fewer than those with *Phaser*. If we consider a cRMSD of <2 Å as successful MR, for example, *MR-REX* failed in 11 cases, whereas *Phaser* failed in 15 cases.

When the decoy r.m.s.d. to the native ranges from 2 to 4 Å the number of failures increases dramatically (Fig. 3*b*). Again,

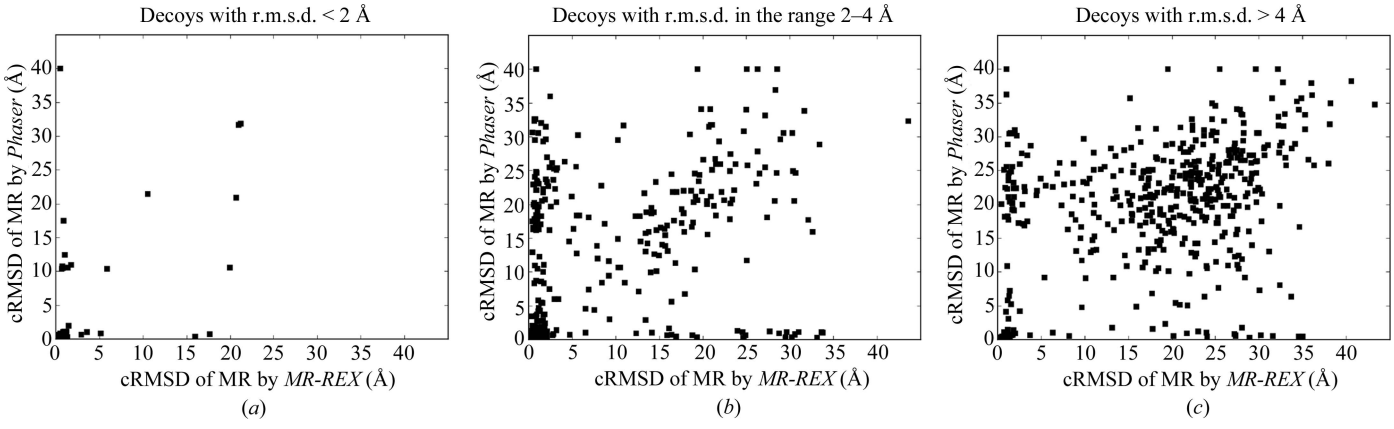


Figure 3
The cRMSDs of MR solutions by *MR-REX versus* those by *Phaser* on 1303 structure decoys. The results are split according to the r.m.s.d. of the initial decoys in the ranges (*a*) <2 Å, (*b*) 2–4 Å and (*c*) >4 Å. For the cases in which *Phaser* did not produce a result, the cRMSD is set to 40 Å for the purposes of the plot.

the number of cases that succeeded using *MR-REX* and failed using *Phaser* far surpasses the number of cases in which *Phaser* succeeded but *MR-REX* failed, *i.e.* in 77 of the 550 medium-accuracy decoys *MR-REX* succeeds but *Phaser* fails, while the reverse occurs in only 31 cases.

For the lowest-accuracy decoys with an r.m.s.d. to the native of >4 Å, the majority of cases failed for MR by both *MR-REX* and *Phaser*. Nevertheless, *MR-REX* could still generate an MR solution with a cRMSD below 2 Å for 60 cases, while this number was 43 for *Phaser*.

Of the 1303 decoys, *MR-REX* generated correct MR solutions for 699 cases, while *Phaser* did so for 632 cases, according to the criterion of a cRMSD of <2 Å. In 123 cases *MR-REX* correctly placed a decoy structure that *Phaser* was unable to place, while the reverse occurs in 56 cases; this shows that the two methods are complementary to each other. Combining the results of the two programs, we obtained 755 cases for which at least one of the programs succeeded. Such a complementary effect is most significant for the low-accuracy protein models. When the r.m.s.d. of the decoy structures is higher than 4 Å, for example, there are only 23 cases that can be commonly solved by both *MR-REX* and *Phaser*, while there are 80 cases that can be solved by either *MR-REX* or *Phaser*.

Similarly, if we use the criteria of $R_{\rm free}$ and TM-score of the *phenix.autobuild* models to assess the success rate of MR, *MR-REX* correctly placed 540 decoy structures while *Phaser* correctly placed 532 decoys. In 70 cases *MR-REX* correctly placed a decoy that *Phaser* was unable to place, while the reverse occurred in 62 cases. If we combined the *MR-REX* and *Phaser* results, we obtained 602 cases for which at least one of the programs succeeded.

Again, the aggregate decoy results based on the two successful criteria are not entirely consistent. There were 39 decoy cases with a cRMSD of >2 Å by *MR-REX* which succeeded according to a *phenix.autobuild* $R_{\rm free}$ of <0.4, most of which are structural decoys from PDB entries 1x8y and 3q6b owing to their approximate rotational and translational symmetries as displayed in Supplementary Fig. S3. There were 198 cases which succeeded according to a cRMSD of <2 Å but failed according to a *phenix.autobuild* $R_{\rm free}$ of <0.4, demonstrating that there is considerable room for the further development of structure-refinement algorithms.

Finally, we note that the median cRMSDs of *MR-REX* and *Phaser* are 1.52 and 2.53 Å, respectively, for the 1303 decoys. This difference is much larger than the difference shown in the number of cRMSDs of <2 Å. This means that nearly half of the MR solutions generated by *MR-REX* have a very low cRMSD (<1.52 Å), even though only 77% of correctly placed models resulted in a successful structural solution. This is important because some of these low-cRMSD MR solutions may not have been solved by the current structure-refinement programs (*phenix.autobuild*) but may be solved by future advanced methods. In contrast, *Phaser* solutions have a higher median cRMSD, indicating that many of the MR solutions have an incorrect placement except for those with a cRMSD of <2 Å, which cannot benefit from potentially advanced model-building methods.

3.4. The major difference between *MR-REX* and *Phaser* is for low-accuracy decoys

To illustrate the dependence of MR on decoy accuracy, in Fig. 4(*a*) we present the success rates of *MR-REX* and *Phaser* as a function of the r.m.s.d. of the initial model to the native when using a cRMSD of <2 Å to assess success, while Fig. 4(*b*) shows the data *versus* the TM-score of the initial models. As expected, the success rate of both programs depends on the quality of the initial probe models, *i.e.* the success rate decreases with increasing r.m.s.d. or decreasing TM-score. Interestingly, the performance of the MR program appears to be more sensitive to the TM-score of the probe models than the r.m.s.d. If we use a TM-score of >0.8 as an indicator of

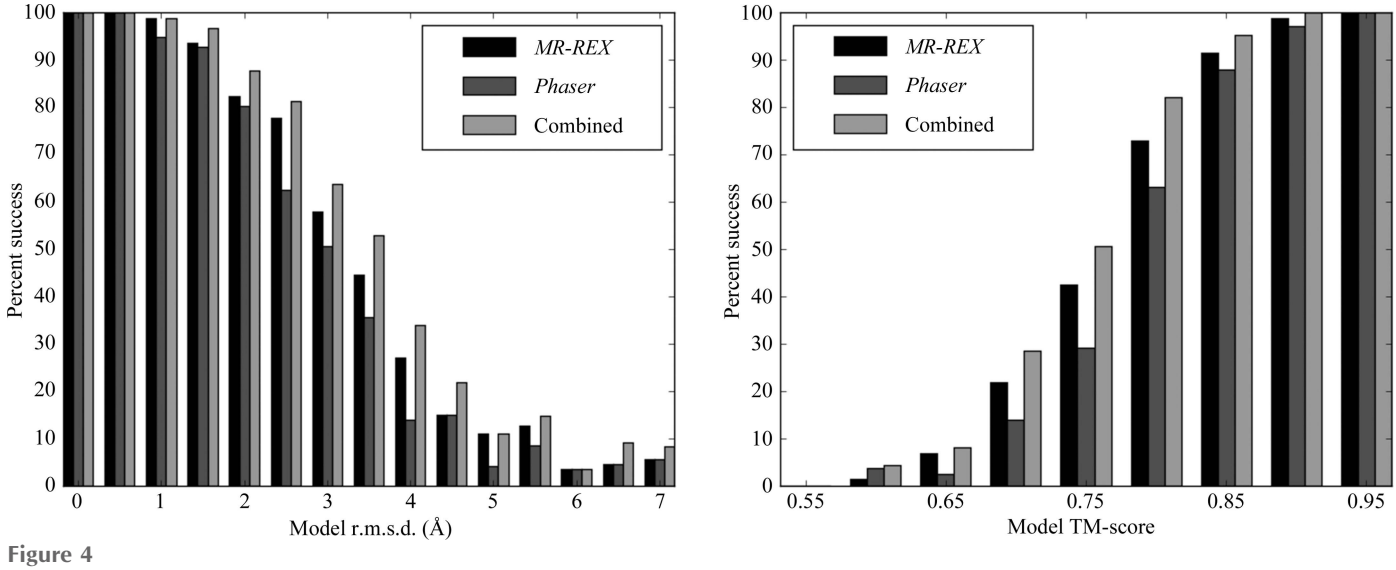


Figure 4
The percentage of successful MR cases as a function of r.m.s.d. and TM-score of the starting model to the native. *MR-REX* is represented in black, *PHENIX* is represented in dark gray and the combined results are represented in light gray.

success, for example, the false-positive (FP) rate and false-negative (FN) rate with a cRMSD of <2 Å are 0.11 and 0.17, respectively. The lowest FP and FN rates are 0.17 and 0.20, respectively, if using an r.m.s.d. of <3.25 Å, which are both higher than those for the TM-score.

Accordingly, when the structure model has a TM-score of >0.95 *MR-REX* and *Phaser* are almost 100% successful. Both *MR-REX* and *Phaser* experience a sharp drop in their success rates when the TM-score is below 0.85, which usually corresponds to the accuracy cutoff for distant *versus* closely homologous structure prediction (Huang *et al.*, 2014; Mariani *et al.*, 2011; Zhang, 2009). However, it is in this region that *MR-REX* obviously outperforms *Phaser*. For the structure models with a TM-score in the range 0.75–0.8, for example, the success rates for *MR-REX* and *Phaser* are 42% and 29%, respectively. Overall, among the 834 cases with a TM-score of <0.85 *MR-REX* succeeded in 249 cases while *Phaser* succeeded in 192 cases; thus, *MR-REX* increased the number of successes by 30%. If we reduce the cutoff to a TM-score of <0.8, this number increases to 49%, which indicates that the difference between *MR-REX* and *Phaser* mainly occurs in cases with a low-accuracy starting model.

3.5. The effects of clash, occupancy optimization and ML score on MR

As mentioned in §2, *MR-REX* implements multiple-step Monte Carlo simulations, in which a new feature is turned on in each step (Fig. 1). In our test, the first round of *MR-REX* simulations generated 536 successful cases with a cRMSD <of 2 Å. In the second round, after the clash score was turned on, the number of successful cases increased to 639; this is 19% higher than the first round, indicating that optimization of the clash score can further improve the success rate of MR. Moreover, if we define an absolute failure as a case in which a placement of the protein with a cRMSD of >2 Å has been generated in the simulations but has a lower XScore than the native placement and any outputted placement of the protein with a cRMSD of <2 Å (*i.e.* these cases will never succeed with the given scoring function and search space), the sum of the number of successes and absolute failures may be used as an indication of the convergence of the MR search. After the first two rounds, we found that the sum of the numbers of successes and absolute failures was equal to the number of decoys, indicating that the simulations had converged, because running additional rounds with the current settings would probably not improve the results.

Starting at the third round, occupancy optimization was turned on and the number of successful cases obtained by *MR-REX* was further increased from 639 to 681 (see Table 2). Here, we note that the accuracy of the residue-quality prediction, which was used to guide segment categorization during occupancy optimization, is still low, with a Matthews correlation coefficient (MCC) of 0.269 to the actual deviation of the residues in the benchmark data, where a residue is considered to be positive when its C^{α} atom deviates from the native by more than 2 Å. Such a substantial improvement in MR is probably not owing to the correct identification of inaccurate segments at the beginning, but rather is owing to the dynamic MC movements and selection process of the optimal occupancy in the REMC search process. Seven iterations have been performed in this round, where each starts from the ending placement of the previous iteration. In the first two iterations using occupancy optimization, 19 new successes were achieved beyond the first two rounds of F-match and clash optimizations. The efficiency of the implementation of further iterations gradually decreases and the last two iterations of occupancy optimization generated only seven additional successes, indicating that the results are nearing convergence at the end of six iterations. The number of successful cases in all iterations and rounds is summarized in Table 2.

Table 2
MR results of *MR-REX* at different rounds and iterations of simulations on 1303 3*DRobot* structure decoys.

Round No.	Search-scoring function	No. of iterations	Median cRMSD (Å)	⟨cRMSD⟩† (Å)	No. of successes‡
1	Match score	1	14.58	4.89	536
2	Match score + clash	1	2.49	4.25	639
3	Match score + clash + occupancy	1	2.15	4.17	641
		2	1.87	4.07	657
		3	1.74	4.02	667
		4	1.63	3.98	674
		5	1.61	3.96	676
		6	1.55	3.94	681
4	Match score + clash + occupancy + ML	1	1.52	3.90	699

† Average cRMSD of all decoys where cRMSD is set to 8 Å if it is higher than 8 Å. ‡ Successful MR is defined if the cRMSD is below 2 Å.

One additional iteration of *MR-REX*, lasting 6 h, was performed with the ML score. In this iteration 18 additional decoys were correctly placed by *MR-REX*.

3.6. What factors affect the success of MR?

In order to understand why the failure point of MR varied from protein to protein, in Supplementary Table S2 we list the worst TM-score structures for which MR succeeds, according to the criterion of a cRMSD of <2 Å, *versus* a number of possible factors that may affect the MR results, including the number of residues in the protein, the volume of the unit cell, the packing density of the unit cell in terms of the number of residues per $Å^3$, the shape of the protein, the number of symmetry mates in the unit cell, the number of HETATMs (excluding waters) per residue and the number of reflections used for MR. Here, the shape of the protein is specified by the elongation, *i.e.* $\varepsilon = (I_c^{1/2} - I_a^{1/2})/(I_a^{1/2} + I_b^{1/2} + I_c^{1/2})$, where I_c is the largest principal moment of inertia, I_a is the smallest principal moment of inertia and I_b is the second largest principal moment of inertia. The elongation of a sphere is 0 and the elongation of a line is 1.

Table 3 lists the correlation coefficients of the MR results with seven feature parameters as obtained by both single- and multi-variable linear regressions. There are many factors that

Table 3
The correlation coefficients between the model features and the failure point of MR in terms of TM-score for both *MR-REX* and *Phaser*.

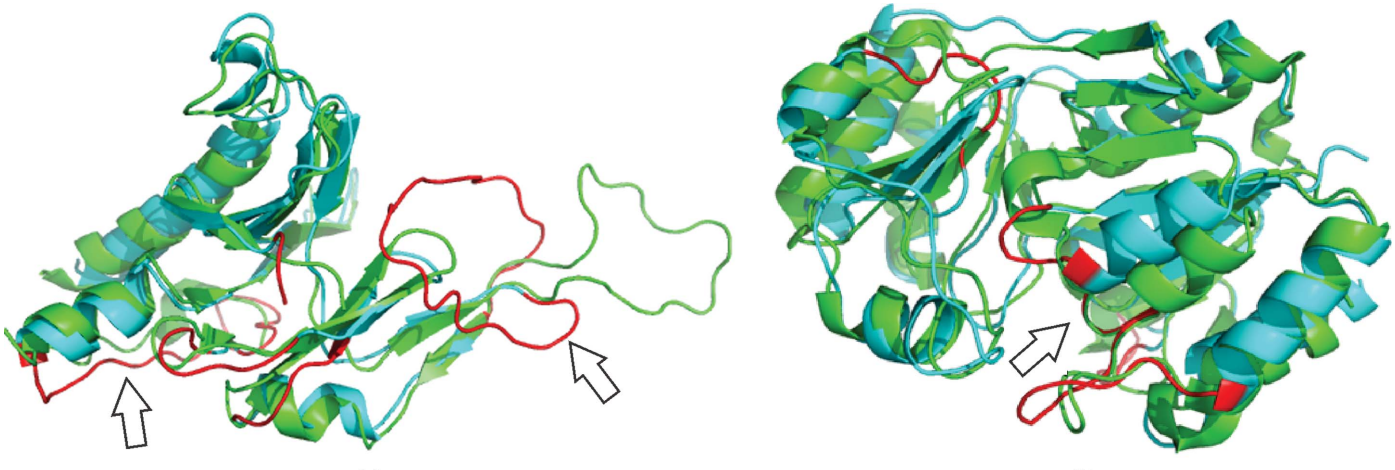
The error bars are estimated by repeatedly and randomly withholding half of the data points, calculating the correlations, recording the correlations and then calculating the standard deviations of the correlations. The error is one standard deviation. The highest correlation in each column is highlighted in bold. No. of residues is the number of residues in the asymmetric unit. Residues per Å^3 is the number of residues per Å^3 of the unit cell.

	MR-REX		*Phaser*	
	Single variable	Multi-variable	Single variable	Multi-variable
HETATM per residue	0.13 ± 0.02	0.03 ± 0.16	−0.03 ± 0.02	−0.02 ± 0.17
Volume of unit cell	0.23 ± 0.06	**0.77 ± 0.14**	0.05 ± 0.03	**0.31 ± 0.11**
No. of symmetry mates	−0.01 ± 0.05	−0.63 ± 0.13	−0.05 ± 0.02	−0.31 ± 0.17
Residues per Å^3	−0.19 ± 0.01	−0.03 ± 0.15	0.04 ± 0.02	0.12 ± 0.11
No. of reflections	0.11 ± 0.05	−0.38 ± 0.18	0.01 ± 0.03	0.06 ± 0.21
Elongation	**−0.38 ± 0.01**	−0.38 ± 0.13	0.16 ± 0.03	0.18 ± 0.17
No. of residues	0.23 ± 0.01	NA	**0.37 ± 0.02**	NA

affect the success or failure of MR and these may confound the results presented here. The multi-variable regression generally gives more meaningful results than the single-variable regression as the multi-variable linear regression reduces confounding effects; but since there are only 38 data points the error bars resulting from multi-variable linear regression are large. Multi-variable regression is not calculated using the number of residues in the protein as it is not an independent variable: it can be determined from the volume, the number of residues per Å^3 and the number of symmetry mates.

For the single-variable linear regression, the strongest correlation for *MR-REX* is with the elongation of the structure, although it is only a weak correlation with $r^2 = -0.38$. The more elongated the protein is, the more likely it is that *MR-REX* will succeed; this is understandable, as having an elongated protein should make it easier to find the correct orientation of the protein. It is therefore expected that *MR-REX* would find it easier to correctly orient proteins that were not approximately spherical in shape. We also found a weak but positive correlation between the elongation and the failure point of *Phaser* with $r^2 = 0.16$, *i.e.* the more elongated the protein is, the harder it would be for *Phaser* to succeed; this is probably because the scoring function used by *Phaser* is related to the Patterson function, which is expected to perform worse for elongated proteins as it is more difficult to separate intramolecular and intermolecular Patterson vectors for elongated proteins. The strongest correlation for *Phaser* is from the number of residues ($r^2 = 0.37$), indicating that *Phaser* has a higher success rate for smaller proteins; but the correlation for *MR-REX* is weaker with $r^2 = 0.23$.

Multi-variable linear regression reveals a strong positive correlation between the unit-cell volume and the failure point for *MR-REX* ($r^2 = 0.77$), and a relatively weaker correlation for *Phaser* ($r^2 = 0.31$). There is also a negative correlation between the number of symmetry mates in the unit cell and the failure point for *MR-REX* ($r^2 = -0.63$) and *Phaser* ($r^2 = -0.31$). Another factor which influences the success or failure of MR but that is not included in Table 3 is the existence of alternate placements of the protein which give similar electron densities as the native but are incorrect. This could occur for proteins that have approximate rotational symmetry. Supplementary Fig. S4(*a*) shows as an example the YmoB protein (PDB entry 2mn2), which has a four-helix bundle fold with an



(*a*) (*b*)

Figure 5
Representative examples in which either *MR-REX* or *Phaser* succeeds in MR. (*a*) A decoy structure of the human Paf1 protein (PDB entry 4m6t) that has an r.m.s.d. of 7.44 Å to the native, which was successfully placed by *MR-REX* with a cRMSD of 0.47 Å and a *phenix.autobuild* R_{free} of 0.27 but was not successfully placed by *Phaser*. (*b*) A decoy structure of the haloalkanoic acid dehalogenase enzyme (PDB entry 2rbk) that has an r.m.s.d. of 2.22 Å, which was successfully placed by *Phaser* but not by *MR-REX*. In both examples the decoy structure is shown in cyan and the native in green. The residues on the decoy that were removed by *MR-REX* in the occupancy optimization are colored red. In (*b*), the arrow marks the structural errors in the middle helix region that were not recognized by the occupancy-optimization algorithm of *MR-REX*.

approximate fourfold rotational axis. For every protein decoy, the MR solution with the highest electron-density correlation to the native, which represents an incorrect MR candidate solution, is found. Supplementary Fig. S4(*b*) presents another example, Rho GDP-dissociation inhibitor (PDB entry 2jhs), which has a β-sandwich fold, where the r.m.s.d. of the decoy structure is 4.4 Å from the native. One candidate solution has an electron-density correlation of 0.45 but with a cRMSD of 28.6 Å, while the correct solution generated but not selected by *MR-REX* has an electron-density correlation of 0.24 and a cRMSD of 0.17 Å. This protein has an approximate twofold rotational axis, and the incorrect placement of PDB entry 2jhs is rotated by 191° with respect to the native but superimposes on it well. In both examples the approximate rotational axis can create an alternative local minimum that can be confused for the correct placement of the protein model, resulting in a failure to solve the protein structure.

3.7. Case studies of the 3*DRobot* set

The test set showed complementary results for *MR-REX* and *Phaser*. Here, we present two typical examples in which either *MR-REX* or *Phaser* succeeds but the other fails in order to further examine the difference in the performance of the two methods.

Fig. 5(*a*) shows an example from the human Paf1 protein (PDB entry 4m6t) in which the initial search model has an r.m.s.d. of 7.4 Å to the native. This structure may seem to be of too low quality to be used for MR, but most of the structure deviations come from two loops (marked by arrows). If the 38 residues that deviate from the native by the greatest amount are removed, leaving 139 of the 177 residues, the r.m.s.d. decreases to 2.08 Å, indicating that the core region of the structure has a high accuracy. *MR-REX* was able to recognize and remove the residues in the two loops in the third iteration of the occupancy-optimization simulation, which resulted in a correct MR solution with a cRMSD of 0.47 Å and a *phenix.autobuild* $R_{\rm free}$ of 0.27. *Phaser* failed to correctly place this structure owing to the high r.m.s.d. However, when the regions found to be bad by *MR-REX* were removed from the model, *Phaser* was able to find the correct solution.

There are only four cases in which *MR-REX* failed but *Phaser* succeeded according to both of the criteria cRMSD < 2 Å and $R_{\rm free} < 0.4$. Fig. 5(*b*) shows one such example, which is from the haloalkanoic acid dehalogenase enzyme (PDB entry 2rbk). The r.m.s.d. of the decoy model is 2.2 Å to the native, for which *Phaser* achieved an MR solution with a cRMSD of 0.46 Å and an $R_{\rm free}$ of 0.28, while the *MR-REX* placement has a cRMSD of 18.27 Å and an $R_{\rm free}$ of 0.52. One reason for the failure is that the error estimation in *MR-REX* missed the major deviation of the models in the middle helix (marked by an arrow in Fig. 5*b*), although it correctly recognized several other high-variation residues. Secondly, there is an intermolecular clash between residue 138 of one copy and residue 261 of another copy in the best placement, which was skipped by *MR-REX* owing to the clash-score penalty. *Phaser* caught the solution since the clash was not severe enough for *Phaser* to rule it out. Despite the clash score overpenalizing the correct placement in this case, we found that the clash score significantly improved the overall performance of *MR-REX*. For another decoy of the same target that had an r.m.s.d. of 2.08 Å, however, *MR-REX* correctly placed the model with a cRMSD of 0.56 Å and an $R_{\rm free}$ of 0.27 (not shown).

Two of the templates from PDB entry 2il5, for which *Phaser* failed and *MR-REX* succeeded, with r.m.s.d.s of 3.25 and 4.44 Å, were pruned according to the segments predicted to be potentially inaccurate by *MR-REX* before the MR search, and *Phaser* was run on the pruned models. *Phaser* still failed to solve these two decoys, but succeeded when pruning the segments determined to be inaccurate by *MR-REX* after the MR search. Randy Read reported that he was able to solve these structures using *Phaser* after using automated model pruning. It is likely that an experienced user could get *Phaser* to succeed on the decoys for which it failed and *MR-REX* succeeded using pruning and alternative parameters. Nevertheless, the variety and complementarity provided by different tools are often helpful for non-experienced users through automated procedures such as online web servers.

3.8. MR results on homologous template structures

The training and testing of *MR-REX* have primarily been based on full-length models created by the 3*DRobot* program (Deng *et al.*, 2016). To examine the methods in the situation of homologous modeling, we created a second set of testing models using *LOMETS* (Wu & Zhang, 2007). *LOMETS* is a meta-threading method containing eight state-of-the-art threading programs which generates template models by matching the query sequence through the PDB library. Since multiple templates can be detected for the same sequence, a total of 320 template models are generated for the 38 test proteins, *i.e.* 8.4 templates on average for each protein, with a sequence identity ranging from 4 to 100%.

MR-REX and *Phaser* are run on each of the 320 templates, and the templates with the worst sequence identity for which MR is successful are recorded for each protein. Here, since the templates are taken from experimental structures, the *B* factor from the template structure is used in *MR-REX* when available. If we consider a solution with a cRMSD of <2 Å as being successful, the average point of failure is a sequence identity of 49% for *MR-REX* and 54% for *Phaser*; the difference is significant, with a *p*-value of 0.04. If we use the criterion of a *phenix.autobuild* $R_{\rm free}$ of <0.4 there is almost no difference between the two programs, which have failure-point sequence identities of 59 and 58%, respectively (Table 1).

However, this does not mean that the *MR-REX* and *Phaser* programs should necessarily be expected to fail for templates with sequence identities below 50–59%, as the sequence identity of the test templates is highly discontinuous owing to the limited number of templates. For example, for PDB entry 1r0u only the template with 100% sequence identity succeeds, where the template with the next lowest sequence identity has a sequence identity of 43% detected by *LOMETS*. If PDB entry 1r0u had a template with a sequence identity between

100 and 43% it might have succeeded and lowered the average point of failure.

For all 320 template models, *MR-REX* generated correct MR solutions for 125 cases, while *Phaser* did so in 118 cases, according to the criterion of a cRMSD of <2 Å. In 14 cases *MR-REX* correctly placed a decoy structure that *Phaser* was unable to place, while the reverse occurred in seven cases; this shows again that the two methods are complementary to each other, where a combination of the two programs results in 132 cases in which at least one of the programs succeeded. Similarly, if we use the criterion of $R_{\rm free}$ and TM-score of the *phenix.autobuild* models, *MR-REX* correctly placed 98 template models and *Phaser* did so for 100 models, where a combination results in 108 cases in which at least one of the programs succeeded. *phenix.autobuild* was not run on the output of the iterations of *MR-REX*, which used the ML score.

Finally, to examine the effect of different scoring functions, we tried to select MR solutions based on the *R* factor (2), the standard deviation (3), the Pearson correlation (4) and the maximum-likelihood (ML) function (Read, 2001*b*) from the MR models generated in the *MR-REX* searching simulations; this resulted in 114, 114, 112 and 123 successful cases, respectively, for which the model has a cRMSD of <2 Å. The result seems to suggest that the ML score is more sensitive to the correct MR solutions. However, we only used the ML score during the last two iterations of *MR-REX*, both of which lasted 6 h. The ML score is computationally expensive. During the 6 h of the last iteration with the ML score an average of 69 cycles were performed, whereas in the 2 h of the last iteration without the ML score 90 cycles were performed. Thus, *MR-REX* runs at about one fourth of the speed when using the ML score compared with when it is not used. Before the two iterations using the ML score during the MR search, *MR-REX* succeeded for 123 templates. Using the ML score during the REMC search made only a small impact, but this is probably owing to the small number of REMC cycles that were performed using the ML score, and if more cycles were run with the ML score it is possible that significantly more templates would be correctly placed.

3.9. Case studies of *LOMETS* set

There are three templates for which *MR-REX* fails according to both cRMSD and *phenix.autobuild*, but *Phaser* succeeds according to both criteria. The first case is PDB entry 1v05 with 96 residues, for which a *LOMETS* template from PDB entry 4m9p was used. This template has an r.m.s.d. of 1.56 Å to the native structure. In this case there are incorrect placements of the template generated by *MR-REX*, with no clashes and slightly lower *R* factors and ML TFs than the placement that maximizes electron-density correlation with the native. As the structure of the model becomes less native-like, the difference between the correct and random placements becomes smaller until there are random placements that give better scores than the correct placement. However, the correct solution found by *Phaser* had a TFZ of 7.2. Nevertheless, there is another template (PDB entry 4p3w) for PDB entry 1v05 identified by *LOMETS* with an r.m.s.d. of 2.35 Å to the native, which was correctly placed by *MR-REX* according to both cRMSD and *phenix.autobuild*.

The second template that *MR-REX* did not correctly place is PDB entry 5ig5, which is a template of target 3b7c. It has 20% sequence identity to the target and an r.m.s.d. of 3.14 Å to the native structure, where *Phaser* finds the correct solution with TFZ = 5.2 without clashes. Finally, *MR-REX* failed to place the template PDB entry 4ouj for target 3vwc that is 146 residues in length. The template has a sequence identity of 26% and an r.m.s.d. of 3.26 Å. The native placement could not be selected by any of the metrics used by *MR-REX*, where the final placement has a cRMSD of 20.4 Å. The TFZ of the solution found by *Phaser* was 4.9, with four clashes for this case. Overall, it seems that in these cases the correct placements of the templates could not be recognized by the *MR-REX* scoring function, including the ML score, although they were generated during the REMC simulations in some cases, while *Phaser* was still able to find a clear difference between the correct and incorrect placements. This might be owing to the anisotropy correction performed by *Phaser*, as well as some difference in the way that the effect of solvent is accounted for.

On the other hand, there are four templates for which the default setting of *Phaser* did not obtain the correct solution according to either criterion but *MR-REX* obtained the correct solution according to both criteria. In three of these cases *Phaser* did not output a solution. For PDB entry 2eed, for example, which is a template of PDB entry 1v05, the r.m.s.d. of the model to the native was 1.21 Å. *Phaser* found a solution with TFZ = 8.8, but rejected it based on the large number of clashes (seven clashes). The template was pruned according to the results from *MR-REX* after the MR search and *Phaser* was then rerun with the new model. The TFZ decreased to 7.8 and the number of clashes was still too great for *Phaser* to accept the solution. When sampling was made finer by setting the rotational sampling angle to 1° and the translational sampling distance to 0.5 Å, *Phaser* was finally able to correctly place the pruned template with TFZ = 8.4 and four clashes.

For PDB entry 4q2n, which is a template of PDB entry 3vqf, *Phaser* found a solution with TFZ = 6.3 but with too many clashes. The template was pruned according to the results from *MR-REX* after the MR search, and *Phaser* was rerun with the new model. The TFZ from *Phaser* increased slightly to 6.5, but the clash score was still too high and no structure was output by *Phaser*. When the sampling was made finer by setting the rotational sampling angle to 1° and the translational sampling distance to 0.5 Å *Phaser* correctly placed the pruned template with TFZ = 5.6 and four clashes. The sequence identity of the template was 22% and the r.m.s.d. to the native was 1.75 Å. The native protein had 85 residues. *MR-REX* found a clear separation between random placements and the native placement in terms of all of the metrics except for the clash score.

For another template of PDB entry 3vqf with an r.m.s.d. of 2.23 Å to the target 5jxb, *Phaser* found a solution with

TFZ = 6.0 but again with too many clashes. When the template was pruned according to the *MR-REX* results after the MR search, *Phaser* successfully placed the template. The TFZ decreased to 4.2 and the number of clashes decreased to four, which was accepted as a solution. Finally, *Phaser* output an incorrect solution for PDB entry 1h8m, which is one of the templates of PDB entry 3bw6. The TFZ was 5.0 and there were four clashes. After pruning the template according to the *MR-REX* results, the TFZ of the top solution increased to 5.8, the clashes were eliminated and the top solution was correct. Overall, among the four cases that failed with the default *Phaser* program, the program succeeded in two cases after the templates were pruned to match the *MR-REX* results; in the other two cases where *Phaser* failed after pruning it was found to succeed after making the sampling finer. It is likely that more successful cases could be identified with careful tuning of the parameters, which suggests the importance of template-model refinement and parameter optimizations.

4. Conclusions

We have developed a novel molecular-replacement tool, *MR-REX*, based on iterative replica-exchange Monte Carlo simulations. The major advantage of the MC-based pipeline is that it enables a cooperative six-dimensional-based translation and rotation search, which allows simultaneous steric clash and structural occupancy optimizations for an extensive conformational search.

MR-REX was mainly tested on a set of 1303 structure decoys of diverse accuracy created by *3DRobot* (Deng *et al.*, 2016) for 38 nonredundant proteins. *MR-REX* can generate correct MR solutions for 699 cases which have a crystallographic r.m.s.d. below 2 Å to the best placement obtained by maximizing the electron-density correlation between the model and the native protein. The average r.m.s.d. (or TM-score) for the worst decoy from which *MR-REX* can generate a correct MR placement is 4.3 Å (0.76) for the 38 proteins, demonstrating the ability of *MR-REX* to place low-accuracy structure models.

The results of *MR-REX* are comparable to (or slightly favorable compared with) those of the state-of-the-art MR tool *Phaser* (McCoy *et al.*, 2007), whereas the number of successful MR cases generated by *MR-REX* is 10% higher than the latter when applied to the same benchmark set. While the performance of the two programs is more comparable for the easy cases where the initial structure models are close to the native (for example an r.m.s.d. of <2 Å or a TM-score of >0.85), the major difference is in placing the poor-quality models. For decoys with a TM-score below 0.85 (or 0.80), for example, the number of successful cases obtained using *MR-REX* is 30% (or 49%) higher than that obtained by *Phaser*. As the accuracy of structure decoys approximately corresponds to the typical quality of distant homologous models, the data demonstrate the potential usefulness of *MR-REX* for exploiting low-resolution protein structure predictions.

The slightly greater success of *MR-REX* can be partly attributed to its introduction of a clash-scoring function into the MR simulations. When the clash score is not used as part of the scoring function during the MR search but is only used in filtering the final models, *MR-REX* is worse than *Phaser*, which also uses a clash score at the end. However, when the clash score is used to guide the MR search the number of successful cases is increased by 19%. In addition, *MR-REX* predicts inaccurate segments of proteins and optimizes the occupancies of these segments, which further increased the number of successes by ∼7%. Finally, using the ML score increased the number of successes by an additional 2.6%.

The median cRMSDs of *MR-REX* and *Phaser* are 1.52 and 2.53 Å, respectively, for the 1303 structure decoys. This difference in median cRMSD indicates that *MR-REX* generates more medium-range MR solutions with a reasonable cRMSD below 2 Å. Even though proteins with MR solutions in this range may not be solved by current structure-construction programs such as *phenix.autobuild*, which solved 77% of the cases correctly placed by *MR-REX*, it may be beneficial when more powerful structure-construction programs (DiMaio *et al.*, 2011) are developed in the future.

To mimic the situation of homologous modeling, we also tested *MR-REX* on a second set of 320 template models created by the meta-threading program *LOMETS* (Wu & Zhang, 2007). While the results of *MR-REX* are still comparable to those of *Phaser*, the difference between the two programs becomes smaller, which is probably owing to the fact that the *MR-REX* parameters have mainly been trained using the *3DRobot* decoys, which are full-length structure models, while *LOMETS* models are homologous models with gapped alignments. Nevertheless, the results of the two programs remain complementary, showing that a combination of both can increase the yield. Based on this decoy set, we also tested the power of different scoring functions in selecting the correct MR solutions; the results showed that the maximum-likelihood score outperforms the *R*-factor-based scores, which is consistent with observations made by other investigators (Read, 2001*b*; McCoy *et al.*, 2005).

Despite the advancements made by *MR-REX*, it should be mentioned that there are still cases where *Phaser* succeeded in MR but *MR-REX* failed. Moreover, *MR-REX* takes up to 20 h when using default parameters, which is much longer than *Phaser*. Nevertheless, given the importance of protein structure determination and the overall acceptable CPU range, the advance in performance is probably sufficient to demonstrate the worth of the time investment of running the method. While the method still has room for optimization, it represents an efficient tool complementary to the current state-of-the-art molecular-replacement methods.

Acknowledgements

We would like to thank Dr Randy Read for help in running the *Phaser* program. JJV and YZ conceived and designed the project, JJV developed the program and performed the data analysis, and JJV and YZ wrote the paper.

Funding information

This work was supported in part by the National Institute of General Medical Sciences (GM-083107 and GM-116960) and the National Science Foundation (DBI-1564756).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Baker, E. N., Anderson, B. F., Dobbs, A. J. & Dodson, E. J. (1995). *Acta Cryst.* D**51**, 282–289.
Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. (2012). *Acta Cryst.* D**68**, 1622–1631.
Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
Brünger, A. T. (1993). *Immunomethods*, **3**, 180–190.
Brünger, A. T. (1997). *Methods Enzymol.* **276**, 558–580.
Bunkóczi, G. & Read, R. J. (2011). *Acta Cryst.* D**67**, 303–312.
Bunkóczi, G., Wallner, B. & Read, R. J. (2015). *Structure*, **23**, 397–406.
Castellano, E. E., Oliva, G. & Navaza, J. (1992). *J. Appl. Cryst.* **25**, 281–284.
Chang, G. & Lewis, M. (1997). *Acta Cryst.* D**53**, 279–289.
Crowther, R. A. (1972). *The Molecular Replacement Method*, 1st ed. New York: Gordon & Breach.
Deng, H., Jia, Y. & Zhang, Y. (2016). *Bioinformatics*, **32**, 378–387.
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.
Drenth, J. (2007). *Principles of Protein X-ray Crystallography*, 3rd ed. New York: Springer.
Evans, P. & McCoy, A. (2008). *Acta Cryst.* D**64**, 1–10.
Frishman, D. & Argos, P. (1995). *Proteins*, **23**, 566–579.
Giorgetti, A., Raimondo, D., Miele, A. E. & Tramontano, A. (2005). *Bioinformatics*, **21**, Suppl. 2, ii72–ii76.
Glykos, N. M. & Kokkinidis, M. (2000). *Acta Cryst.* D**56**, 169–174.
Glykos, N. M. & Kokkinidis, M. (2001). *Acta Cryst.* D**57**, 1462–1473.
Grosse-Kunstleve, R. W. & Adams, P. D. (2001). *Acta Cryst.* D**57**, 1390–1396.
Huang, Y. J., Mao, B., Aramini, J. M. & Montelione, G. T. (2014). *Proteins*, **82**, Suppl. 2, 43–56.
Jamrog, D. C., Zhang, Y. & Phillips, G. N. (2003). *Acta Cryst.* D**59**, 304–314.
Jogl, G., Tao, X., Xu, Y. & Tong, L. (2001). *Acta Cryst.* D**57**, 1127–1134.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Liu, Q., Weaver, A. J., Xiang, T., Thiel, D. J. & Hao, Q. (2003). *Acta Cryst.* D**59**, 1016–1019.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
Mariani, V., Kiefer, F., Schmidt, T., Haas, J. & Schwede, T. (2011). *Proteins*, **79**, Suppl. 10, 37–58.
McCoy, A. J. (2007). *Acta Cryst.* D**63**, 32–41.
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.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
McCoy, A. J., Oeffner, R. D., Wrobel, A. G., Ojala, J. R. M., Tryggvason, K., Lohkamp, B. & Read, R. J. (2017). *Proc. Natl Acad. Sci. USA*, **114**, 3637–3641.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). *J. Chem. Phys.* **21**, 1087–1092.
Navaza, J. (1987). *Acta Cryst.* A**43**, 645–653.
Navaza, J. (1990). *Acta Cryst.* A**46**, 619–620.
Navaza, J. (1993). *Acta Cryst.* D**49**, 588–591.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Read, R. J. (2001*a*). *Acta Cryst.* D**57**, 1373–1382.
Read, R. J. (2001*b*). *Acta Cryst.* D**57**, 1373–1382.
Rez, D., Rez, P. & Grant, I. (1994). *Acta Cryst.* A**50**, 481–497.
Rodríguez, D. D., Grosse, C., Himmel, S., González, C., de Ilarduya, I. M., Becker, S., Sheldrick, G. M. & Usón, I. (2009). *Nature Methods*, **6**, 651–653.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Sammito, M., Meindl, K., de Ilarduya, I. M., Millán, C., Artola-Recolons, C., Hermoso, J. A. & Usón, I. (2014). *FEBS J.* **281**, 4029–4045.
Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. (2004). *Acta Cryst.* D**60**, 1229–1236.
Shrestha, R. & Zhang, K. Y. J. (2015). *Acta Cryst.* D**71**, 304–312.
Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.
Swendsen, R. H. & Wang, J.-S. (1986). *Phys. Rev. Lett.* **57**, 2607–2609.
Terwilliger, T. (2004). *J. Synchrotron Rad.* **11**, 49–52.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Tong, L. (1996). *Acta Cryst.* A**52**, 782–784.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wang, G. & Dunbrack, R. L. (2003). *Bioinformatics*, **19**, 1589–1591.
Wang, Y., Virtanen, J., Xue, Z., Tesmer, J. J. G. & Zhang, Y. (2016). *Acta Cryst.* D**72**, 616–628.
Wang, Y., Virtanen, J., Xue, Z. & Zhang, Y. (2017). *Nucleic Acids Res.* **45**, W429–W434.
Weiser, J., Shenkin, P. S. & Still, W. C. (1999). *J. Comput. Chem.* **20**, 217–230.
Wu, S. & Zhang, Y. (2007). *Nucleic Acids Res.* **35**, 3375–3382.
Xu, J. & Zhang, Y. (2010). *Bioinformatics*, **26**, 889–895.
Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. & Zhang, Y. (2015). *Nature Methods*, **12**, 7–8.
Zhang, Y. (2009). *Curr. Opin. Struct. Biol.* **19**, 145–155.
Zhang, Y. & Skolnick, J. (2004*a*). *J. Comput. Chem.* **25**, 865–871.
Zhang, Y. & Skolnick, J. (2004*b*). *Proteins*, **57**, 702–710.