Title	Crystallographic phasing with NMR models: an envelope approach
Author(s)	ZHANG, W; Zhang, T; Zhang, H; Hao, Q
Citation	Acta Crystallographica Section D Biological Crystallography, 2014, v. 70 n. 7, p. 1977-1982
Issued Date	2014
URL	http://hdl.handle.net/10722/199211
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Crystallographic phasing with NMR models: an envelope approach

Weizhe Zhang,[a] Tao Zhang,[b] Hongmin Zhang[a] and Quan Hao[a]*

[a]Department of Physiology, University of Hong Kong, Hong Kong, People's Republic of China, and [b]Institute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China

Correspondence e-mail: qhao@hku.hk

X-ray crystallography and NMR are complementary tools in structural biology. However, it is often difficult to use NMR structures as search models in molecular replacement (MR) to phase crystallographic data. In this study, a new approach is reported utilizing a molecular envelope of NMR structures for MR phasing with the program *FSEARCH* at low resolution (about 6 Å). Several targets with both crystallographic and NMR structures available have been tested. *FSEARCH* was able to find the correct translation and orientation of the search model in the crystallographic unit cell, while conventional MR procedures were unsuccessful.

1. Introduction

One of the most critical steps in X-ray crystallography is to determine the phases of the diffraction data, which can be a bottleneck in the use of this technology and is called the phase problem. Molecular replacement (Rossmann, 1990; Rossmann & Blow, 1962) is a rather effective and economical method of solving the phase problem without the need to prepare heavy-atom or selenomethionine derivatives as is required in the isomorphous replacement method (Blow & Rossmann, 1961; Perutz, 1956) and the anomalous dispersion method (Pannu & Read, 2004; Hendrickson, 1991; Pähler *et al.*, 1990). In general, molecular replacement can be straightforward if the search model shares at least 30% sequence identity with the unknown structure. A common notion in molecular replacement is that a correct solution requires the root-mean-squared deviation (r.m.s.d.) of the C^α atoms between the search model and the target structure to be no greater than 1.5 Å. Hence, a good starting model is a crucial part of a successful molecular-replacement process.

Traditionally, molecular replacement is performed using the Patterson-based approach (Rossmann & Blow, 1962), which divides the vectors in the Patterson map into intermolecular vectors and intramolecular vectors, thus enabling separation of the rotation search and translation search. The Patterson function is essentially the Fourier transform of the intensities rather than the structure factors. In applying the Patterson method, in addition to the conventional use of atomic structures as search models, cases of molecular replacement using envelopes have also been reported (Urzhumtsev & Podjarny, 1995). In these cases, an artificial reduction in resolution could improve the accuracy of the target function; this was first studied by Urzhumtsev & Podjarny (1995) and later utilized in the program *SoMoRe* (Jamrog *et al.*, 2003).

In structural biology, nuclear magnetic resonance (NMR) is a powerful tool complementary to X-ray crystallography which is able to provide structural information in solution and

in the solid state. The number of structures solved by NMR has increased rapidly, which makes a large pool of resources available from which crystallographers can choose molecular-replacement (MR) search models. However, using NMR structures as search models to solve a crystal structure by MR is not always successful, even if the NMR structure is of the same protein. There is sometimes large global structural disagreement between X-ray and NMR structures, mainly owing to the fact that solution NMR data represent molecules in solution while X-ray diffraction data reflect an average over molecules arranged in a periodic crystal lattice.

In 1987, Brünger and coworkers first showed that information derived from solution NMR structures can be employed in solving crystal structures by MR using the known structure of crambin (Brünger *et al.*, 1987). Since then, several successful cases have been reported of using NMR structures

as MR search models. Chen and coworkers have summarized a general protocol of how to apply MR using NMR models by studying a number of test cases (Chen *et al.*, 2000; Chen, 2001). Discussions regarding how to utilize NMR models in molecular replacement have mainly focused on the preparation of NMR search models. The most intuitive design is to remove structurally disordered regions such as long flexible chains or loops or to combine flexibility information into search models by using a composite search model (Müller *et al.*, 1995; Kleywegt *et al.*, 1994) or artificial temperature factors (Wilmanns & Nilges, 1996; Baldwin *et al.*, 1991; Anderson *et al.*, 1996). A different method was developed to improve the accuracy of NMR structure models by energy-based refinement (Qian *et al.*, 2007). Another NMR model-preparation protocol called *FindCore* has been reported that identifies one or more sets of 'core atoms' with well defined positions (Mao *et al.*, 2011; Snyder & Montelione, 2005), and recently the program *AMPLE* has been used to process NMR ensembles into MR search models by a cluster-and-truncate method (Bibby *et al.*, 2013).

Here, we present a different approach to handle the disagreement between the NMR and crystallographic data. In addition to the usual search-model preparation from NMR data, we build a molecular envelope around the C^{α} atoms and search at low resolution. Envelope phasing is a method that utilizes protein envelope information to solve the crystallographic phase problem. The program *FSEARCH* (Hao *et al.*, 1999; Ockwell *et al.*, 2000; Hao, 2001) is capable of correctly positioning the envelope in a crystallographic unit cell. *FSEARCH* uses a six-dimensional search on structure factors instead of the conventional Patterson search, because the density inside the envelope is uniform, with the result that the intra-envelope vectors are uniformly distributed and therefore do not match the intramolecular (atom-to-atom) vectors represented by the Patterson function. In order to find the best match between the observed and calculated structure factors, *FSEARCH* performs a six-dimensional search on orientation and translation simultaneously. *FSEARCH* can use a
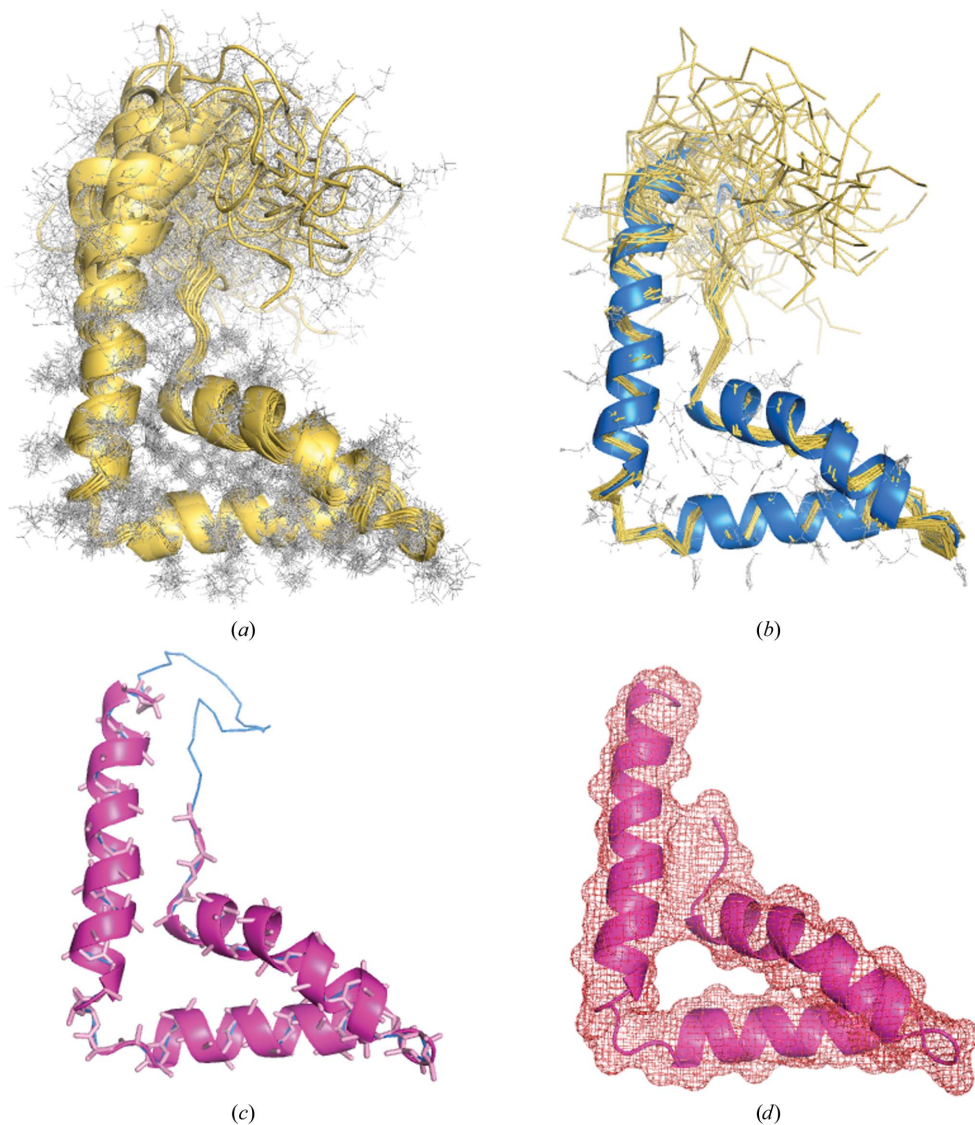

Figure 1
Search-model preparation. (*a*) NMR ensemble displayed in cartoon representation (yellow). (*b*) NMR ensemble in ribbon representation (yellow) and averaged model in cartoon representation (blue). (*c*) Averaged model in ribbon representation (blue) and modified polyalanine model in cartoon representation (magenta). (*d*) Modified model in cartoon representation (magenta) and the envelope (mesh).

Table 1
Molecular-replacement test results using averaged models.

The r.m.s.d. was calculated between the averaged model and the superimposed crystal structure. The *Phaser* LLG was obtained using the averaged model and also the NMR ensemble with default parameters at full resolution as well as at 5, 6 and 7 Å resolution. The correct solution was identified by the lowest *FSEARCH R* factor.

Target	Crystal PDB code	NMR PDB code	Crystal data resolution (Å)	Space group	Length (residues)	Main-chain r.m.s.d. (Å)	*Phaser* LLG	*FSEARCH* R factor	*phenix.autobuild* R/free R (%)
BeR31	3cpk	2k2e	2.5	$P4_32_12$	150	1.23	<86	0.535	26/36
CcR55	2o0q	2jqn	1.8	$C222$	115	1.31	<66	0.460	20/22
CtR107	3e0h	2kcu	1.8	$P2_12_12_1$	158	1.58	<42	0.491	26/32
HMG	2hdz	1k99	2.5	$P6_5$	91	1.60	<20	0.509	26/30

pre-determined envelope from any source, such as electron microscopy (EM), small-angle X-ray scattering (SAXS) or the atomic coordinates of a homologous model structure (Hao, 2006).

2. Methods

The program *FSEARCH* is able to perform a six-dimensional search to locate the envelope in the unit cell, which provides a starting point for further model building and refinement. The result of further model building will be regarded as an indicator of whether the search method is successful. The general workflow of our method is as follows.

(i) Model preparation. As there are always several models in one NMR PDB file, the first step is to obtain an averaged model from all models. Side-chain truncation is then applied such that any side chain longer than alanine is truncated to alanine, with flexible main-chain regions removed. The poly-alanine model is placed onto a three-dimensional grid of 1 Å spacing. A molecular envelope is constructed where the grid nodes closest to each non-H atom and its six nearest neighbours are set to 1 and the other grid nodes are set to 0. An example is shown in Fig. 1.

(ii) Six-dimensional search. The prepared envelope is delivered to *FSEARCH* in order to locate the correct envelope position in the unit cell. The *FSEARCH* result is given as a list of translations and orientations sorted in ascending order by *R* factor. To save computational time, it is suggested that an initial coarse search should be performed first to find a rough solution (usually 5° steps in the Eulerian angles α, β and γ and 2 Å steps in *x*, *y* and *z*); a finer search (usually 1° steps in α, β and γ and 1 Å steps in *x*, *y* and *z*) based on the rough solution can then be used in order to find the global minimum *R* factor.

(iii) Applying the solution. The polyalanine model prepared in step (i) is translated and oriented by the operation obtained from step (ii) with the lowest *R* factor.

(iv) Model building and refinement. The new model from step (iii) is delivered to an autobuilding program and is then subjected to further refinement. Since the model obtained in step (i) is an averaged model, chemical restraints such as bond lengths or bond angles might not be correct. Therefore, the autobuilding program is needed to fix this problem and also to complete parts that were missing in the initial model.

3. Test results

In order to test our method, several models with both NMR and crystal structures available were used as test cases. The NMR structures were used as the search models, while the crystal structures were regarded as targets. Experimental structure factors were used in *FSEARCH*. The aim was to test whether *FSEARCH* was able to correctly locate the envelope in the unit cell. The results that *FSEARCH* gave were passed to an autobuilding program. Success of the autobuilding program would indicate that *FSEARCH* gave a correct solution. The test protocol was based on the general workflow as described above in §2. A search model based on the NMR structures was obtained using *CHAINSAW* (Stein, 2008). The target solution (translation and rotation) was obtained by superposing the search model onto the crystal structure using *Coot* (Emsley & Cowtan, 2004). The *CCP*4 version of the *FSEARCH* program was used for the six-dimensional search (Hao, 2006). *phenix.autobuild* (Adams *et al.*, 2010) was used for model building.

Three NMR structures with 100% homologous crystal structures presented in Mao *et al.* (2011) were used in our test (BeR31, CcR55 and CtR107; Mao *et al.*, 2011). Also included in the test was the HMG box 5 crystal structure (PDB entry 2hdz; Rong *et al.*, 2007), which shows only 20% homology to the HMG box 1 NMR structure (PDB entry 1k99; Xu *et al.*, 2002). The transformation of the prepared atomic model to an envelope is performed within *FSEARCH*. The *R* factor at 5–7 Å resolution between the structure factors calculated from the atomic model and the envelope is in the range 14–23%. The test results are summarized in Table 1.

For all four test cases, the conventional molecular-replacement procedure *Phaser* (McCoy *et al.*, 2007) was not successful in yielding a correct solution when tested using data at both full and low resolution or using full NMR ensembles, as demonstrated by the low LLG values and also by manual examination. In contrast, the six-dimensional envelope search approach described in §2 produced the correct structures for all four cases.

It is worth noting the following.

(i) In the HMG case, the NMR model (PDB code 1l8y; Yang *et al.*, 2003) with 100% sequence identity to the crystal structure could not be placed in the correct position in the crystallographic unit cell by either *Phaser* or *FSEARCH* because the r.m.s.d. between the NMR structure and the crystal structure is too large (2.5 Å).

(ii) *FSEARCH* was tested using data resolutions of 5, 6 and 7 Å for each case. The first three data sets led to similar results at different resolutions, while the HMG case showed resolution sensitivity in envelope phasing (Fig. 2). For the HMG case, only the result using 6 Å resolution data led to successful autobuilding, while the tests with 5 or 7 Å resolution data failed. A detailed analysis shows that the 6 Å resolution solution is closer to the correct position than the 5 and 7 Å resolution solutions.

3.1. Using centroid models instead of averaged models in the search

For the four cases, we also used centroid models to test our approach. The centroid models were obtained by choosing the most representative models among the NMR ensembles *via*

the *OLDERADO* web service (http://www.ebi.ac.uk/pdbe/ NMR/olderado/; Kelley *et al.*, 1997). Other procedures in model preparation such as side-chain modification and envelope generation were performed in the same way as in the test with the averaged models. The test results are summarized in Table 2 and only the *FSEARCH* solutions of two cases, BeR31 and CcR55, led to a successful autobuilding. Compared with the results with the averaged models in Table 1, the centroid models have larger r.m.s.d.s against the crystal structures.

3.2. Evaluating the results

The r.m.s.d. between the main-chain C^{α} atoms of the *FSEARCH* solution and the known crystal structure is used
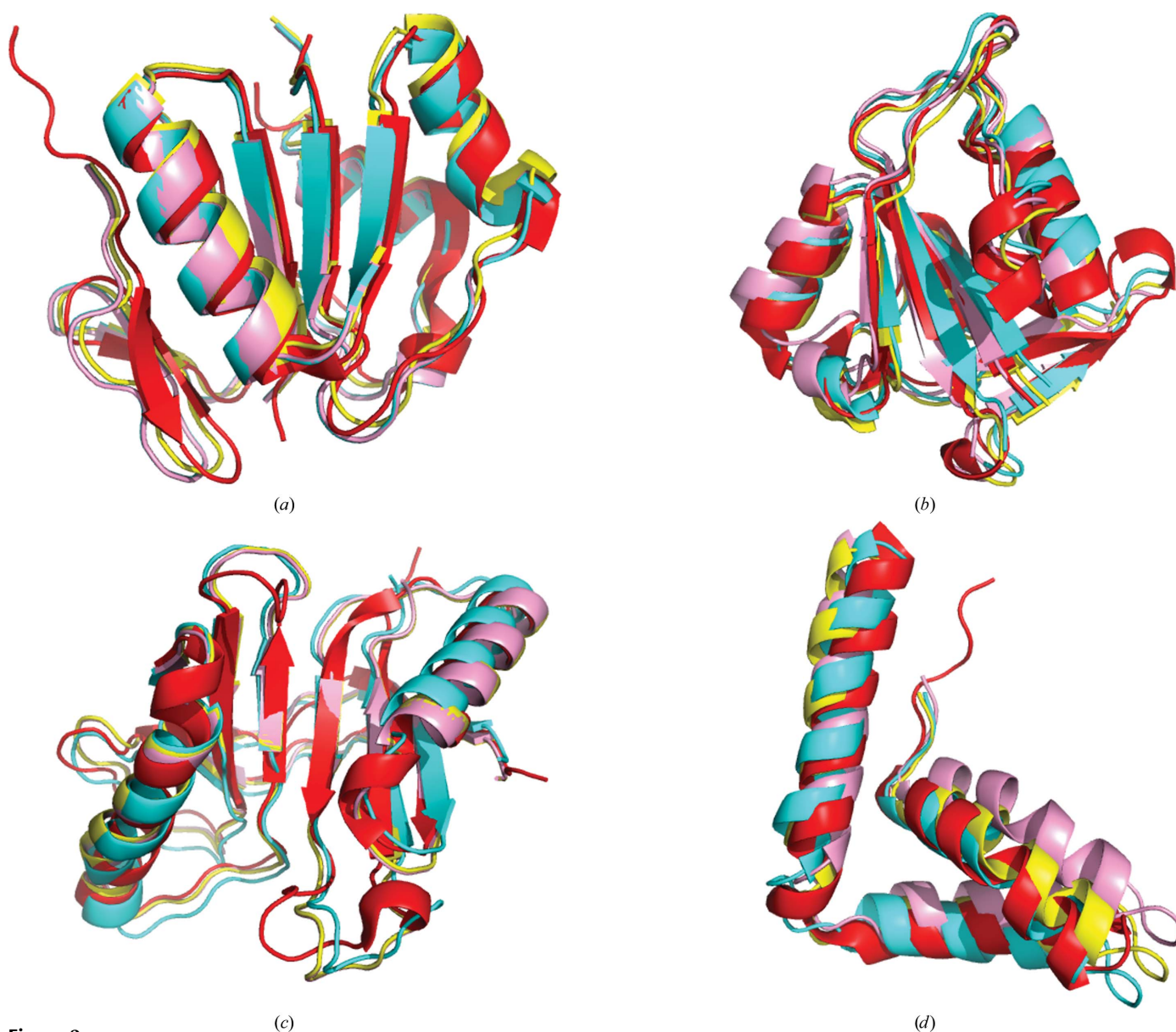

Figure 2
FSEARCH results with the averaged models at different resolutions. The target crystal structures are coloured red. The averaged models after applying the *FSEARCH* solution at 5 Å resolution are coloured pink, at 6 Å yellow and at 7 Å cyan. The test cases are (*a*) BeR31, (*b*) CcR55, (*c*) CtR107 and (*d*) HMG.

Table 2
Molecular-replacement test results using centroid models.

The r.m.s.d. was calculated between the NMR centroid model and the superimposed crystal structure. The *Phaser* LLG was obtained using the centroid model with default parameters at full resolution and also at 5, 6 and 7 Å resolution. The correct solution was identified by the lowest *FSEARCH R* factor.

Target	Main-chain r.m.s.d. (Å)	*Phaser* LLG	*FSEARCH* R factor (%)	*phenix.autobuild* R/free R (%)
BeR31	1.41	<59	0.533	25/33
CcR55	1.60	<43	0.475	26/28
CtR107	1.84	<16	0.514	—
HMG	1.80	<19	0.515	—

Table 3
R.m.s.d. (Å) between the *FSEARCH* solution and the target crystal structure.

R.m.s.d.-A is for the averaged model and r.m.s.d.-C is for the centroid model. Solutions from both the averaged models and the centroid models were analysed at different resolutions (5, 6 and 7 Å). The number of reflections used and the data completeness at each resolution are also shown.

Target		5 Å	6 Å	7 Å
BeR31	R.m.s.d.-A	1.56	1.43	1.60
	R.m.s.d.-C	1.97	1.96	1.97
	No. of reflections	614	370	240
	Completeness (%)	100	100	100
CcR55	R.m.s.d.-A	1.71	1.59	1.73
	R.m.s.d.-C	1.95	2.05	1.98
	No. of reflections	991	585	374
	Completeness (%)	98.6	97.8	96.9
CtR107	R.m.s.d.-A	1.74	1.69	1.95
	R.m.s.d.-C	2.84	2.56	2.56
	No. of reflections	671	397	255
	Completeness (%)	99.2	98.8	98.0
HMG	R.m.s.d.-A	3.43	1.78	2.13
	R.m.s.d.-C	3.76	2.14	2.23
	No. of reflections	573	327	200
	Completeness (%)	96.1	93.4	89.7

to evaluate the accuracy of *FSEARCH* results under different reductions in resolution (Table 3). The r.m.s.d. values given by the averaged models are consistently smaller than the values given by the centroid models, indicating that *FSEARCH* could yield a more accurate result using averaged models.

Among the four cases that we have tested, the HMG case is the most difficult not only owing to the relatively large r.m.s.d. and low sequence identity between the search model and the target crystal structure but also because this protein is the smallest (with only 91 amino acids), making this case the most sensitive to the search parameters. To investigate whether the experimental data quality or completeness could play a role in determining the success of the method, we repeated the HMG tests using the structure factors calculated from the crystal structure as the *FSEARCH* input. Again, the 6 Å resolution data led to successful autobuilding while the test with 5 Å resolution data failed. At 7 Å resolution, using the same number of reflections with the experimental data failed, but a complete data set with or without bulk-solvent correction was successful. The reason could be that at 5 Å resolution the (wrong) details of an envelope may prevent one from finding the solution and these details become less significant below

6 Å, but at a resolution of 7 Å the data completeness is crucial to ensure a sufficient data-to-parameter ratio.

4. Discussion

The conformational differences between an NMR search model and a crystal structure may result in the failure of conventional molecular-replacement procedures. In this study we have demonstrated that a low-resolution envelope would be less sensitive to these differences and gives a more intuitive view of structural similarity. Utilizing the envelope as a search model and searching at low resolution based on structure factors could provide a new method for solving difficult molecular-replacement cases when there are large discrepancies between the search model and the target structure.

References

Adams, P. D. et al. (2010). *Acta Cryst.* D**66**, 213–221.
Anderson, D. H., Weiss, M. S. & Eisenberg, D. (1996). *Acta Cryst.* D**52**, 469–480.
Baldwin, E. T., Weber, I. T., St Charles, R., Xuan, J.-C., Appella, E., Yamada, M., Matsushima, K., Edwards, B. F. P., Clore, G. M., Gronenborn, A. M. & Wlodawer, A. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 502–506.
Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. (2013). *Acta Cryst.* D**69**, 2194–2201.
Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.
Brünger, A. T., Campbell, R. L., Clore, G. M., Gronenborn, A. M., Karplus, M., Petsko, G. A. & Teeter, M. M. (1987). *Science*, **235**, 1049–1053.
Chen, Y. W. (2001). *Acta Cryst.* D**57**, 1457–1461.
Chen, Y. W., Dodson, E. J. & Kleywegt, G. J. (2000). *Structure*, **8**, R213–R220.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Hao, Q. (2001). *Acta Cryst.* D**57**, 1410–1414.
Hao, Q. (2006). *Acta Cryst.* D**62**, 909–914.
Hao, Q., Dodd, F. E., Grossmann, J. G. & Hasnain, S. S. (1999). *Acta Cryst.* D**55**, 243–246.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Jamrog, D. C., Zhang, Y. & Phillips, G. N. (2003). *Acta Cryst.* D**59**, 304–314.
Kelley, L. A., Gardner, S. P. & Sutcliffe, M. J. (1997). *Protein Eng.* **10**, 737–741.
Kleywegt, G. J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K. & Jones, T. A. (1994). *Structure*, **2**, 1241–1258.
Mao, B., Guan, R. & Montelione, G. T. (2011). *Structure*, **19**, 757–766.
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.
Müller, T., Oehlenschläger, F. & Buehner, M. (1995). *J. Mol. Biol.* **247**, 360–372.
Ockwell, D. M., Hough, M. A., Grossmann, J. G., Hasnain, S. S. & Hao, Q. (2000). *Acta Cryst.* D**56**, 1002–1006.
Pähler, A., Smith, J. L. & Hendrickson, W. A. (1990). *Acta Cryst.* A**46**, 537–540.
Pannu, N. S. & Read, R. J. (2004). *Acta Cryst.* D**60**, 22–27.
Perutz, M. F. (1956). *Acta Cryst.* **9**, 867–873.
Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007). *Nature (London)*, **450**, 259–264.
Rong, H., Li, Y., Shi, X., Zhang, X., Gao, Y., Dai, H., Teng, M., Niu, L., Liu, Q. & Hao, Q. (2007). *Acta Cryst.* D**63**, 730–737.

Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.

Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Snyder, D. A. & Montelione, G. T. (2005). *Proteins*, **59**, 673–686.

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

Urzhumtsev, A. & Podjarny, A. (1995). *Acta Cryst.* D**51**, 888–895.

Wilmanns, M. & Nilges, M. (1996). *Acta Cryst.* D**52**, 973–982.

Xu, Y., Yang, W., Wu, J. & Shi, Y. (2002). *Biochemistry*, **41**, 5415–5420.

Yang, W., Xu, Y., Wu, J., Zeng, W. & Shi, Y. (2003). *Biochemistry*, **42**, 1930–1938.