Acta Cryst. (1996). D**52**, 973–982

Molecular Replacement with NMR Models Using Distance-Derived Pseudo *B* Factors

MATTHIAS WILMANNS* AND MICHAEL NILGES

European Molecular Biology Laboratory, Postfach 102209, D-69012 Heidelberg, Germany.
E-mail: wilmanns@embl-heidelberg.de

(*Received 11 January 1996; accepted 7 March 1996*)

Abstract

The statistical significance of molecular-replacement solutions with models derived from NMR data is strongly enhanced if pseudo *B* factors that reflect expected atomic coordinate errors are introduced. These *B* factors are derived from atomic distances of an ensemble of NMR models and an averaged model. A recently determined X-ray structure of a Pleckstrin homology domain:ligand complex has been used as a test case for molecular replacement with NMR templates. The feasibility of the models for molecular replacement has been studied in two steps: (i) correctness of solutions, verified by correct rotational angles and translations; (ii) statistical significance (detection) of solutions, measured by *R* factors and correlation coefficients in the translation function. None of the models with uniform *B* factors were detectable in the translation function whereas two models with distance-derived *B* factors gave statistically significant *R* factors and correlation coefficients. The potential impact of distance-derived *B* factors on the detectability of molecular-replacement solutions was further tested by using *B* factors that were derived from spatial distances between each NMR model and the X-ray structure. It is concluded that the introduction of distance-derived *B* factors can be an essential component of many future molecular-replacement problems that use NMR models as templates.

1. Abbreviations

CC, correlation coefficient; PH, Pleckstrin homology; IP3, D-myo-inositol-1,4,5-trisphosphate; *B* factor, atomic displacement parameter; r.m.s.d., root-mean-square deviation; NOE, nuclear Overhauser effect.

2. Introduction

In the last decade a substantial fraction of the structures for proteins with molecular weight up to ~20 kDa have been solved by NMR methods. Fig. 1 compares the number of structures solved by X-ray crystallography and NMR which have been published in the most recent annual atlas of macromolecular structures (Hendrickson & Wüthrich, 1994). It has been predicted that the number of new protein structures solved by NMR will increase rapidly and extend to larger proteins in the range of 30–40 kDa (Wagner, 1993; Wüthrich, 1995).

The NMR structures available represent an important pool of models to be explored as templates for the molecular-replacement method in X-ray crystallography. So far, however, only a few X-ray structures phased with NMR templates have been published, among them are Tendamistat (Braun, Epp, Wüthrich & Huber, 1989) and crambin (Brünger *et al.*, 1987) as model simulations, interleukin-8 (Baldwin *et al.*, 1991), interleukin-4 (Müller, Oehlenschläger & Buehner, 1995), endothelin (Janes, Peapus & Wallace, 1994), Er-1 (Weiss *et al.*, 1995) and a Pleckstrin homology (PH) domain:D-myo-inositol-1,4,5-trisphosphate (IP3) complex (Hyvönen *et al.*, 1995). We have reinvestigated the molecular-replacement solution of the PH:IP3 complex structure as a test case which will be described in this publication.

PH domains occur in a variety of signal transduction and cytoskeletal proteins. A function of PH domains seems to be targeting of adjacent protein modules to the membrane (for review see Saraste & Hyvönen, 1995; Pawson, 1995; Cohen, Ren & Baltimore, 1995). The original structure of the unliganded PH domain of β-spectrin (106 residues) was solved by NMR (Macias *et al.*, 1994). The β-spectrin PH domain is folded as a seven-stranded β-sandwich, followed by a C-terminal α-helix. Another α-helix is located in the loop between β-strands 3 and 4.

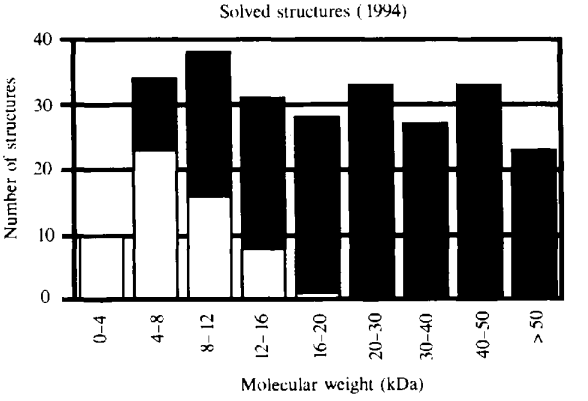

Fig. 1. Number of structures of X-ray structures (black bars) and NMR structures (white bars) published in the annual atlas of macromolecular structures 1994 (Hendrickson & Wüthrich, 1994).

ISSN 0907-4449

Subsequently, we have obtained crystals of the same PH domain complexed with IP3 in space group $P42_12$ and one molecule per asymmetric unit (Hyvönen *et al.*, 1995). The X-ray structure was solved by molecular replacement with the program *AMoRe* (Navaza, 1994) using ten available NMR models* that resulted from the molecular dynamics refinement. The correct solution was obtained from a model that contained distance-derived atomic displacement (*B*) parameters. These atomic pseudo *B* factors were calculated and averaged from the atomic distances between each of the ten NMR models and the averaged NMR model. Each averaged spatial atomic distance was converted into a pseudo *B* factor by the tenfold of the squared distance. The correct solution scored with 4.0σ (rank 7) in the rotation function. The correct solution of the translation function had an *R* factor of 54.3% and correlation coefficient (CC) of 27.1% as defined in the program *AMoRe* (Navaza, 1994).† The correctly positioned model was refined as rigid body to an *R* factor of 52.8% and a CC of 30.9%. We were not able to detect a correct solution from models with uniform *B* factors. The PH:IP3 complex was refined by a combined and iterative procedure that included model rebuilding with the software *O* (Jones, Zou, Cowan & Kjeldgaard, 1991), molecular dynamics refinement with *X-PLOR* (Brünger, 1993) and map improvement with the *ARP* software (Lamzin & Wilson, 1993). The *R* factor and R_{free} of the refined structure were 20.5 and 28.6%, respectively. The final X-ray structure contained all 864 non-H atoms of the PH domain, the IP3 molecule and 77 water molecules. A schematic drawing of the crystal structure is shown in Fig. 2. For further details and structural description see Hyvönen *et al.* (1995).

The central goal in our reinvestigation of the molecular replacement is to identify 'best' NMR models. The quality of a NMR model for molecular replacement is primarily based on its fit to the unknown X-ray coordinates. For the core regions, defined by intramolecular packing constraints, the root-mean-square deviations (r.m.s.d.) between correct NMR and X-ray structures are generally below 2 Å (Billeter, 1992; MacArthur, Laskowski & Thornton, 1994). Exposed loop regions of an NMR structural ensemble frequently show large local deviations because of scarcity of NOE constraints (Wüthrich, 1995). These deviations reflect the true state of a soluble protein as its surface is structurally disordered through contacts with rapidly fluctuating solvent molecules. In contrast, parts of the protein surface are locked into an unique conformation through protein–protein lattice contacts in crystals. These contacts do not generally alter the overall conformation of a protein but can create local surface rigidity which has to be treated as a crystallographic artefact.

The contribution of each scatterer as function of resolution depends on the its amount of vibration. This thermal motion is modelled by the *B* factor,

$$B = 8\pi^2\langle\mu^2\rangle,$$

where $\langle\mu^2\rangle$ is the mean-square displacement of a scatterer. Here we propose a scheme for defining pseudo *B* factors that reflect expected coordinate errors of each scatterer of an NMR template used for molecular replacement. These pseudo *B* factors are derived from atomic distances between an ensemble of individual NMR models and the averaged NMR model.

Independently, distance-derived *B* factors have been applied to an average, energy-minimized, NMR model (Anderson, Weiss & Eisenberg, 1996) that was used for the solution of the Er-1 crystal structure by molecular replacement (Weiss *et al.*, 1995). With this model the solution was found using the molecular-replacement modules of *X-PLOR* including PC refinement (Brünger, 1993). However, no analysis on the improvement of the molecular-replacement solution through application of distance-derived *B* factors has been described in this study.

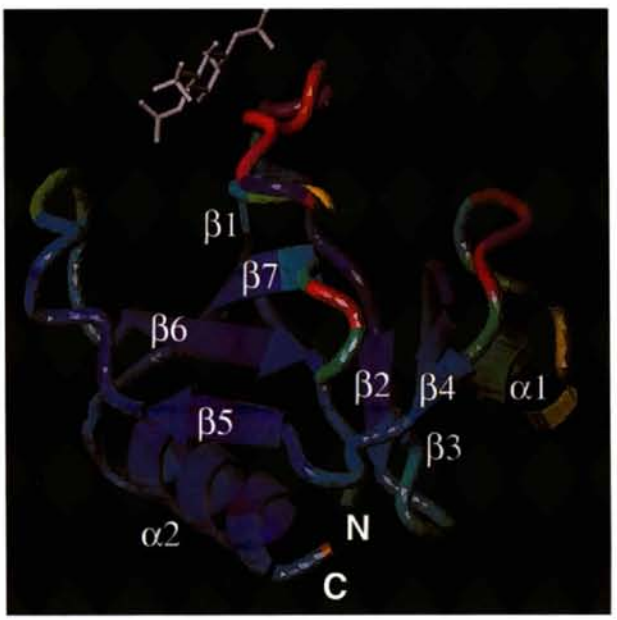

Fig. 2. Ribbon drawing of the PH:IP3 crystal structure. The β-strands are shown by arrows and the α-helices are shown by spirals. The average spatial distances between the individual NMR models and the averaged NMR model, which have been converted into distance-derived pseudo *B* factors, are imposed by colouring the ribbon. The colours gradually change from blue to red with increasing distances. The IP3 molecule is shown in white. The secondary-structural elements are located as follows: β1, 3–11; β2, 24–31; β3, 34–38; α1, 41–46; β4, 55–57; β5, 62–65; β6, 75–79; β7, 85–89; β2, 93–103.

* Structures that are used as templates for molecular replacement are denoted 'models'. This notation does not imply any quality statement.
† *R* factors and correlation coefficients are given as percentages These values follow the conventions of the program *AMoRe* (Navaza, 1994).

In this paper we demonstrate that the use of distance-derived *B* factors can strongly enhance the detectability of a correct structure solution by molecular replacement. The amount of improvement crucially depends on the correlation of the predicted deviations with the real deviations between each known NMR model and the unknown X-ray structure. Well predicted error estimates can be a key element for recognizing a correct solution from NMR models.

3. Methods

3.1. *Refinement of NMR structures*

The NMR structure calculations made use of ambiguous distance restraints for ambiguous NOE's. Three lists of upper distance limits were used to calculate the structures: (i) a list of manually chosen NOE's, which were mostly unambiguously assigned; (ii) a peak list generated from the raw NOE spectra by an automated peak-picking program (Aurelia, Bruker); (iii) a list of hydrogen bonds, specified as hydrogen–acceptor distances (2.2 Å) and donor–acceptor distances (3.2 Å). In case of doubt several choices for the acceptor were specified and the hydrogen-bond distance was treated like an ambiguous NOE.

The NOE volumes on the automatically generated peak list were calibrated and converted into distance classes by an automated calibration program (M. Nilges, unpublished results) interfaced with *X-PLOR* (Brünger, 1993). The upper distance limits were set to 6.0 Å for weak peaks; 3.3 Å for medium peaks; and for 2.7 Å for strong peaks. Lower bounds were set to the van der Waals limit or to 0. Floating assignment (Holak, Nilges & Oschkinat, 1989; Weber, Morrison & Hare, 1988) was used for methylene groups and propyl groups with distinguishable chemical shifts. NOE's for protons were treated like ambiguous NOE's.

Structures were generated using simulated-annealing protocols similar to those published (Nilges, Clore & Gronenborn, 1988; Brünger & Nilges, 1993) with an extended version of *X-PLOR* (Brünger, 1993). Modifications were introduced in the protocols to include the ambiguous distance restraints, floating assignments for prochiral groups, and a reduced representation for non-bonded interactions for a part of the calculation to increase efficiency.

50 structures were calculated starting from conformations with random backbone torsion angles. These structures were refined in an iterative procedure. In each iteration, the ambiguous NOE's of 20 structures were partially assigned by keeping only those assignments observed in the seven structures with the lowest energy for the next iteration of refinement. Once the initial list of NOE peaks with all possible assignments has been generated this refinement/assignment procedure is completely automatic and does not require any human intervention.

3.2. *Molecular replacement*

AMoRe (Navaza, 1994) as implemented in the *CCP*4 suite version 2.6 (Collaborative Computational Project, Number 4, 1994). Each molecular-replacement procedure included five steps called *SORTING* (sort observed data), *TABLING* (reorientation of model and preparation of continuous Fourier coefficients), *ROTING* (fast rotation function; Navaza, 1987), *TRAING* (translation function; Crowther & Blow, 1967; Harada, Lifchitz, Berthou & Jollès, 1981) and *FITING* (rigid-body least-squares refinement; Huber & Schneider, 1985). The ten available NMR models were superimposed on each other with the program *LSQKAB* (Kabsch, 1976) in identical orientation.

In *TABLING* the *NOROTATE* option was used to allow comparison of the molecular-replacement results. Optimization of some *ROTING* parameters resulted in the following values: resolution 15–3.5 Å; unit cell of model 50 $Å^3$; maximum integration radius, 19.0 Å; angular step size, 2.5°. For the published structure solution (Hyvönen *et al.*, 1995), a resolution range from 8 to 3.0 Å was used. Extending the low-resolution end from 8 to 15 Å considerably improved the detectability of the rotation solutions as estimated by the number of standard deviations above the mean and rank in the solution list. Default values for all other *ROTING* parameters were used. All peaks above 0.5 of the maximum peak were analysed. In *TRAING* and *FITING* default values were used except for the resolution range which was set to 15–3.0 Å.

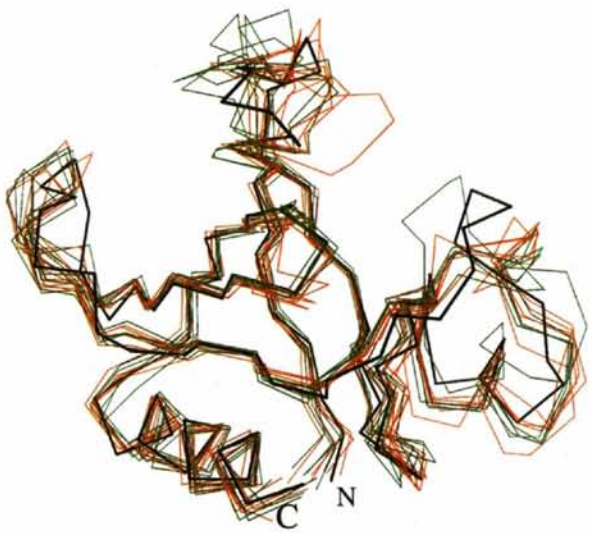

Fig. 3. Cα drawings of the superimposed ten NMR models and the X-ray structure of the PH domain (black). The five NMR models (*A*, *B*, *C*, *D*, *I*) with distance-derived *B* factors that gave correct translation-function solutions are shown in red, the remaining NMR models are shown in green. The orientation of the NMR models is identical to the orientation in Fig. 2.

Table 1. *Molecular-replacement statistics of NMR models with distance-derived B factors*

The ten NMR models are labelled *A–K* and ranked according to their weighted r.m.s.d.'s with the X-ray structure. The r.m.s.d.'s have been determined with the program *LSQKAB* (Kabsch, 1976) for all 864 matching non-H atoms of the PH domain. Only correct translation solutions were submitted to rigid-body refinement. All values of the rotation function, translation function and rigid-body refinement are taken from outputs of the *AMoRe* package (Navaza, 1994).

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*K*
Assessment of NMR models										
R.m.s.d. (Å)	3.06	3.14	3.14	3.23	3.21	3.29	3.28	3.23	3.31	3.21
Weight. r.m.s.d. (Å)*	1.68	1.73	1.73	1.76	1.77	1.81	1.82	1.83	1.85	1.92
R.m.s. (average NMR) (Å)	1.34	1.41	1.26	2.10	1.55	1.60	1.65	1.33	1.91	1.56
CC (distances)†	0.436	0.575	0.412	0.457	0.469	0.591	0.472	0.324	0.682	0.247
E(tot) (kcal mol^{-1})‡	487	525	470	547	496	554	490	495	553	544
Rotation										
α (°)	28.79	31.98	32.83	30.46	35.78	34.29	32.91	34.88	28.54	23.51
β (°)	46.32	42.69	45.33	45.18	44.49	46.15	43.63	49.54	48.21	50.72
γ (°)	205.29	203.95	201.84	204.37	204.12	206.23	204.73	201.76	203.48	206.98
Peak height (s.d.)	3.30	4.12	4.06	4.14	3.21	4.24	4.34	4.39	4.38	3.05
Rank	12	2	2	2	12	2	1	1	12	15
Translation										
x	0.0412	0.0416	0.0405	0.0407	0.0515	0.2306	0.7378	0.1510	0.0445	0.0386
y	0.7896	0.7970	0.7928	0.7865	0.2779	0.4835	0.2644	0.2194	0.7982	0.7926
z	0.2399	0.2419	0.2434	0.2479	0.2410	0.2074	0.0116	0.4301	0.2432	0.2425
CC (%)	30.7	28.8	28.3	25.7	22.2	21.6	22.0	22.8	28.8	26.6
R factor (%)	54.2	54.7	55.4	56.6	57.5	56.7	57.6	57.0	55.8	56.2
Rigid-body refinement										
α (°)	27.84	29.41	28.93	28.17					27.47	24.14
β (°)	45.81	43.16	45.03	46.01					47.39	50.86
γ (°)	205.61	204.67	204.62	206.34					205.01	206.72
x	0.0388	0.0427	0.0423	0.0409					0.0446	0.0368
y	0.7907	0.7994	0.7942	0.7881					0.7965	0.7950
z	0.2390	0.2443	0.2440	0.2479					0.2412	0.2396
CC (%)	33.5	32.9	33.3	28.1					30.3	26.6
R factor (%)	52.8	52.9	52.9	55.1					54.8	55.2

* Each spatial distance used for calculation of the weighted r.m.s.d. is weighted by $1/B$(dist). † CC(dist) is the correlation coefficient (%) between the spatial distances of each NMR model with the averaged NMR model and the X-ray structure. ‡ E(tot) = w(cov)E(cov) + w(exv)E(exv) + w(NOE)E(NOE). For further details see Nilges (1995).

4. Results

4.1. *Evaluation of NMR models*

Ten refined NMR models, denoted *A–K* in the following, were available for the X-ray structure solution of the PH:IP3 complex (Fig. 3). None of these structures showed large violations of the restraints. However, the conformational energies were relatively high (Table 1). More recently these structures were further refined after a fully automated version of the assignment software became available (M. Nilges, M. Macias & H. Oschkinat, unpublished results). This refinement revealed that three out of 568 NOE restraints that had been manually assigned to be unambiguous were overinterpretated. These NOE's affect residues around helix 1 (Fig. 2).

There are genuine differences between the X-ray structure and the NMR structures probably resulting from different physical environment, pH and the presence of the IP3 ligand in the crystal. The most notable structural differences are: (i) the salt bridge between R7 and E53, found in the X-ray structure, could not be detected in any of the NMR models; (ii) the long loop between strands β1 and β2 has virtually no long-range NOE's. Therefore, its conformation with respect to the rest of the molecule is not well determined in the NMR structures. Some residues of this loop specifically interact with the IP3 ligand (Hyvönen *et al.*, 1995). Sharpening of line widths of NOE's for residues R21, W23 and Q33 and some side chains involved in binding to IP3 has been observed in NMR measurements of the PH domain in the presence of either IP3 or L-α-glycerophospho-D-myo-inositol-4,5-bisphosphate. No conformational change of the loop β1–β2 could be detected upon ligand binding. The electron density of the crystal structure is unambiguous for the entire loop and implies a defined conformation of this loop; (iii) the X-ray structure shows an 1.5-turn helix between strands β3 and β4 (residues 41–48). This region is embedded in numerous contacts with symmetry-related molecules in the crystal. The NMR data also indicate a helix in this region. However, no slow exchange for amide protons of these residues could be detected suggesting

increased mobility of these residues. Furthermore, some key NOE's indicative for the presence of a helix are either absent or weak.

The quantitative evaluation of the ten NMR models has been divided to data that were available prior to the X-ray structure solution of the PH:IP3 complex and data that could only be obtained with the knowledge of the X-ray structure. The results are summarized in the first part of Table 1.

Prior to the crystallographic structure solution each of the ten NMR models within the ensemble was analysed by the total energy resulting from the molecular-dynamics refinement. The mean energy resulting from molecular dynamics refinement was 516±34 kcal mol^{-1} (1 kcal mol^{-1} = 4.184 kJ mol^{-1}) and ranged between 470 kcal mol^{-1} (model *C*) and 553 kcal mol^{-1} (model *I*). The mean r.m.s.d. between each model and the averaged model, calculated for all ten NMR models, is 1.57±0.29 Å for all 864 non-H atoms of the PH domain. The observed range is between 1.26 Å (model *C*) and 2.10 Å (model *D*). In these evaluations, model *C* scored best and was successfully used for the original structure solution (Hyvönen *et al.*, 1995).

After the X-ray structure became available the deviations between each NMR model and the X-ray structure were calculated. The r.m.s.d.'s of all ten models are above 3 Å (mean r.m.s.d. = 3.21±0.08 Å). If the superpositions are limited to Cα positions, the r.m.s.d.'s reduce to 1.3–1.8 Å (not shown in Table 1). Plots of the spatial differences of the Cα positions of a representative set of pairs of the X-ray structure and NMR models (*A*, *C*, *I*, *K*) are shown in Fig. 4. All four models show peak distances above 4 Å for the loops β1–β2, β5–β6 and the region β3–α1–β4 (*cf.* Figs. 2 and 3). These peak areas are also found in the other X-ray structure/NMR model pairs. Differences in the spatial deviations of these four models that influence the detectability of molecular-replacement solutions will be further discussed below.

With the solved X-ray structure we have calculated the linear correlation coefficients (CC) between the spatial distances of each NMR model (i, $i = A, \ldots, K$) with the average NMR model and each NMR model

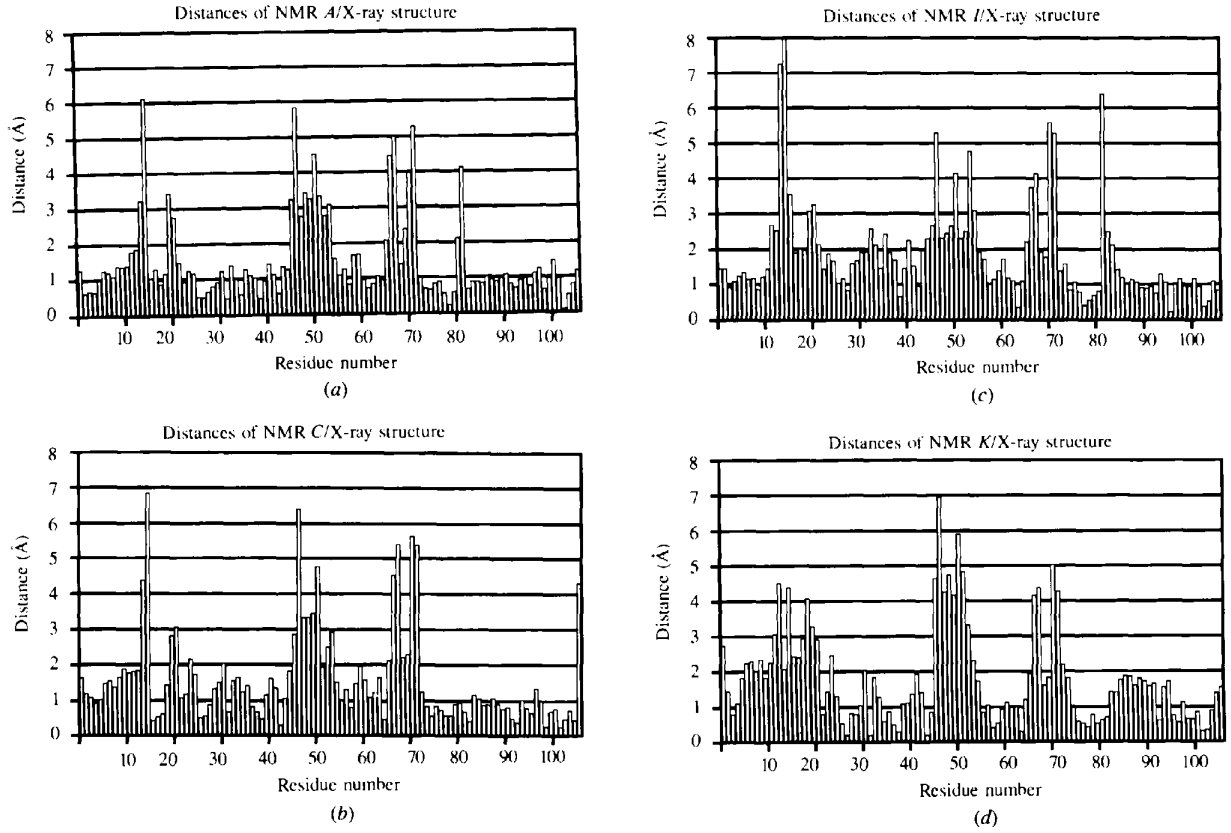

Fig. 4. Spatial distances between NMR models *A*, *C*, *I*, *K* (see Table 1) and the X-ray structure of the PH domain. For simplicity, only the distances of the Cα positions are shown. Model *A* has the lowest r.m.s.d. to the X-ray structure and gives the best statistics in the correct translation-function solution. Model *C* gives a correct translation-function solution and was used for the original X-ray structure solution (Hyvönen *et al.*, 1995). Model *I* has a relative high r.m.s.d. but the best correlation of expected distances with the observed distances to the X-ray structure. This model also gives a correct translation-function solution. Model *K* has the highest r.m.s.d. with the X-ray structure and the lowest fit of distances. This model gives a correct rotation, which, however, does the not satisfy the detection criteria.

(*i*) with the X-ray structure (Table 1). The CC's range between 24.7% (model *K*) and 68.2% (model *I*). The CC values show that spatial distances with respect to the NMR average have a potential for correct coordinate error prediction with respect to the X-ray structure to be solved.

4.2. *Distance-derived B factors*

Atomic *B* factors are normally derived from the mean-square displacement of atoms, $8\pi^2\langle u^2\rangle$. Here, the average atomic distances between each individual NMR model (*A–K*) and the average NMR model were converted into pseudo *B* factors by,

$$B(\text{dist}) = 10\langle \text{dist}(i_{\text{ave}},i_{\text{model}})^2\rangle,$$

where $\text{dist}(i_{\text{ave}},i_{\text{model}})$ is the averaged spatial distance between atom *i* of the averaged NMR model and all individual models. This equation resulted from an empirical optimization of its multiplicator (10) and exponent (2) and thus is related to the standard definition of atomic *B* factors. *B*(dist) values greater than 99 Å^2 were reset to 99 Å^2. Spatial differences of 0.2, 1 and 5 Å, result in *B*(dist) values of 0.4, 10 and 250 Å^2 (reset to 99 Å^2), respectively. Fig. 5 shows the averaged distances and distance-derived *B* factors applied to each of the ten NMR models. The largest deviations are observed for the two loops $\beta1$–$\beta2$ and $\beta6$–$\beta7$ and the region around helix $\alpha1$ (*cf.* Fig. 2) resulting in distance-derived *B* factors higher than 50 Å^2 for residues 13–21, 51–53 and 82, respectively. The refined *B* factors of the crystal structure of the PH:IP3 complex are also shown for comparison. The correlation between the refined *B* factors of the X-ray structure and distance-derived *B* factors is 58% on all Cα positions. This value indicates correlation between the atomic displacements in the refined X-ray structure and coordinate deviations within the used set of ten NMR templates.

The r.m.s.d.'s between each NMR model and the X-ray structure were recalculated by weighting each spatial atomic distance by 1/*B*(dist) (Table 1). The average weighted r.m.s.d. is 1.79±0.07 Å, slightly more than half the mean value of the unweighted r.m.s.d. The weighted r.m.s.d.'s are more relevant for evaluating the molecular-replacement solutions because they account for the contributions of scatterers weighted by distance-derived *B* factors. The models *A–K* have been ranked according to their weighted r.m.s.d. values.

In the Fourier transformations, however, the structure-factor amplitude contribution also depends on reciprocal resolution (or Bragg angle) in the $\exp[-(\sin^2\theta/\lambda^2 B)]$ term where θ is the Bragg angle and λ the wavelength of the X-rays. These effects are illustrated by a few examples in Table 2. At 3.0 Å resolution, the high-resolution limit used in the translation function, the contribution of an atomic scattering factor *f* with *B* = 1 Å^2 is 0.98*f*. The contribution of an atomic scattering factor with *B* = 99 Å^2, in contrast, is only 0.10*f* at 3.0 Å resolution.

4.3. *Molecular-replacement tests*

We have tested three different versions of each NMR model *A–K*: (i) uniform *B* factor *s* of 20 Å^2 (the average *B* factor for all main-chain atoms in the X-ray structure is 20.1 Å^2; Hyvönen *et al.*, 1995); (ii) distance-derived *B* factors as described in the previous section; (iii) *B*

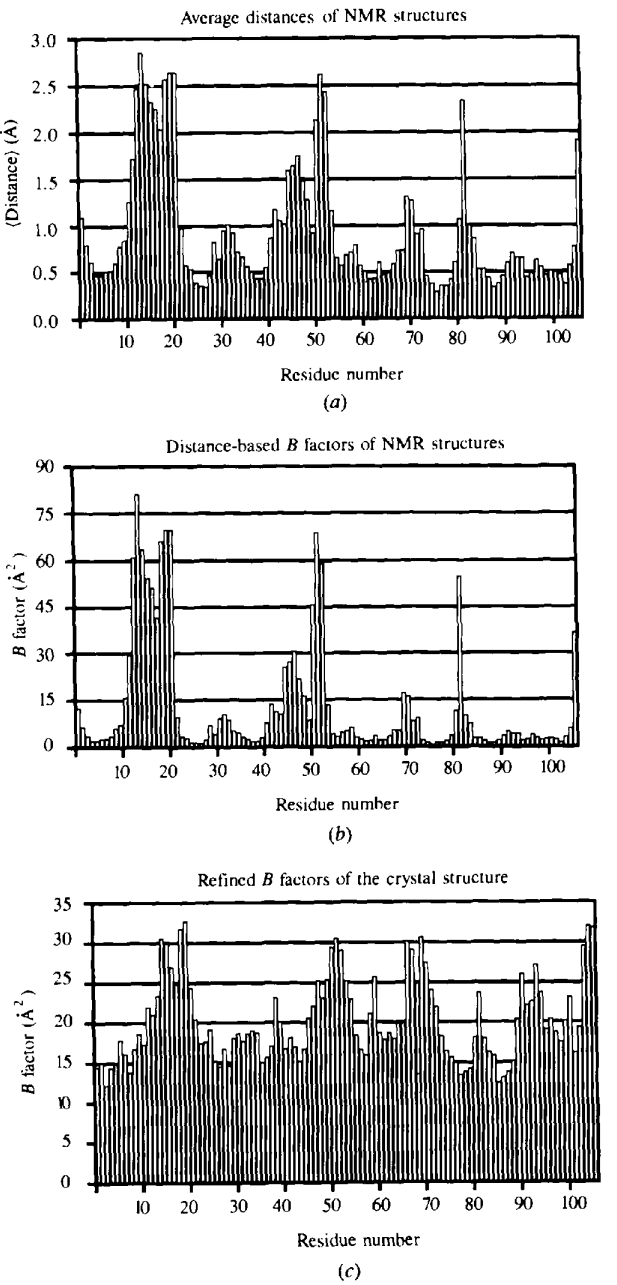

Fig. 5. (*a*) Average atomic distance of ten NMR models and the averaged NMR model. (*b*) Distance-derived *B* factors applied to NMR templates. Please note that the same distance-derived *B* factors have been applied to each NMR model. (*c*) Atomic *B* factors of the X-ray structure of the PH:IP3 complex (Hyvönen *et al.*, 1995). For simplicity only the values of the Cα positions are shown.

Table 2. *Structure-factor amplitudes as function of B factors and resolution*

The values are calculated from $\exp[-(\sin^2\theta/\lambda^2)B]$.

Resolution (Å)	8.0	4.5	3.0
B factor (Å^2)			
1.0	1.00	0.99	0.98
20.0	0.94	0.78	0.63
99.0	0.72	0.36	0.10

Table 3. *Molecular-replacement statistics of NMR model versions*

Version (*B* factor)	Uniform	Distance	Simulated
Rotation function			
Correct solutions	8	9	10
⟨relative peak height⟩ (σ)	4.08	4.11	4.51
Translation function			
Correct solutions	4	6	9
⟨*R* factor⟩ (%)*	57.9	55.6	55.2
⟨ΔR factor⟩ (%)*†	−1.2	−1.7	−2.9
⟨*R* factor⟩ (%)‡		55.5	55.7
⟨ΔR factor⟩ (%)†‡		−1.8	−2.4
Detectable solutions (*R*-factor test)§	0	2	3
⟨CC⟩ (%)*	26.5	27.9	31.9
⟨ΔCC⟩ (%)*†	4.1	5.6	8.5
⟨CC⟩ (%)‡		28.2	30.3
⟨ΔCC⟩ (%)†‡		6.0	7.2
Detectable solutions (CC test)¶	1	5	7
Rigid-body refinement			
⟨*R* factor⟩ (%)*	54.8	54.0	52.7
⟨ΔR factor⟩ (%)*†	−2.0	−2.3	−3.4
⟨*R* factor⟩ (%)‡		54.1	53.0
⟨ΔR factor⟩ (%)†‡		−2.2	−3.2
⟨CC⟩ (%)*	38.8	36.0	44.1
⟨ΔCC⟩ (%)*†	4.7	6.0	8.7
⟨CC⟩ (%)‡		35.8	43.1
⟨ΔCC⟩ (%)†‡		6.0	8.2

* Statistics for solutions which were correct for all three model versions (*B, C, D, I*). † ⟨value⟩ (best solution) − ⟨value⟩ (incorrect solutions); ⟨value⟩ (best solution) is the top-ranking solution; ⟨value⟩ (incorrect solutions) is the average of solutions 2–11 of the sorted list. ‡ Statistics for all correct solutions. § Threshold for successful detection: $R = 55\%$. ¶ Threshold for correct solution: $\Delta CC = 5\%$.

factors derived from the atomic distances between each NMR model and X-ray structure, which are referred to as 'simulated' below. The last version of each NMR model was used as simulation to explore the limits of distance-derived *B* factors. The statistics of the three molecular-replacement steps (rotation function, translation function, rigid-body refinement) were analysed for each NMR model. The full statistics including the solution values are given for the models with distance-derived *B* factors in Table 1. The average values for all model versions are shown in Table 3.

4.4. *Rotation function*

Correct solutions for each version of the ten NMR models except model *A* with uniform *B* factors have been obtained in the fast rotation function (Fig. 6). In general, a peak height of 4 standard deviations above the mean was required to detect the correct solution at rank 1 or 2 of the peak list. Using a peak height of 4σ as threshold all three versions of six models (*B*, *C*, *F*, *G*, *H*, *I*) were detectable. The relative peak heights of the correct solutions (all models except *A*) improved on average with respect to the versions with uniform *B* factors: $\Delta 0.03\sigma$ for the versions with distance-derived *B* factors and $\Delta 0.43\sigma$ for the versions with simulated *B* factors. We conclude that at the step of the rotation function there was no significant improvement by distance-derived *B* factors. The versions with simulated *B* factors, however, demonstrate that good estimates of *B* factors have a strong potential to improve the significance of a correct rotation-function solution.

4.5. *Translation function and rigid-body refinement*

Correct translation solutions were obtained with four, six and nine NMR models of the versions with uniform, distance-derived and simulated *B* factors (Fig. 7). The correct solutions were submitted to rigid-body refinement. The significance of the translation function and rigid-body refinement solutions was measured in terms of CC and *R* factors as given in the *AMoRe* package (Navaza, 1994). The differences between CC and *R* factor of the best solution and the next ten solutions were also used as a measure of statistical significance.

In the translation function the average *R* factor for the model versions with uniform *B* factors that gave correct solutions is above random, 57.9% (Table 3). The average *R* factors of the correct solutions of the model versions with distance-derived and simulated *B* factors are between 55 and 56%. Fig. 7(*a*) shows that model *A* gives the lowest *R* factors, 54.0% (distance-derived *B* factors) and 52.6% (simulated *B* factors). The *R* factors of all 'best' incorrect solutions are above random

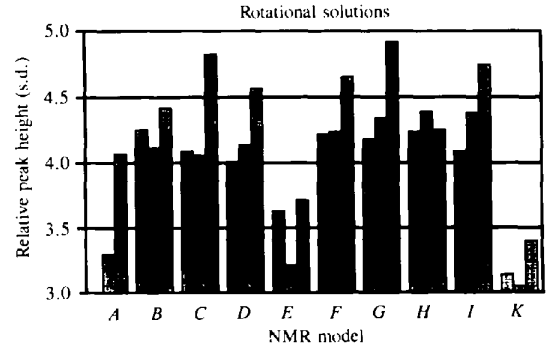

Fig. 6. Correct rotation-function solutions of three versions of NMR models *A*–*K* (see Table 1). Models with uniform *B* factors, left bars; models with distance-derived *B* factors, middle bars; models with simulated *B* factors, right bars. Please note that there is no bar for model *A* with uniform *B* factors because the correct rotation solution was not detected.

(57%). The average *R* factors of the correct solutions are, however, only 1.2% (uniform *B* factors) to 2.4% (simulated *B* factors) lower then those of the next ten incorrect solutions (Fig. 7*b*). In the rigid-body refinement step the average *R* factors of the correct solutions drop to 54.8% (uniform *B* factors), 54.1% (distance-derived *B* factors) and 53.0% (simulated *B* factors). An *R* factor of about 52% appears to be the limit for the available NMR models and relates to their deviations from the X-ray structure. The *R*-factor differences between correct and incorrect solutions increase, however, only by less than 1% for all three model versions.

In contrast to *R* factors there is no objective random estimate for CC's. In this study the CC's of correct and incorrect solutions of the model versions with uniform *B* factors were slightly higher than those of the other model versions. Applying high *B* factors to uncertain parts of the model results in a *de facto* reduction of the completeness of the models. This probably has a reducing effect on the absolute CC values regardless of the correctness of a solution. For this reason the discussion on CC's will focus on CC differences (Fig. 7*d*) rather than on absolute CC values (Fig. 7*c*). The CC differences were estimated from the best solution and the average CC of the following ten solutions of the sorted solution list.

In the translation function the CC differences are 4.1, 6.0 and 8.7% for models with uniform, distance-derived and simulated *B* factors, respectively. The differences between 'best' incorrect solutions and the following ten solutions were below 3% in all test cases. Rigid-body refinement increases the absolute CC's by about 10% (Table 3). However, the improvement of the average CC differences is below 1%.

5. Discussion

5.1. *Detection of molecular-replacement solutions*

We have shown that in the given test case four, six and nine out of ten NMR models give a correct translation-function solution, depending on whether they contain uniform *B* factors, distance-derived *B* factors or simulated *B* factors. These solutions were detected using the knowledge about the correct rotation angles and translations. The statistical significance of the solutions has not been used as a success criterion so far.

No straightforward success criterion exists for rotation-function solutions. It is quite common in

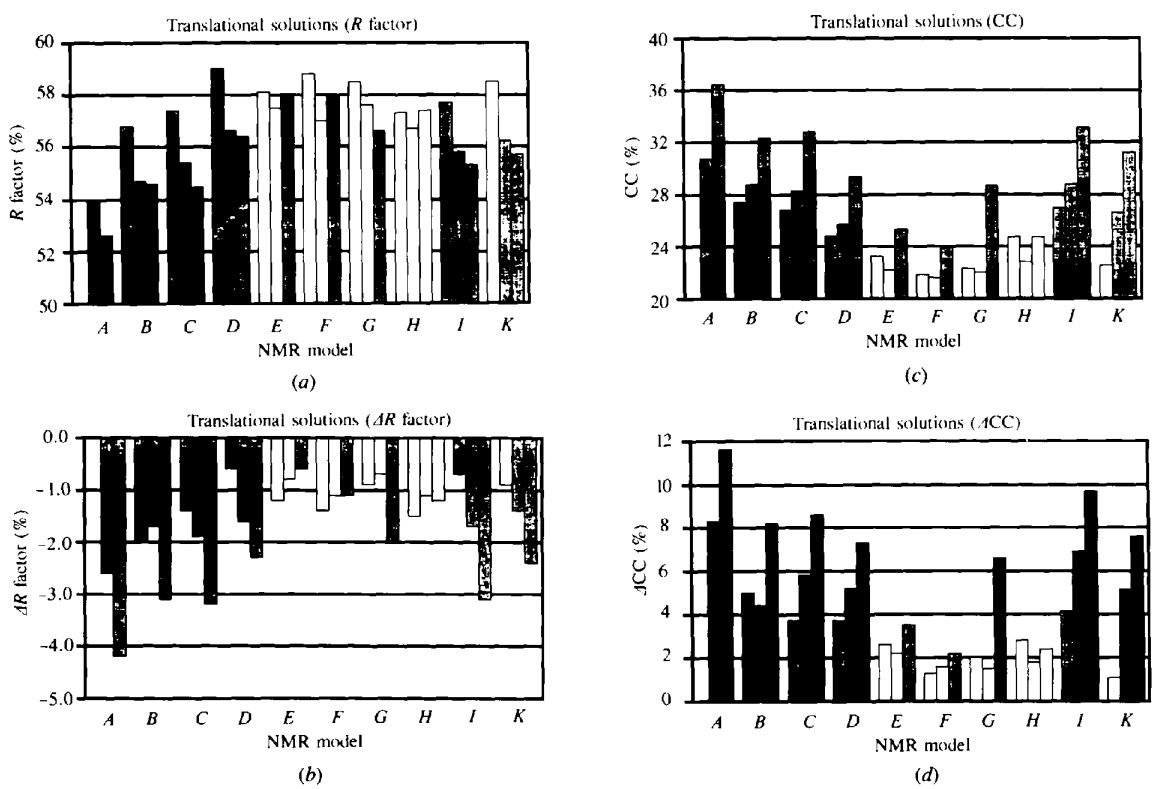

Fig. 7. Translation-function solutions of NMR models *A*–*K* (see Table 1). The bars of correct and detectable solutions are in black, bars of correct but undetectable solutions are in grey, bars of incorrect solutions are white. There is no bar for model *A* with uniform *B* factors due to missing rotation solution. (*a*) *R* factors; (*b*) *R*-factor differences between the best, top ranking solutions and the average *R* factor of solutions 2–11; (*c*) CC; (*d*) CC differences between the best, top-ranking solution and solutions 2–11.

molecular replacement that the correct solution is not top ranking. In our experience, it is most important to have the possibility for rapid screening of a large number of solutions. The program *AMoRe* (Navaza, 1994) is well suited for this task.

The statistical significance of translation-function solutions has been analyzed here by the *R* factor and CC. In a real experiment without the knowledge of the solution confidence criteria for successful detection of a solution are required. Here we use threshold values for *R* factor = 55.0% (2% below the random *R* factor) and a CC difference of 5% between the top-ranking solution and the next ten solutions. If applying the *R*-factor test, none of the solutions with uniform *B* factors, two solutions (models *A*, *B*) with distance-derived *B* factors and three solutions (models *A*, *B*, *C*) with simulated *B* factors pass (Fig. 7*a*). In the CC test, one model (*B*) with uniform *B* factors is successful, five models (*A*, *C*, *D*, *I*, *K*) with distance-derived *B* factors and seven models with simulated *B* factors (*A*, *B*, *C*, *D*, *G*, *I*, *K*) are successful (Fig. 7*d*).

These confidence tests demonstrate that application of distance-derived *B* factors was an absolute requirement to find any statistically relevant molecular-replacement solution from the available NMR models. The translation solution of the only model with uniform *B* factors, model *B*, that passes the CC confidence test still has a practically random *R* factor of 56.8%.

5.2. *Performance of individual models*

Four of six NMR models with distance-derived *B* factors that give correct solutions (*A*–*D*) have the lowest weighted r.m.s.d.'s (1.68–1.6 Å) from the X-ray structure (Table 1). Only models *A* and *B* pass both confidence tests for detection. The spatial distances of all Cα positions except some local loop regions (shown for model *A* in Fig. 4*a*) are below 1.5 Å. Model *I*, which also gives a correct solution, has a higher weighted r.m.s.d. (1.85 Å). However, for this model the correlation of the expected spatial distances with the observed distances with respect to the X-ray structure (68.2%) is better than for all other NMR models. The good fit of the predicted distances might be a reason why relative high spatial differences between residues 15 and 55 (Fig. 4*c*) do not disguise a detectable solution in the translation function. The translation solution of model *K* is also correct but not detectable. In this model the spatial differences of the C-terminal helix are higher than in the other shown models (Fig. 4*d*).

Neither of the two parameters available prior to the X-ray structure solution, total energies from molecular dynamics refinement and r.m.s.d. from the average NMR model, can be correlated with success rates of molecular-replacement solutions. In the original structure solution of the PH:IP3 complex, model *C* was used because it had the lowest energy and closest fit to the averaged NMR model.

5.3. *Recommendations*

Our recommendations for molecular replacement from NMR models are as follows.

Use all available NMR models unless that there are indications for diverging overall qualities (*e.g.* number of NOE violations).

In our experience the absolute *R* factor and CC differences between the top-ranking translation solution and a number of following solutions are the most sensitive detection criteria. These values only slighty improve in the rigid-body refinement.

Introduction of pseudo *B* factors in NMR models that account for expected coordinate errors strongly enhances the detectability of molecular-replacement solutions. We have derived expected atomic coordinate errors from average spatial distances between individual NMR models and an averaged NMR model. More sophisticated approaches that could *e.g.* include the number of NOE's and band widths of NOE's per residue might further improve the fit.

6. Outlook

NMR structures provide an important pool of structural templates that can be used for molecular-replacement solutions of X-ray structures. It is expected that the number of solved NMR structures will increase tremendously in the near future. NMR structures are normally available as an ensemble of related structures exploring the conformational space which is constrained by the measured NOE's. The wealth of this information can be exploited to calculate pseudo *B* factors derived from expected coordinate errors. In our test case, the structure solution of the PH domain:IP3 complex, use of distance-derived *B* factors was essential to solve the structure. This molecular-replacement solution used NMR templates with identical sequence. To what extent NMR models with non-identical sequences are feasible as templates for molecular replacement remains to be investigated.

Another application of distance-derived *B* factors could be in the field of molecular replacement with X-ray models. Coordinate deviations mostly depend on the amount of sequence similarity between the structural template and the structure to be solved (Chothia & Lesk, 1986). The structural deviations of identical structures are generally low (below 0.5 Å r.m.s.d.) if they are refined.

A requirement would be the availability of more then one template structure which is homologous with the sequence of the structure to be determined. Pseudo *B* factors could be derived from spatial distances of superimposed structures and positional differences of the amino-acid sequences. We would expect increasing accuracy of expected coordinate errors with increasing number of homologous template structures.

We thank Maria Macias and Hartmut Oschkinat for their contribution in original structure determination of the NMR structure of the β-spectrin PH domain, Marko Hyvönen for his involvement in the X-ray structure solution of the PH:IP3 complex and Matti Saraste for the overall support and ideas. We further thank Kristina Djinovic, Ralf Ficner and Paul Tucker for critical reading of the manuscript. The coordinates and observed structure-factor data set of the PH:IP3 complex have been deposited with the Protein Data Bank.*

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1BTN, R1BTNSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: LI0225).

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.

Billeter, M. (1992). Quart. *Rev. Biophys.* **25**, 325–377.

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

Brünger, A. T. (1993). *X-PLOR Version* 3.1, Yale University Press, New Haven, CT, USA.

Brünger, A. T., Campbell, R. L., Clore, G. M., Gronenborn, A. M., Karplus, M., Petsko, G. A. & Teeter, M. M. (1987). *Science*, **235**, 458–460.

Brünger, A. T. & Nilges, M. (1993). *Quart. Rev. Biophys.* **26**, 49–125.

Chothia, C. & Lesk, A. M. (1986). *EMBO J.* **5**, 823–826.

Cohen, G. B., Ren, R. & Baltimore, D. (1995). *Cell*, **80**, 237–248.

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

Crowther, R. A. & Blow, D. M. (1967). *Acta Cryst.* **23**, 544–548.

Harada, Y., Lifchitz, A., Berthou, J. & Jollès, P. (1981). *Acta Cryst* A**37**, 398–406.

Hendrickson, W. A. & Wüthrich, K. (1994). Editors. *Macromolecular Structures* 1994. London: Current Biology.

Holak, T. A., Nilges, M. & Oschkinat, H. (1989). *FEBS Lett.* **242**, 218–224.

Huber, R. & Schneider, M. (1985). *J. Appl. Cryst.* **18**, 165–169.

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.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

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

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.

MacArthur, M. W., Laskowski, R. A. & Thornton, J. M. (1994). *Curr. Opin. Struct. Biol.* **4**, 731–737.

Macias, M. J., Musacchio, A., Ponstingl, H., Nilges, M., Saraste, M. & Oschkinat, H. (1994). *Nature (London)*, **369**, 675–677.

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

Navaza, J. (1987). *Acta Cryst.* A**43**, 645–653.

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

Nilges, M. (1995). *J. Mol. Biol.* **245**, 645–660.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988). *FEBS Lett.* **229**, 129–136.

Pawson, T. (1995). *Nature (London)*, **373**, 573–580.

Saraste, M. & Hyvönen, M. (1995). *Curr. Opin. Struct. Biol.* **5**, 403–408.

Wagner, G. (1993). *J. Biomol. NMR*, **3**, 375–385.

Weber, P. L., Morrison, R. & Hare, D. (1988). *J. Mol. Chem.* **204**, 483–487.

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.

Wüthrich, K. (1995). *Acta Cryst.* D**51**, 249–270.