BIOINFORMATICS

Vol. 21 Suppl. 2 2005, pages ii72–ii76
doi:10.1093/bioinformatics/bti1112

Protein Structure and Function

Evaluating the usefulness of protein structure models for molecular replacement

Alejandro Giorgetti[1,†], Domenico Raimondo[1,†], Adriana Erica Miele[1] and Anna Tramontano[1,2,*]

[1]Department of Biochemical Sciences and [2]Istituto Pasteur—Fondazione Cenci Bolognetti, University of Rome 'La Sapienza', 00185 Rome, Italy

ABSTRACT

Motivation: We investigate the relationship between the quality of models of protein structure and their usefulness as search models in molecular replacement, a widely used method to experimentally determine protein structures by X-ray crystallography.

Results: We used the available models submitted to the Critical Assessment of Techniques for Protein Structure Prediction to verify in which cases they can be automatically used as search templates for molecular replacement. Our results show that there is a correlation between the quality of the models and their suitability for molecular replacement but that the traditional method of relying on sequence identity between the model and the template used to build it is not diagnostic for the success of the procedure.

Availability: Additional data are available at http://cassandra.bio.uniroma1.it/mr-results-casp.html

Contact: anna.tramontano@uniroma1.it

1 INTRODUCTION

X-ray crystallography is the most used and most effective technique for obtaining the structure of proteins and protein complexes. As of today, the X-ray structures of tens of thousand of proteins are deposited in the Protein Structure Data Base (PDB) (Bernstein *et al.*, 1977) and this number is also continuously increasing owing to the efforts of structural genomics projects aimed at providing representative examples of the protein structural space.

In an X-ray diffraction experiment, crystals of the protein of interest are irradiated with X-rays, and interference effects give rise to a characteristic diffraction pattern. The electron density of the protein, i.e. the positions of the protein atoms, determines the diffraction pattern of the crystal, i.e. the magnitudes and phases of the X-ray diffraction waves, and vice versa, through a Fourier transform function. In practice

$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl}\vec{F}_{hkl} = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}F(h,k,l)\mathrm{e}^{-2\pi i(hx+ky+lz)}$$

where $\rho(x,y,z)$ is the electron density at position (x,y,z), $\vec{F}(h,k,l)$ is the vector describing the diffracted waves in terms of their amplitudes $F(h,k,l)$ and phases (the exponential complex term). The electron density at each point depends upon a sum of all of the amplitudes and phases of each reflection.

In a diffraction experiment, a crystal is irradiated with a particular X-ray wavelength and the resulting diffracted waves are collected on physical or electronic devices. However, in this passage from 3D to 2D all the information on the phases is lost, and this is one of the fundamental problems of structural science.

By and large, there are three approaches to solving the phase problem: direct methods, interference-based methods and molecular replacement (MR) methods. The first type of method consists of using all the possible values for the phases in the Fourier transform equation until an interpretable electron density is found; this is feasible only for small proteins. The second method consists of measuring the effect of adding a known set of waves to the system. This can be done either by adding one or more high atomic number atoms (the so-called heavy atoms) to the crystal, thus solving the phase problem by multiple isomorphous replacement, or by substituting some atoms with specific heavy atoms able to scatter anomalously the incident wave, thus solving the phase by multiwavelength anomalous dispersion. In other words collecting the data from one single crystal at three different incident wavelengths is equivalent to collecting three datasets with three different heavy atoms. Finally, there is MR, which is based on the observation that, if the atomic coordinates x, y, and z are known, the structure factors can be computed back by inverse Fourier transform. This implies that one can identify the correct model among a large number of possible solutions by comparing the observed and computed intensities and is indeed the procedure that has been used for solving the structure of many small molecules, but it is unfeasible for proteins where a large number of atoms makes the exploration of the conformational space simply impossible.

If the relative positions of the atoms in the protein structure are known (or can be estimated) the problem is reduced to finding the position of the whole molecule in the unit cell rather than that of each of its atoms. The prior knowledge of a protein structure thereby simplifies the solution of a different crystal form of the same molecule. In some cases, the structure of a homologous protein or a model of the target protein can be sufficient to approximate the relative position of the atoms in the structure and allow the structure factors to be computed. This strategy is known under the name 'molecular

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author 2005. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

replacement' and was introduced by Michael G. Rossmann and David M. Blow in 1962 (Rossmann and Blow, 1962).

Recently there has been an explosive growth in the number of protein structures solved using this technique, and automated packages can make the application quite straightforward. Progress has been made in using models from NMR or modelling (Turkenburg and Dodson, 1996; Wilmanns and Nilges, 1996; Jones, 2001).

The predicted structure of a protein can be used in MR when the computational model is sufficiently accurate for a reasonably large fraction of the structure in the crystal. A generally accepted rule of thumb is that MR is effective if the model is reasonably complete and shares at least 40–50% sequence identity with the unknown structure. On the other hand, techniques for predicting protein structures have matured in recent years, most of them taking advantage of the increased availability of sequences and structures of homologous members of a protein family; thus sequence identity is not necessarily an appropriate parameter to estimate the likelihood that a given model is suitable as an MR search template.

At present, a clear rule relating the quality of a model and its suitability for MR is lacking, and filling this gap is the aim of the present analysis.

It should be mentioned that the ability of the crystallographer, several rounds of trial and error and manual intervention in the intermediate steps of the process can also increase the rate of success of MR in very difficult cases. For example, it has been shown that in some cases, screening for MR solutions with a large number of models might produce suitable solutions (Claude *et al*., 2004) and that modelling can be used to improve the initial model after initial MR solutions have been found (Collaborative Computational Project Number 4, 1994; Potterton *et al*., 2003). Here we use only completely automatic procedures without manual intervention in any of the steps; therefore, our estimate represents the 'safe' limit above which a model can be used for MR, but it should be kept in mind that models of lower quality might be successful in specific cases.

2 METHODS

We used the collection of models submitted to the Critical Assessment of Techniques for Protein Structure Prediction (CASP) (Moult *et al*., 1995, 2001, 1997, 2001, 2003). In this assessment crystallographers and NMR spectroscopists are asked to make available the sequences of proteins whose structure is likely to be solved in the near future and predictors are asked to submit 3D models for the proteins before their experimental structure is made available. Subsequently, the experimental structures and the models are compared and the latter assessed in terms of their ability to reproduce the native protein structure.

The CASP website (http://predictioncenter.llnl.gov) contains, for each target structure, hundreds of models together with values reflecting their structural similarity to the target structure. The most relevant parameters are the root mean square deviation (RMSD) between corresponding atoms of the target and model structures and the GDT-TS value. The latter is a distance-based measure defined as

1/4 [(fraction of Cα atoms within a 1 Å distance) + (fraction of Cα atoms within a 2 Å distance) + (fraction of Cα atoms within a 4 Å distance) + (fraction of Cα atoms within an 8 Å distance)].

We selected target proteins for which the structure factors (i.e. the collected diffraction intensities) have been deposited in the PDB database and ran an automatic procedure for MR using a set of their models as search templates.

The MR procedure was run in a completely automatic fashion using two of the most popular programs which perform automated searches using similar algorithms: MolRep (Vagin and Teplyakov, 1997) and AMoRe (Navaza, 1994), both included in the CCP4 suite (Collaborative Computational Project Number 4, 1994). The procedure consisted of the following steps:

1. Each template (including its side chains) was given to MolRep to perform 10 rotational searches to orient it into the experimental data and 10 translational searches for each rotational solution.
2. The best solution was converted into a poly-Alanine model, then given as input to AMoRe to perform 20 cycles of automatic solution searching. When the first solution did not allow the map to be reconstructed, subsequent ones were tried. In no case the subsequent attempts were successful when the first was not.
3. The best solution from AMoRe was given as input to Refmac5 (version 5.0.32, Murshudov *et al*., 1997) to perform 10 cycles of restrained refinement without prior phase information and the default weighting matrix (0.3).
4. Maps were generated and displayed for inspection.
5. Structure building was performed automatically using the program ArpWarp, version 6.1.1 (Morris *et al*., 2002), as implemented in the CCP4 suite.

All the calculations were run through the graphical interface CCP4i of the CCP4 suite (Collaborative Computational Project Number 4, 1994) on Linux machines. Although other programs are available for solving structures by MR, we chose those most widely used by crystallographers.

For each MR experiment we monitored the following data:

1. The residual factor or agreement factor R defined as

$$R\text{-factor} = \frac{\sum_{h,k,l} \| F_{\text{obs}}(h,k,l) | - | F_{\text{calc}}(h,k,l) \|}{\sum_{h,k,l} |F_{\text{obs}}(h,k,l)|}$$

   where F_{obs} and F_{calc} are the observed structure factors and the structure factors computed on the basis of the molecular template, respectively. We computed the R-factor at the beginning ($R\text{-factor}_{\text{i}}$) and at the end ($R\text{-factor}_{\text{f}}$) of the Refmac procedure (see additional data).
2. R_{free}, i.e. the R-factor calculated on a fixed fraction (5%) of the experimental reflections not used in the refinement of the structure, therefore not biased by the procedure (see additional data).
3. The correlation coefficient:

$$C = \frac{\sum \left(|F_{\text{obs}}(h,k,l)|^2 - \overline{|F_{\text{obs}}(h,k,l)^2|}\right)\left(|F_{\text{calc}}(h,k,l)|^2 - \overline{|F_{\text{calc}}(h,k,l)|^2}\right)}{\sqrt{\sum \left(|F_{\text{obs}}(h,k,l)|^2 - \overline{|F_{\text{obs}}(h,k,l)^2|}\right)^2 \sum \left(|F_{\text{calc}}(h,k,l)|^2 - \overline{|F_{\text{calc}}(h,k,l)|^2}\right)^2}}$$

   where all sums are over all the reflections in the reciprocal lattice (h,k,l). We computed the value of C at the beginning (C_{i}) and at the end (C_{f}) of the Refmac procedure (see additional data).
4. C_{free}, i.e. the correlation coefficient calculated on the same fraction of reflections as for R_{free} (see additional data).

When the procedure was able to automatically build an experimental structure of the protein we also computed (Table 1)

5. The fraction of built amino acids.
6. The GDT-TS between the automatically built structure and the deposited experimental structure obtained using the LGA package (Zemla, 2003).
7. The RMSD between the automatically built structure and the deposited experimental structure.
8. The GDT-TS between the automatically built structure and the CASP model used in the MR procedure.
9. The RMSD between the automatically built structure and the CASP model used in the MR procedure.

3 RESULTS

Table 1 shows the details of the results of the MR procedure for 34 models covering 7 of the CASP targets. The resolution of the

Table 1. Results of the MR procedure

CASPm id	GDT CASPm	RMSD CASPm	%Id	%Built	GDT MRmX	RMSD MRmX	GDT MRmCASP	RMSD MRmCASP	Fold
Target T0143 PDBid 1QY6 resolution 1.9 Å									
TS453	83.8	3.36	25.89	84.72	100	0.14	89.34	1.27	All β
TS084	82.18	2.79		94.44	94.67	0.3	84.8	1.32	
TS425	82.06	3.62		25.5	100	0.27	100	0.58	
TS427	81.94	3.27		93.5	99.38	0.4	84.95	1.27	
TS093	81.71	3.55		79.16	79.53	0.15	67.65	1.38	
TS028	81.6	2.55		93.51	81.28	1.05	68.78	1.7	
Target T0153 PDBid 1MQ7 resolution 1.9 Å									
TS028	88.25	2.57	34.38	97.7	100	0.13	89.42	1.16	All β
TS299	86.94	1.3		88.3	72.02	0.72	66.9	1.21	
TS329	86.75	1.29		76.74	100	0.21	85.1	0.92	
TS020	85.63	1.34		90.6	72.2	0.4	66.9	1.21	
TS471	84.7	5.64		90.6	100	0.13	90.3	1.27	
TS169	83.39	1.44		0	—	—	—	/	
Target T0182 PDBid 1O0X resolution 1.9 Å									
TS329	94.18	1.28	42.34	99.19	100	0.12	93.62	1.02	$\alpha + \beta$
TS112	93.28	1.19		99.19	99.58	0.17	93.61	1.03	
TS513	92.77	1.32		99.19	100	0.11	92.34	0.97	
TS067	92.67	1.27		99.19	99.58	0.17	92.68	0.97	
TS029	91.67	1.36		99.19	100	0.12	91.7	1.06	
TS008	85.54	1.68		99.19	99.58	0.17	84.72	1.31	
TS529	77.71	1.8		0	—	—	—	/	
Target T0233 PDBid 1VQU resolution 1.85 Å									
TS450	85.92	1.81	42.5	39.13/54.4[a]	56.85/44.96[a]	0.68/1.14[a]	55.93/41.35[a]	0.86/0.74[a]	All α
TS011	84.72	1.85		33.14/23.05[a]	58.9/40.2[a]	0.85/1.05[a]	57.8/39.01[a]	1.17/0.64[a]	
TS176	83.43	1.93		—	—	—	—	/	
Target T0246 PDBid 1VLC resolution 1.9 Å									
TS319	89.05	1.29	57.1	80.28	94.55	0.7	85.54	1.17	α/β
TS532	88.56	1.4		75.14	66.9	0.36	74.61	1.05	
TS289	86.8	1.43		94.28	74.39	1.34	67.35	1.6	
TS100	83.05	1.69		72.87	99.61	0.2	86.12	1.37	
TS096	82.77	1.99		66.57	75	0.17	67.13	1.12	
TS561	81.85	1.79		—	—	—	—	/	
TS092	81.36	1.93		—	—	—	—	—	
TS079	81.14	2.25		—	—	—	—	—	
TS472	78.75	1.97		—	—	—	—	—	
Target T0274 PDBid 1WGB resolution 2.0 Å									
TS450	80.93	3.59	23.45	—	—	—	—	—	All β
Target T0275 PDBid 1WYG resolution 2.1 Å									
TS591	81.11	2.88	28.57	—	—	—	—	—	α/β
TS166	71	2.77		—	—	—	—	—	

For each target, the first line reports the target id in the CASP database, its PDB code and the resolution of the X-ray structure. The first column reports the identification code of the model (CASPm id). The next columns show the GDT-TS and the RMSD values between the model and the target structure (GDT CASPm and RMSD CASPm), the percentage of identical amino acids between the CASP target structure and the best structural template available in the 'core', defined as the regions where corresponding atoms of the model and template do not deviate by >5 Å (%Id). The number of amino acids built by the automatic procedure is reported in the column labelled '%Built'. We also list the GDT-TS and RMSD values for the superposition of the model built by the MR procedure on the deposited structure (GDT MRmX and RMSD MRmX) and of the model built by the MR procedure on the CASP model (GDT MRmCASP and RMSD MRmCASP). The last column lists the fold type of the target (Fold).
[a]T0233 is a homodimer and the results are shown for both monomers.

experimental data ranged from 2.1 to 1.85 Å. Successful cases are those where a substantial fraction of the structure has been built obtaining a model closer to the experimental structure than to the CASP model initially used in MR.

The case of the T0233 structure is peculiar. This protein is a homodimer and we were able only to partially reconstruct each of the monomers (Table 1). However, we were unable to build a significant portion of the structure even using the deposited experimental coordinates (from which several amino acids are missing).

The targets shown here have at least one CASP model with GDT-TS >80 (which implies that most of the main chain structure of the protein has been reasonably modelled). They all belong to the comparative model category; i.e. there is at least one homologous protein of known structure that could be used as template for building the model. Their sequence identity with the best template available ranges between 23 and 57%, and they belong to different fold classes.

In all but two cases, at least one CASP model was successful as a search template, indicating that the above-mentioned threshold of

40–50% sequence identity as a minimum requirement for MR is overestimated. Our data show that models built on the basis of a 30% sequence identity are already sufficiently accurate to be used in the MR procedure, and in some cases models based on sequence identity as low as 25% can be successfully used. Interestingly, in all but one case, the best available structural template was not successful in reconstructing any portion of the experimental structure. For T0153 alone, the best template allowed 14% of the structure to be reconstructed. Note that the best model for this target allowed the reconstruction of 97% of the experimental coordinates.

In CASP experiments, both the RMSD between the model and the structure and the GDT-TS values defined above are used as parameters for evaluating the quality of a model. Interestingly, when different CASP models are available for the same protein structure and only one or some of them work as search models, the successful ones are not always those with a lower RMSD, but in all cases they are those with higher GDT-TS. The RMSD value, owing to its quadratic form, is more sensitive to the extent of local deviations of the models from the structure, whereas GDT-TS is indicative of the percentage of correctly predicted structure and is less sensitive to larger errors localized in a few regions of the model.

Our data point to the fact that the RMSD between the model and the experimental structure is not the correct measure to use for defining the usefulness of a model for applications such as MR and, therefore, to evaluate the quality of a model. In this specific application of modelling, what really counts is the overall ability of the model to produce interpretable electron density maps of the protein, which helps in building most of it, rather than the details of the less well predicted parts.

In all cases, a GDT-TS >84 is sufficient to guarantee the success of the procedure regardless of the sequence identity between the target and template structure, of the method used for producing the model and of the structural class of the protein under examination. In our automatic procedure, models with GDT-TS <80 are never successful. For models of intermediate quality, the results vary. Most of the time a large fraction of the structure can be automatically built with respectable quality, and it is likely that, in these cases, more iterations and, most of all, manual intervention can lead to success.

Another important observation is that even limited improvements in the quality of a model can be instrumental in the success of an MR experiment. The case of the CASP models built for Target T0246 is instructive. This is an 'easy' target, sharing 57% sequence identity with its best template. Of the nine CASP models that we used for the experiment, five were successful and four were not. The former are of higher quality, as can be seen from their GDT-TS values (ranging from 89 to 83), but the latter are only slightly worse, with GDT-TS values between 82 and 79 (Fig. 1). The same observation holds for targets T0152 and T0153, indicating that even a minor improvement in the quality of the final model can be relevant for its ability to be used as a search model in MR. This observation can explain why it has been so difficult so far to predict beforehand when a model can be successfully used in MR solely on the basis of the sequence identity between a model and the structural template used to build it.

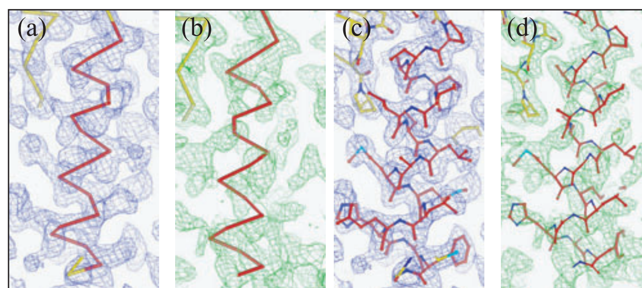

Fig. 1. Example of the results of the MR procedure using CASP models of different quality as search models in the MR procedure. (**a**) Electron density map obtained using the structure factors for target T0246 (PDB ID: 1vlc) phased using the CASP model TS319 (GDT-TS between the CASP model and the experimental structure = 89.05). The automatic procedure described in 'Methods' was used to obtain the model shown in the figure as a Cα trace. (**b**) Electron density map obtained using the structure factors for target T0246 (PDB ID: 1vlc) phased using the CASP model TS472 (GDT-TS between the CASP model and the experimental structure = 78.75). The automatic procedure could not build any model in the map. The model obtained using the map in (a) is shown superimposed on the density for reference only. Parts (**c**) and (**d**) show the same maps as in (a) and (b), respectively, including all the heavy atoms of the built model.

4 DISCUSSION AND CONCLUSIONS

MR is a very cost-effective method for solving the 3D structure of a protein by X-ray crystallography. It is expected to play an increasing role in the phasing of protein X-ray diffraction data, given the ever increasing number of proteins experimentally solved and deposited. However, the MR approach requires the availability of a reasonably good and complete model, and we have shown here that there is a clear relationship between the quality of a model and its suitability as a search model for MR experiments.

The large efforts in protein structure prediction might be very useful in pushing the limits of structure solving using MR. Models of proteins can be powerful tools for structural genomics projects and are expected to reduce the need for expensive and time-consuming phasing experiments. An increased usefulness of predicted structures as MR search models would have a substantial impact on our ability to cover the protein structural space. However, as we show here, minor differences in the quality of the model can make a substantial difference in the outcome. This points to the fact that more efforts should be devoted to improving the initial model, since even minor improvements can be important for practical applications such as the one discussed here. This issue has been addressed in the recent CASP6 experiment, where the community agreed that in the future more importance should be given to the details of the produced models (Cozzetto *et al*., 2005; Valencia, 2005).

The results shown here demonstrate that this is not only an intellectual and methodological issue; it also has important practical applications. We plan to run this analysis continuously on different datasets of models and to evaluate whether the rate of success can be improved by using combinations of models or limiting the search model to secondary structure elements or to the core of the predicted structure.

The possible use of a model is directly related to its quality, and more efforts should be devoted to precisely assessing the requirements that any given application poses on the quality of the models: even marginal improvements in prediction methods can be instrumental for important applications.

In a real setting, the results of the comparison of the model with the experimental structure are not available. It would be very useful

to be able to identify beforehand features of the model that correlate with its ability to be used as an MR search model. Unfortunately, the available quality evaluation measures, such as those provided by Verify3d (Eisenberg *et al*., 1997), Whatcheck (Hooft *et al*., 1996) and Procheck (Laskowski *et al*., 1993) do not correlate with the ability of the models to reconstruct the experimental structure (data not shown). Nevertheless, our results do show that it is advisable to attempt MR with protein models, even when the sequence identity with their best templates is low. More important, models can be successful also in cases when the structural templates used to build them are not.

We are currently investigating the possibility of using sets of different superimposed models in the MR search. Preliminary results indicate that ensembles obtained by superimposing both successful and unsuccessful models can be used to reconstruct the experimental structure in a number of cases. Such a strategy would clearly speed up the whole process, and a study aimed at defining its range of applicability is ongoing.

ACKNOWLEDGEMENTS

This work was supported by Istituto Pasteur—Fondazione Cenci Bolognetti and by the Bio-Sapiens Network of Excellence funded by the European Commission FP6 Programme, contract number LHSG-CT-203-503265. We are grateful to Dr Domenico Cozzetto for helpful discussions.

Conflict of Interest: none declared.

REFERENCES

Bernstein,F.C. *et al*. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur. J. Biochem.*, **80**, 319–324.

Claude,J. *et al*. (2004) CaspR: a web server for automated molecular replacement using homology modelling. *Nucleic Acids Res.*, **32**, W606–W609.

Collaborative Computational Project Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst.*, **D50**, 760–763.

Cozzetto,D. *et al*. (2005) Ten years of predictions . . . and counting. *FEBS J.*, **272**, 1–2.

Eisenberg,D. *et al*. (1997) Verify3d: assessment of protein models with three-dimensional profiles. *Methods Enzymol*., **277**, 396–404.

Hooft,R.W.W. *et al*. (1996) Errors in protein structures. *Nature*, **381**, 272–272.

Jones,D.T. (2001) Evaluating the potential of using fold-recognition models for molecular replacement. *Acta Crystallogr.*, **D57**, 1428–1434.

Laskowski,R.A. *et al*. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Morris,R.J. *et al*. (2002) ARP/wARP's model building algorithms. I. The main chain. *Acta Crystallogr.*, **D58**, 968–975.

Moult,J. *et al*. (1995) A large-scale experiment to assess protein structure prediction methods. *Proteins*, **23**, ii–v.

Moult,J. *et al*. (1997) Critical assessment of methods of protein structure prediction (CASP): round II. *Proteins*, (suppl. 1), 2–6.

Moult,J. *et al*. (2001) Critical assessment of methods of protein structure prediction (CASP): round III. *Proteins*, (suppl. 3), 2–6.

Moult,J. *et al*. (2001) Critical assessment of methods of protein structure prediction (CASP): round IV. *Proteins*, (suppl. 5), 2–6.

Moult,J. *et al*. (2003) Critical assessment of methods of protein structure prediction (CASP)-round V. *Proteins*, **53** (suppl. 6), 334–339.

Murshudov,G.N. *et al*. (1997) Refinement of macromolecular structure by maximum likelyhood method. *Acta Crystallogr.*, **A50**, 1131–1137.

Navaza,J. (1994) AMoRe: and automatic package for molecular replacement. *Acta Crystallogr.*, **D53**, 240–255.

Potterton,E. *et al*. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr.*, **D59**, 1131–1137.

Rossmann,M.G. and Blow,D.M. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.*, **15**, 24–31.

Turkenburg,J.P. and Dodson,E.J. (1996) Modern developments in molecular replacement. *Curr. Opin. Struct. Biol.*, **6**, 604–610.

Vagin,A.A. and Teplyakov,A. (1997) MOLREP: an automated program for molecular replacement. *Appl. Crystallogr.*, **30**, 1022–1025.

Valencia,A. (2005) Protein refinement: a new challenge for CASP in its 10th anniversary. *Bioinformatics*, **21**, 277.

Wilmanns,M. and Nilges,M. (1996) Molecular replacement with NMR models using distance-derived pseudo *B* factors. *Acta Crystallogr.*, **D52**, 937–982.

Zemla,A. (2003) LGA: a method for finding 3D similarities in protein structures. *Nucleic Acids Res.*, **31**, 3370–3374.