Solution solution: using NMR models for molecular replacement

Yu Wai Chen

Cambridge University Chemical Laboratory and Centre for Protein Engineering, MRC Centre, Hills Road, Cambridge CB2 2QH, England

Correspondence e-mail:
ywc@mrc-lmb.cam.ac.uk

NMR structures can serve as a good source of search models in crystal structure determination by molecular replacement. However, owing to the inherent problems of NMR models, this procedure is not always straightforward. Here, an updated overview is presented with particular emphasis on the preparation of NMR search models and the latest trends in methodology. An experimental protocol developed recently is described and results on its use in solving a new structure as well as its test against a difficult published case are presented.

1. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a powerful alternative to X-ray crystallography for determining structures of small macromolecules and contributes to a substantial fraction of the depositions in the Protein Data Bank. Brünger *et al.* (1987) first provided a proof-of-principle that NMR structures, many of them being compact protein structural modules, can serve as trial models to solve crystal structures by molecular replacement (MR). The first real application was reported ten years ago (Baldwin *et al.*, 1991). Owing to the inherent deficiencies of NMR models, their use in MR is not always straightforward, even if the search model and the target crystal structure are of the same macromolecule. Nevertheless, the number of successful MR solutions has risen steadily in the past few years. Up to the time when this article was written, 26 cases (Table 1, cases 3–28) have been reported in the literature, demonstrating that the difficulties in using NMR search models could be readily overcome. The readers are referred to a comprehensive review on this topic (Chen *et al.*, 2000). Here, based on this recent review, I present a concise overview supplemented by new information appearing in the past few months and new test results of a recommended MR protocol.

2. Overview

2.1. Problems of NMR models

The success of MR in macromolecular crystallography depends in general largely on the sequence (and structural) homology between the target structure and the search model. Using NMR search models in MR challenges this view: MR can often fail, even when the search model has 100% sequence identity to that of the target crystal structure. This can be attributed to three inherent problems of NMR models: (i) inaccuracies in NMR structures that are mainly based on short distance restraints, (ii) imprecision of search models as a result

Table 1
A list of structures surveyed in this study.

The accuracy column shows r.m.s. deviation of backbone (unless otherwise stated) atomic positions between the crystal and NMR structures in the well defined core region of the respective proteins involved, usually with loops and temini excluded. If the protein is oligomeric, the comparison is for the oligomeric state of the search model used in MR. This list is arranged in chronological order of publication. Some articles reported more than one MR calculation (*e.g.* case 7) and the tabulated parameters in this table are those for a representative calculation.

	Year	Protein	Software	Accuracy (Å)	Model content	Reference
1	1987	Crambin	Other	1.1	Mean	Brünger *et al.* (1987)
2	1989	Tendamistat	Other	1.03	Individual	Braun *et al.* (1989)
3	1991	IL-8	Other	0.96	Mean	Baldwin *et al.* (1991)
4	1992	CD8α	*X-PLOR*	—	Composite	Leahy *et al.* (1992)
5	1994	Endothelin	Other	2.0	Individual	Janes *et al.* (1994)
6	1994	CRABP II†	*AMoRe*	—	Composite	Kleywegt *et al.* (1994)
7	1995	hIL4	Others/*X-PLOR*	1.2–1.4 C$^\alpha$	Individual/ensemble	Müller *et al.* (1995)
8	1995	PH†	*AMoRe*	1.6	Individual	Hyvönen *et al.* (1995); (1996)
9	1995	Er-1	*X-PLOR*	1.17 all C$^\alpha$	Mean	Weiss *et al.* (1995); Anderson *et al.* (1996)
10	1996	ACBP	*AMoRe*	1.9 C$^\alpha$	Ensemble	Kleywegt (1996*b*); van Aalten *et al.* (in preparation)
11	1996	p53 TET	*X-PLOR*	0.73 C$^\alpha$	Mean	Miller *et al.* (1996)
12	1997	MCP-1	*AMoRe*	0.9/1.0 C$^\alpha$	Mean	Lubkowski *et al.* (1997)
13	1998	p53 TET	*AMoRe*	1.4 C$_r$	Ensemble	Mittl *et al.* (1998)
14	1998	p14$^{\text{TCL1}}$	*AMoRe*	1.1 C$^\alpha$	Mean	Hoh *et al.* (1998)
15	1998	IM7	*AMoRe*	1.7 C$^\alpha$	Ensemble	Dennis *et al.* (1998)
16	1998	RBI†	*AMoRe*	2.0	Mean	Strobl *et al.* (1998)
17	1998	p19†	*AMoRe*	1.7 C$^\alpha$	Not reported	Brotherton *et al.* (1998)
18	1998	CHFI	*EPMR*	1.61	Single calculated	Behnke *et al.* (1998)
19	1999	RATI	*AMoRe*	2.5 all C$^\alpha$	Individual	Gourinath *et al.* (1999)
20	1999	NK1	*AMoRe*	0.4 C$^\alpha$	Mean	Chirgadze *et al.* (1999)
21	1999	CTLA-4	*AMoRe*	2.9 all C$^\alpha$	Mean	Sheriff *et al.* (1999)
22	1999	Protein S	*AMoRe*	1.1	Mean	Wenk *et al.* (1999)
23	1999	CV-N	*AMoRe*	0.55	Mean	Yang *et al.* (1999)
24	1999	hCD2:hCD58	*AMoRe*	1.04 all C$^\alpha$	Mean	Wang *et al.* (1999)
25	1999	p85α SH2	*AMoRe*	0.92 C$^\alpha$	Ensemble	Hoedemaeker *et al.* (1999)
26	2000	Met-Rantes	*AMoRe*	Not reported	Not reported	Hoover *et al.* (2000)
27	2000	p73α SAM	*AMoRe*	1.38	Ensemble	Wang *et al.* (2000, 2001)
28	2001	Mss4	*AMoRe*	1.3 C$^\alpha$	Mean	Zhu *et al.* (2001)
29	2001	p85α SH2†	*AMoRe*	1.0 C$^\alpha$	Ensemble	Pauptit *et al.* (2001)
30	2001	L30e	*AMoRe*	1.62	Ensemble	Chen & Wong (unpublished results)

† MR was performed on this component of a complex structure.

of limited data (low observation-to-parameter ratio) and (iii) the difficulty in representing the relative reliability of atomic positions in an NMR model.

2.1.1. Accuracy and precision. For the updated list in Table 1, when the well defined regions of a crystal structure and those of its NMR search model are compared, the r.m.s. deviations of the backbone atoms (C$^\alpha$, N, C, O) have a mean value of 1.3 ± 0.2 Å (Chen *et al.*, 2000). The overall structures of the two states agree well and large structural differences are generally local and limited to flexible loops, the termini and exposed side chains. Advancements in NMR instrumentation and methodology have led to much improved agreement between NMR and crystal structures (see, for example, Kuszewski *et al.*, 1999). Modern NMR structures show excellent internal consistency: the overall backbone r.m.s. deviation of individual conformers in well defined regions is usually less than 1 Å, with a mean value of 0.5 ± 0.1 Å for the cases listed in Table 1 (Chen *et al.*, 2000). In one study, the authors reported that only the best-refined final NMR model can lead to successful MR, while intermediate models that are lesser in precision all failed (case 7; Müller *et al.*, 1995). This is echoed by our results on one test study (see §3.2.2). In many cases, modern NMR structures are good for MR calculations,

provided that they are carefully prepared. The major task of model preparation is to remove (or down-weight) regions where large local structural variations are likely to occur and to use the best structural representation of the well defined regions in the MR calculations.

2.1.2. Reliability information. In MR calculations, crystallographic *B* factors are important in supplying weights to the atomic contributions of the scattering factors and reflecting the precision of atomic positions. An individual NMR conformer lacks this information. However, the equivalent information is embodied in an NMR ensemble — poorly defined regions supply fewer experimental restraints and these regions exhibit larger variations in atomic positions.

To exploit the reliability information in an NMR structure for use in MR, two approaches have been developed. The first method involves using a single model with artificially calculated *B* factors based on atomic r.m.s. deviations from the mean structure (Baldwin *et al.*, 1991; Anderson *et al.*, 1996; Wilmanns & Nilges, 1996). The second approach simply involves using the whole ensemble as a composite model (Leahy *et al.*, 1992; Kleywegt *et al.*, 1994; Müller *et al.*, 1995; Kleywegt, 1996*a*), with all atoms assigned uniform *B* factors. An ensemble supplies an inherent weighting scheme

according to the mutual agreement of equivalent atomic positions.

Of the 22 cases in Table 1 where a crystal structure was solved with an NMR model, 13 were solved with single models and seven were solved with ensemble models; the remaining case 7 made use of both types, while the model used in case 26 was not detailed. Most (11) of the single model cases made use of a representative model, usually the restrained minimized average structure. Individual conformers rarely led to success (cases 5 and 19 only). Single models were often preferable because they are easier to handle and the calculations are faster to perform. However, some very difficult cases could only be solved using ensemble models (Dennis *et al.*, 1998; Mittl *et al.*, 1998; Hoedemaeker *et al.*, 1999; Chen & Clore, 2000; Wang *et al.*, 2001). An ensemble is probably a more realistic representation of the 'true' (time- and space-averaged) structure than a single model (Sutcliffe, 1993). Use of ensemble models is becoming more popular.

2.2. Software

Since 1994, *AMoRe* (Navaza, 1994; Navaza & Saludjian, 1997) has led to the solution of more structures than any other program (Table 1). The major advantages of this program are its speed and the fact that many potential solutions can be tested in a single run. It may not offer the best signal-to-noise discrimination at every stage, but a correct rotation-function solution that is buried among noise peaks is still useful in the subsequent translation search and rigid-body refinement.

Continued increases in computing power have brought about some new excitements. Recently developed partial six-dimensional searches (Chang & Lewis, 1997; Kissinger *et al.*, 1999; Glykos & Kokkinidis, 2000) can be performed with a speed comparable to conventional two-step MR. Two of these programs are described in this issue (Glykos & Kokkinidis, 2001; Kissinger *et al.*, 2001).

A new approach in simultaneous search for multiple copies of a molecule in the unit cell by MR has been developed recently and implemented into the program *MOLREP* (Collaborative Computational Project, Number 4, 1994; Vagin & Teplyakov, 2000). The authors reported success in a previously failed MR case searching with an NMR model.

Another new molecular-replacement program, *Beast* (Read, 2001), has been designed to use multiple possible molecular-replacement models and is thus particularly suited to the use of ensemble NMR models in MR. We have performed preliminary trials with this program on a very difficult test case, that of CHFI (see §3.2.1), and found it to be successful (Read & Chen, unpublished results).

3. Recommended protocol and case studies

3.1. Experimental

It is customary to delete unstructured residues in a search model. A set of tools to help with this task is described in Kleywegt (1996a) and is available at the URL http://xray.bmc.uu.se/usf/factory_6.html. Recent NMR structures usually contain around 20 conformers in a bundle and are generally good enough for MR (Chen & Clore, 2000). Obvious 'outliers' can be removed from the set. A script called *multi_probe* (ftp://xray.bmc.uu.se/pub/gerard/omac/multi_probe) was found to be most useful for preparing a set of three ensemble models, with varying extents of side-chain truncation. The script first aligns members of the ensemble and then prepares an all-atom model, a poly Ser/Ala/Gly (poly SAG; all non-glycine/alanine side chains changed to serine) model and a poly AG model. A detailed description of the procedure can be found on the internet (http://imsb.au.dk/~mok/o/ofaq/Q.879.html). If a single NMR model is used, artificial *B* factors can be assigned using an empirical formula based on atomic r.m.s. deviations from the mean structure (Wilmanns & Nilges, 1996). This procedure has been implemented in a Perl script and is available for download (http://www.mrc-cpe.cam.ac.uk/~ywc/rmsdB.html).

The whole NMR ensemble, as prepared by the *multi_probe* script, was input into *AMoRe* as a single trial model. For most cases in Table 1, the high-resolution cut-off of the data used for searching falls in a narrow range from 3.5 to 4.5 Å; commonly used low-resolution limits are 10.0 and 15.0 Å. All the MR calculations in this work were performed with data in the resolution range 15–3.5 Å. This protocol has been tested on three difficult published cases (Table 1, cases 9, 13 and 18) and applied to solving two structures (Table 1, cases 27 and 30). I summarize previous findings and present new results in the following section.
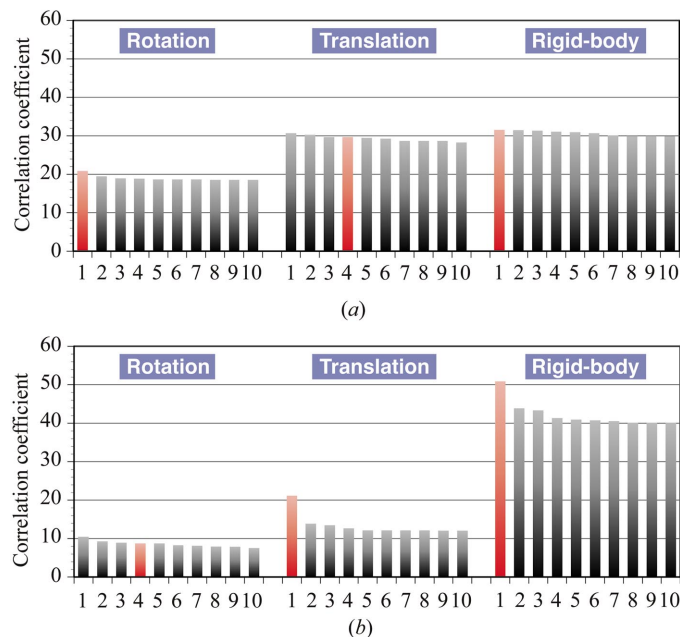
Figure 1
Results of MR calculations with NMR models. In each graph, (*a*) CHFI and (*b*) L30e, the top 10 peaks of the three MR steps: rotation function, translation function and rigid-body refinement are shown, in descending order. The correct solutions are coloured red and noise peaks are coloured black.

3.2. Case studies

3.2.1. Three published cases tested

In two earlier articles, we have reported the successful application of this recommended protocol in re-solving two published problems: namely, the p53 TET domain and the Er-1 pheromone (Chen & Clore, 2000; Chen *et al.*, 2000). In both cases, a poly SAG model is found to be the most successful and results are very clear at every stage.

We also studied a third test case: the corn hageman factor inhibitor (CHFI). This structure was originally solved with *EPMR* (Kissinger *et al.*, 1999), a program which implements six-dimensional searches with evolutionary programming. Neither *X-PLOR* nor *AMoRe* led to an MR solution (Behnke *et al.*, 1998), making it particularly challenging. Our earlier attempt to solve the structure with this recommended protocol also failed (Chen *et al.*, 2000). Jorge Navaza, the author of *AMoRe*, kindly advised that a solution could be obtained with slight modification to the procedure. With the poly SAG ensemble model, the correct rotation solution is obtained without any problem as the top peak. Interestingly, the subsequent translation search only worked when the correlation coefficient (CC) in terms of intensities target and not the Crowther–Blow translation function (default in *AMoRe*) was used (Jorge Navaza, personal communications). The search results are noisy (Fig. 1*a*): the correct translation-function peak only ranked fourth, with a CC of 29.5. After rigid-body refinement, the correct peak is promoted to the top (CC = 31.4), but is hardly distinguishable from the highest noise peak (CC = 31.2). Nevertheless, the top peak is associated with the lowest *R* factor and corresponds to the correct solution.

3.2.2. Two new cases solved

The recommended protocol was employed in solving the crystal structure of the human p73α SAM domain and has already been reported (Chen *et al.*, 2000; Wang *et al.*, 2000).

Here, I present fresh results of the application of this protocol to the solution of a new structure: that of the ribosomal protein L30e (Wong *et al.*, 2001). Structure determination by NMR and by MR took place in parallel. MR using an NMR model of a homologous protein did not yield a solution and neither did NMR models of the same protein at the early stages of refinement give positive results. It was not until a good quality near-completion NMR model was available that we could solve the structure. This ensemble contains 25 conformers and has an internal precision (backbone r.m.s. deviation to the mean structure) of 0.6–0.7 Å. With this intermediate model, we generated the set of all atoms, poly AG and poly SAG models, but none of these gave an MR solution. We then generated an ensemble model preserving all the hydrophobic core side chains while the long surface side chains (Lys, Arg, Met, Gln, Glu, Asp and Asn) were changed to serine. Using this search model, a prominent peak appeared after the translation search that was clearly discriminated from the noise (Fig. 1*b*). This solution has the highest CC of 50.7, which is well above the highest noise peak of 43.7 and has the lowest *R* factor of 0.46 among noise peaks having *R* factors ranging from 0.49 to 0.52. This solution is only obtained in $P6_1$ but not in the enantiomorphic space group. Checking backward, the correct orientation only ranked fourth in the rotation-function search (Fig. 1*b*). Successful refinement confirmed that the MR solution is correct: the current *R* factor and free *R* factor are both below 0.3 (Chen & Wong, unpublished results).

4. Discussion

The recommended protocol was found to be successful in solving two new structures as well as offering improvements over published results for three difficult test cases. It is interesting to compare the structural differences in these cases. The r.m.s. deviation of well defined backbone atoms for the p53 TET domain is 0.4 Å, that for Er-1 is 1.2 Å and that for the p73α SAM domain is 1.4 Å, while that for CHFI and L30e is 1.6 Å. The clarity of the results is correlated somewhat with the structural differences between the search model and the target structure, *i.e.* the accuracy of the respective search models. During the re-examination of p53 TET domain, it was found that a more accurate structure led to substantially improved search results (Chen & Clore, 2000). Only a near-completely refined NMR models can lead to solution of the L30e. In the case of CHFI, one learned that MR calculations can be sensitive to the target translation function used.

Obtaining an MR solution is only halfway through the problem. Subsequent structure refinement and rebuilding can be very tedious and frustrating but these are outside the scope of this work. Some practical experiences can be found in another manuscript in this issue (Pauptit *et al.*, 2001).

References

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.

Behnke, C. A., Yee, V. C., Trong, I. L., Pedersen, L. C., Stenkamp, R. E., Kim, S.-S., Reeck, G. R. & Teller, D. C. (1998). *Biochemistry*, **37**, 15277–15288.

Braun, W., Epp, O., Wüthrich, K. & Huber, R. (1989). *J. Mol. Biol.* **206**, 669–676.

Brotherton, D. H., Dhanaraj, V., Wick, S., Brizuela, L., Domaille, P. J., Volyanik, E., Xu, X., Parisini, E., Smith, B. O., Archer, S. J., Serrano, M., Brenner, S. L., Blundell, T. L. & Laue, E. D. (1998). *Nature (London)*, **395**, 244–250.

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.

Chang, G. & Lewis, M. (1997). *Acta Cryst.* D**53**, 279–289.

Chen, Y. W. & Clore, G. M. (2000). *Acta Cryst.* D**56**, 1535–1540.

Chen, Y. W., Dodson, E. J. & Kleywegt, G. J. (2000). *Structure*, **8**,

R213–R220.

Chirgadze, D. Y., Hepple, J. P., Zhou, H., Byrd, R. A., Blundell, T. L. & Gherardi, E. (1999). *Nature Struct. Biol.* **6**, 72–79.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dennis, C. A., Videler, H., Pauptit, R. A., Wallis, R., James, R., Moore, G. R. & Kleanthous, C. (1998). *Biochem. J.* **333**, 183–191.

Glykos, N. M. & Kokkinidis, M. (2000). *Acta Cryst.* D**56**, 169–174.

Glykos, N. M. & Kokkinidis, M. (2001). *Acta Cryst.* D**57**, 1462–1473.

Gourinath, S., Srinivasan, A. & Singh, T. P. (1999). *Acta Cryst.* D**55**, 25–30.

Hoedemaeker, F. J., Siegal, G., Roe, S. M., Driscoll, P. C. & Abrahams, J. P. (1999). *J. Mol. Biol.* **292**, 763–770.

Hoh, F., Yang, Y.-S., Guignard, L., Padilla, A., Stern, M.-H., Lhoste, J. M. & van Tilbeurgh, H. (1998). *Structure*, **6**, 147–155.

Hoover, D. M., Shaw, J., Gryczynski, Z., Proudfoot, A. E. I., Wells, T. & Lubkowski, J. (2000). *Protein Peptide Lett.* **7**, 73–82.

Hyvönen, M., Macias, M. J., Nilges, M., Oschkinat, H., Saraste, M. & Wilmanns, M. (1995). *EMBO J.* **14**, 4676–4685.

Janes, R. W., Peapus, D. H. & Wallace, B. A. (1994). *Nature Struct. Biol.* **1**, 311–319.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.

Kissinger, C. R., Gehlhaar, D. K., Smith, B. A. & Bouzida, D. (2001). *Acta Cryst.* D**57**, 1475–1479.

Kleywegt, G. J. (1996a). *CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **32**, 32–36.

Kleywegt, G. J. (1996b). *Acta Cryst.* D**52**, 842–857.

Kleywegt, G. J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K. & Jones, T. A. (1994). *Structure*, **2**, 1241–1258.

Kuszewski, J., Gronenborn, A. M. & Clore, G. M. (1999). *J. Am. Chem. Soc.* **121**, 2337–2338.

Leahy, D. J., Axel, R. & Hendrickson, W. A. (1992). *Cell*, **68**, 1145–1162.

Lubkowski, J., Bujacz, G., Boqué, L., Domaille, P. J., Handel, T. M. & Wlodawer, A. (1997). *Nature Struct. Biol.* **4**, 64–69.

Miller, M., Lubkowski, J., Rao, J. K. M., Danishefsky, A. T.,

Omichinski, J. G., Sakaguchi, K., Sakamoto, H., Appella, E., Gronenborn, A. M. & Clore, G. M. (1996). *FEBS Lett.* **399**, 166–170.

Mittl, P. R. E., Chène, P. & Grütter, M. G. (1998). *Acta Cryst.* D**54**, 86–89.

Müller, T., Oehlenschläger, F. & Buehner, M. (1995). *J. Mol. Biol.* **247**, 360–372.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Navaza, J. & Saludjian, P. (1997). *Methods Enzymol.* **276**, 581–594.

Pauptit, R. A., Dennis, C. A., Derbyshire, D. J., Breeze, A. L., Weston, S. A., Rowsell, S. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 1397–1404.

Read, R. (2001). *Acta Cryst.* D**57**, 1373–1382.

Sheriff, S., Klei, H. E. & Davis, M. E. (1999). *J. Appl. Cryst.* **32**, 98–101.

Strobl, S., Maskos, K., Wiegand, G., Huber, R., Gomis-Rüth, F. X. & Glockshuber, R. (1998). *Structure*, **6**, 911–921.

Sutcliffe, M. J. (1993). *Protein Sci.* **2**, 936–944.

Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.

Wang, J.-H., Smolyar, A., Tan, K., Liu, J.-H., Kim, M., Sun, Z. Y., Wagner, G. & Reinherz, E. L. (1999). *Cell*, **97**, 791–803.

Wang, W. K., Bycroft, M., Foster, N. W., Buckle, A. M., Fersht, A. R. & Chen, Y. W. (2001). *Acta Cryst.* D**57**, 545–551.

Wang, W. K., Proctor, M. R., Buckle, A. M., Bycroft, M. & Chen, Y. W. (2000). *Acta Cryst.* D**56**, 769–771.

Weiss, M. S., Anderson, D. H., Raffioni, S., Bradshaw, R. A., Ortenzi, C., Luporini, P. & Eisenberg, D. (1995). *Proc. Natl. Acad. Sci. USA*, **92**, 10172–10176.

Wenk, M., Baumgartner, R., Holak, T. A., Huber, R., Jaenicke, R. & Mayr, E.-M. (1999). *J. Mol. Biol.* **286**, 1533–1545.

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

Wong, K. B., Wang, W. K., Proctor, M. R., Bycroft, M. & Chen, Y. W. (2001). *Acta Cryst.* D**57**, 865–866.

Yang, F., Bewley, C. A., Louis, J. M., Gustafson, K. R., Boyd, M. R., Gronenborn, A. M., Clore, G. M. & Wlodawer, A. (1999). *J. Mol. Biol.* **288**, 403–412.

Zhu, Z., Dumas, J. J., Lietzke, S. E. & Lambright, D. G. (2001). *Biochemistry*, **40**, 3027–3036.