Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

LAFIRE: software for automating the refinement process of protein-structure analysis

Min Yao,* Yong Zhou and Isao Tanaka

Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

Correspondence e-mail: yao@castor.sci.hokudai.ac.jp

Received 7 July 2005
Accepted 23 November 2005

Manual intervention is usually required in the multiple rounds of refinement of protein crystal structures, including linking and/or extending the fragments of the initial model and rebuilding (fitting) ill-matched residues using computer-graphics software. Such manual modification is both time-consuming and requires a great deal of expertise in crystallography. Consequently, the refinement process becomes the bottleneck for high-throughput structure analysis. A program, *Local correlation coefficient-based Automatic FItting for REfinement* (*LAFIRE*), has been developed to achieve manual intervention-free refinement. This program was designed to perform the entire process of protein structural refinement automatically using the refinement programs *CNS*1.1 (*CNS* v.1.1) or *REFMAC*5. The automatic process begins from an initial model, which can be approximate, fragmentary or even only main-chain, and refines it to the final model including water molecules, controlled by monitoring the $R_{\rm free}$ factor. More than 30 structures have now been refined successfully in a fully or semi-automatic manner within a few hours or days using *LAFIRE*.

1. Introduction

Recent advances in protein crystallography, such as the development of the multiwavelength anomalous diffraction method using selenomethionine cloned proteins and high-throughput protein preparation and crystallization, have made it possible to solve protein structures very rapidly. A number of automatic and semi-automatic programs are available for individual steps in structure determination. Moreover, automated structure-solution software suites such as *PHENIX* (Adams *et al.*, 2002), *ACrS* (Brunzelle *et al.*, 2003), *CRANK* (Ness *et al.*, 2004), *ELVES* (Holton & Alber, 2004) and *SGXPro* (Fu *et al.*, 2005) *etc*. are also available. However, the refinement process still requires human intervention and is therefore the most time-consuming step requiring the most skill in the whole process of structure analysis.

Structural refinement of macromolecules is difficult because the observation-to-parameter ratio (typically about 3–5) is usually low. A number of different methods have been developed to overcome this difficulty, such as stereochemical restrained least squares (Hendrickson, 1985), molecular dynamics including simulated annealing (Brünger *et al.*, 1987) and maximum likelihood (Murshudov *et al.*, 1997). However, in most cases manual intervention is required as a complementary function to the refinement algorithms; the model is fitted into the electron-density map between rounds of refinement using computer-graphics programs such as *O* (Jones *et al.*, 1991), *TURBO-FRODO* (http://afmb.cnrs-mrs.fr), *XtalView* (McRee, 1999), *MAIN* (Turk & Guncar,

1999) or *QUANTA* (Accelrys Inc.). Moreover, although automatic model-building programs such as *ARP/wARP* (Perrakis *et al.*, 1999; Morris *et al.*, 2002), *RESOLVE* (Terwilliger, 2002, 2003), *MAID* (Levitt, 2001) and *CAPRA/TEXTAL* (Ioerger & Sacchettini, 2002) *etc.* are now available, these programs usually provide only 60–95% of the whole structure at ~3 Å resolution, depending on the quality of the experimental electron-density map. Therefore, it is still necessary to complete the initial model by linking and/or extending the fragments during refinement.

For rapid and effortless structure analysis, we have developed a new automatic refinement software package, *LAFIRE* (*Local correlation coefficient-based Automatic FItting for REfinement*), designed to perform the whole process of protein structural refinement automatically with the refinement programs *CNS*1.1 (*CNS* v.1.1; Brünger *et al.*, 1998) or *REFMAC*5 (Murshudov *et al.*, 1997). The automatic process begins from an initial model that can be approximate, fragmentary or even only main chain, which is then refined to the final model, including water molecules, which can then be subjected to a final check by crystallographers using a graphics program. Here, we present an overview and processing strategy of *LAFIRE* and provide results obtained by its application to actual samples.

2. Overview of *LAFIRE*

The automatic refinement system *LAFIRE* was developed on SGI computers running IRIX 6.5; versions for Linux (Red Hat 9.0, Fedora 2, Mandrake) are also available. The program consists of four parts: partial model building, model modification (fitting) including evaluation of the current model, a graphic monitor system (*LAFIRE_moleview*) and a process-control system that includes interfaces with refinement programs (*CNS*1.1, *REFMAC*5; Fig. 1). The program for building and fitting was written in C and Fortran. A C-shell script is used as a job-control system and as an interface between programs.

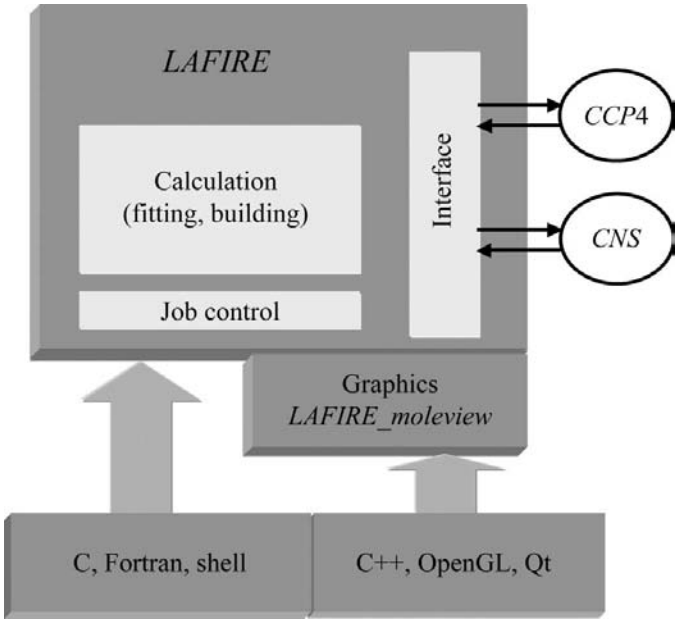

Figure 1
Architecture of the *LAFIRE* program.

Fig. 2 shows the general flowchart of the refinement process by *LAFIRE*. The program checks and replaces amino acids in the initial model with reference to the sequence file. This function is used for unassigned models, such as polyalanine/glycine/serine models or models obtained by molecular replacement. *LAFIRE* then builds the missing parts, such as loops or terminal regions, fits the model to the experimental or σ_A-weighted $2F_o - F_c$ maps and runs the refinement programs. The processes of *LAFIRE* are controlled by monitoring the R_{free} factor and are repeated until there is no further improvement.

2.1. Building and fitting

Owing to their lack of regular structure, loops and terminal regions are more difficult to build automatically compared with α-helix and β-strand regions. Furthermore, the electron-density maps in these areas are usually less clear owing to conformational flexibility and automatic model-building programs often fail to build these irregular parts. *LAFIRE* was designed to construct these missing regions iteratively during refinement. Assuming that the initial partial model is essentially correct with the exception of a small number of residues at both ends of the fragments, we have developed a new algorithm for linking and/or extending the current fragments (map-segment pruning method; Zhou *et al.*, 2006). In this algorithm, the chain connections are first analyzed for existing fragments and map segments for missing parts of the model are extracted as continuous regions of electron density higher than a given threshold value. These pruned map segments include the C^α positions of the terminal residues of existing fragments. Then, based on the connection information of existing fragments and C^α positions as anchor points, missing residues are built individually using the peptide (C^α_i, C_i, O_i, N_{i+1}, C^α_{i+1}) as a unit. As the whole process is run automatically, *LAFIRE* can build increasing numbers of residues during the progress of refinement.

To perform fitting, ill-matched regions should be detected first. In the present study, we developed a modified grouped local correlation coefficient (GLCC) as shown in (1), which is similar to that reported by Pavelcik *et al.* (2002). GLCC_i for residue i is calculated separately for the main-chain group including C^β (N, C^α, C^β, C, O) and the side-chain group excluding C^β,

$$\mathrm{GLCC}_i = g_i \frac{\langle \rho_{\mathrm{obs}} \rho_{\mathrm{cal}} \rangle_i}{[(\langle \rho^2_{\mathrm{obs}} \rangle_i)(\langle \rho^2_{\mathrm{cal}} \rangle_i)]^{1/2}}, \quad (1)$$

$$g_i = \frac{1}{N} \int_{\rho_{\mathrm{obs}\_m}}^{\langle \rho_{\mathrm{obs}} \rangle_i} h(\rho_{\mathrm{obs}})\, \mathrm{d}\rho_{\mathrm{obs}}. \quad (2)$$

Here, g_i is a weighting factor that accounts for the quality of the density map of the ith residue, $h(\rho_{\mathrm{obs}})$ is the number of grid points that have a density value of ρ_{obs}, ρ_{obs_m} is the minimum

density value of the map and N is the number of grid points in the map. The average value, $\langle\rho_{\rm obs}\rangle_i$, is calculated from grouped atoms of the ith residue. Integration of (2) is performed in the protein region estimated from the current model (10 Å beyond the molecular boundary). (1) has the advantage of being insensitive to map quality, especially for poor or low-resolution maps.

Based on the GLCC, the main chain and side chains are modified separately, except for proline residues, which are included as rigid bodies. While the main chain is fitted as a rigid body (Luo *et al.*, 1992), the side chain is fitted in several groups corresponding to rotations of χ_1, χ_2, χ_3, ... The *cis*-peptide conformation is also considered for the *X*–Pro peptide bond.

2.2. Multi-level strategy for building and fitting

In most cases, linking and/or extending the current fragments using the map-segment pruning method and fitting based on the grouped local correlation coefficients (as evaluation functions) are effective. However, as there is a wide degree of variation in model quality during refinement, no single algorithm can deal effectively with all possible variations. To overcome this difficulty, we adopted a multi-level strategy for building and fitting, *i.e.* a slightly different algorithm is applied to building and fitting in each stage of the refinement process.

On the assumption that the initial fragments are essentially correct with the exception of both ends of each fragment, in the first level of building ($N_{\rm mls}$ = 1) the missing parts are constructed based on the pruned map that excludes segments occupied by all atoms in the current model except terminal residues of the fragments. In the second level of building ($N_{\rm mls}$ = 2), the map is pruned only for the main chain including the C^{β} atoms of the current model, because side chains (especially large side chains) are frequently misplaced in the electron-density map. Indeed, side chains positioned erroneously in the main-chain region often lead to failure in main-chain tracing. In the third level ($N_{\rm mls}$ = 3), the program attempts to build short missing fragments of 1–3 residues even when the density map is poor or the space is insufficient to accommodate these short missing fragments. In the latter case, the peptide backbone of the short missing fragment is built on a scaled size corresponding to the space of the missing part. This part is then corrected by fitting and refinement. These three levels of building are carried out in order during the refinement cycles until the $R_{\rm free}$ factor shows no further changes (Fig. 2).

Two levels are prepared for the fitting process; level 1 fitting is employed during the refinement process in the first and second levels of building, while level 2 fitting is performed during the refinement process in the third level of building. Level 1 fitting is the standard method used by *LAFIRE* as described above (§2.1). In this fitting level, the program attempts to fit main-chain torsion angles (φ, ψ) of all non-glycine and non-proline residues at least into the allowed region of the Ramachandran plot (Ramachandran *et al.*, 1963). In level 2 fitting, the program attempts to fit the main-chain torsion angles of all non-glycine and non-proline residues into the favoured region of the Ramachandran plot. Flipping of the peptide plane is also employed in this level.

2.3. Other protocols

The *LAFIRE* refinement procedure was designed based on our experience with refinement of over 40 structures using the program *CNS*1.1; the cross-validation $R_{\rm free}$ factor is used as a criterion to monitor refinement. The standard refinement protocol consists of positional refinement followed by *B*-factor refinement in each round. Rigid-body refinement is carried out in the first round of the refinement process. When using *CNS*1.1, simulated annealing from 2000 K can be used in the first two rounds. These refinement procedures can be adapted to the user's own problem.

For MIR, MAD or SAD methods, the experimental map used to build the initial model is used for building and fitting at the initial rounds of the first and second level of building in our program. After the structure has been refined to an $R_{\rm free}$ factor of below 30% or to convergence, the $\sigma_{\rm A}$-weighted $2F_{\rm o} - F_{\rm c}$ map is used for building and fitting. In the current version of *LAFIRE*, a $\sigma_{\rm A}$-weighted $F_{\rm o} - F_{\rm c}$ map is used only to check fitting, but in future versions it will be used to aid building (see §4).

As protein molecules related by non-crystallographic symmetry (NCS) may

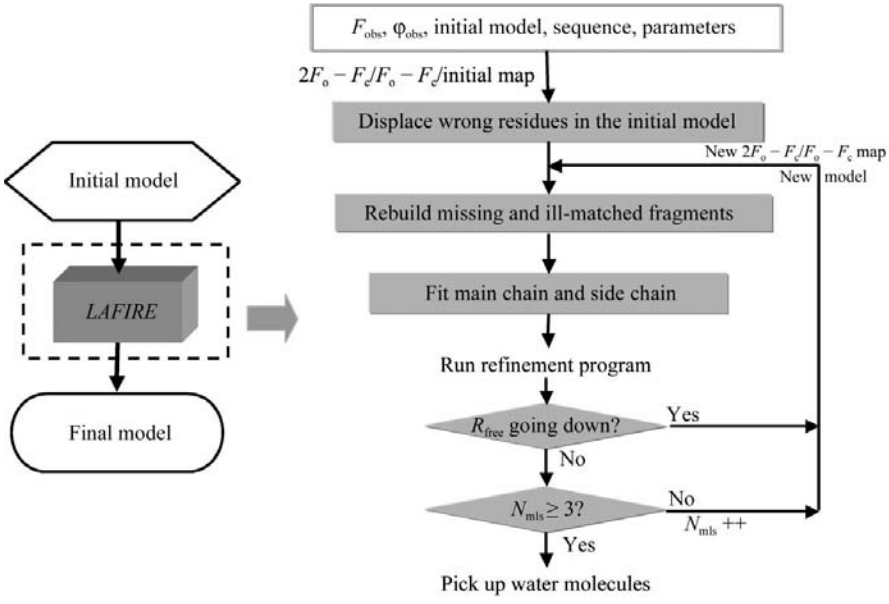

Figure 2
Flowchart of the *LAFIRE* procedure. $N_{\rm mls}$ is a counter for controlling multi-level strategy.

Table 1
Result of full or semi-automatic refinement using *LAFIRE* with *CNS*1.1 for four tests and 14 applications.

				LAFIRE model				Deposited model	
Protein	Resolution (Å)	N_{res}†	Initial model N_{res}‡ (%)	N_{res}§ (%)	R_{free}/R (%)	cns_b (Å)/ cns_a¶ (°)	Manual intervention††	N_{res}§ (%)	R_{free}/R (%)
1ub9 (PH1061)	2.05	100	92.0/A	100 (52)	24.5/21.5	0.005/1.08	No	100 (57)	24.2/21.6
1v71 (IPMI)	1.98	163 × 2	97.5/m	99.4 (205)	26.4/23.1	0.005/1.26	No	99.7 (294)	24.7/20.6
1iz6 (aIF5A)	2.0	138 × 3	97.3/m	99.3 (219)	24.6/20.0	0.005/1.32	No	98.3 (279)	23.6/18.5
1ucg (MC1)	1.65	190 × 2	100/M	100 (100)	21.4/19.0	0.005/1.25	No	100 (235)	19.6/18.1
1uly (PH1932)	2.5	192	99.0/m	99.0 (25)	28.7/21.7	0.007/1.14	No	100 (25)	26.8/20.4
1v7b (CGL2612)	1.85	177 × 2	97.2/m	98.6 (225)	24.3/21.6	0.005/1.02	No	98.0 (154)	24.1/20.8
1vaj (PH0010)	1.82	205	99.0/A	99.0 (156)	23.5/20.4	0.005/1.42	No	99.0 (165)	23.2/20.2
1ve0 (ST2072)	2.0	134	61.2/R	100 (66)	24.9/20.3	0.006/1.24	No	100 (45)	22.2/18.1
1v7o (PH0574)	2.62	157 × 2	98.1/R + m	98.1	30.5/27.4	0.009/1.45	No	98.1 (82)	27.6/23.4
1vgj (PH0099)	1.94	184	78.8/A	99.5 (94)	27.6/23.2	0.006/1.18	Yes	100 (41)	26.5/21.6
1wls (PH0066)	2.16	328 × 2	64.0‡‡/A + m	99.5 (137)	26.3/22.8	0.007/1.63	Yes	100 (250)	25.3/21.1
1wle (merRS)	1.65	501 × 2	89.2/A	92.6 (563)	23.9/22.3	0.005/1.26	No	93.2 (614)	22.6/21.0
1wzz (CMCax)	2.3	322	67.7/R	95.0 (129)	23.2/19.5	0.007/1.29	No	98.8 (201)	21.2/17.7
1wmi (RelEB)	2.3	(90 + 67) × 2	61.5/R	92.7	29.4/23.2	0.008/1.36	Yes	89.81 (36)	27.6/22.8
1wu7 (Ta0099)	2.4	434 × 2	83.4/R	87.0	29.6/25.9	0.006/1.33	No	97.2 (244)	26.3/21.0
1wy7 (PH1948)	2.6	207 × 4	85.5/A	94.8 (103)	28.9/24.0	0.007/1.37	No	96.5 (345)	26.8/24.2
1wv3 (SAV0287)	1.75	242	65.0/R	76.8 (189)	23.6/21.3	0.005/1.34	No	76.9 (295)	22.2/19.5
PF0475	2.9	276 × 4	99.3/M	100	27.6/24.8	0.008/1.43	No	Not finished	Not finished

† Number of residues in the asymmetric unit. The residues of the His tag and their linker are excluded because they could not be built in most cases. ‡ Percentages of residues in the initial model. M, the initial model was the search model of the MR method; A, the initial model was built automatically by *ARP/wARP*; R, the initial model was built automatically by *RESOLVE*; m, the initial model was built manually using *O*. § The values in parentheses indicate the numbers of water and other molecules. ¶ cns_b and cns_a are the root-mean-square deviations of bond lengths and bond angles calculated by *CNS*, respectively. †† Refinement was carried out semi-automatically using *LAFIRE* with manual intervention; wrongly built residues were removed (PDB codes 1wls, 1wmi) and/or one or two resides were built in the region where the electron density was very poor (PDB codes 1vgj, 1wmi). ‡‡ The initial model was built using 2.6 Å data.

differ in some side chains or loops, these protein molecules are built and fitted without NCS restraints, even when NCS restraints are applied to the refinement process.

2.4. Input and output

LAFIRE does not have a graphical user interface, but the parameters required for input have been simplified and minimized. The coordinates of the initial model, diffraction data with experimental phases (if available), sequence information and several parameters for process control are required. With the exception of parameters that are used only for controlling processes, all parameters used in the refinement process are defined in our program as the default state. *LAFIRE* produces refined coordinate files before and after pickup of water molecules and reports the model quality indicators, including the Ramachandran plot calculated using *PROCHECK* (Laskowski *et al.*, 1993), as well as the list of missing residues in the final model of *LAFIRE*.

2.5. Monitor program

To monitor the progress of refinement, a graphics program, *LAFIRE_moleview*, was developed using the Qt (TrollTech Inc.) OpenGL module (OpenGL Inc.). To maintain a good view of the structure, *LAFIRE_moleview* displays only a fragment of three, five or seven residues for both the structure and electron density in the main window. Concurrently, a sub-window shows the whole structure and the fragment shown in the main window is indicated in a different colour. This program also displays the GLCC, R_{free} factors and Ramachandran plots.

3. Results

LAFIRE has been used for fully automatic refinement from an initial to a final model including water molecules for test structural refinements that had already been performed manually. Furthermore, *LAFIRE* has been applied to more than 30 new structures with or without slight human intervention, of which 14 are shown in Table 1. The operations requiring human intervention were mostly to detect and remove wrongly built residues in the initial model and to locate ligands. In most cases, the initial models were fragmented (the maximum number of fragments was 12 for 1wls). The longest fragment built by *LAFIRE* during refinement in these applications had 24 residues. Applications to one test and three actual samples are described in detail below.

3.1. Automatic refinement for test samples

The program was developed using four test samples: PH1061 (PDB code 1ub9) and IPMI (Yasutake *et al.*, 2004; PDB code 1v7l) for the Se-MAD method and aIF5A (Yao *et al.*, 2003; PDB code 1iz6) and MC1_N71T (Numata *et al.*, 2003; PDB code 1ucg) for the MR method. These structures had been refined using the program *CNS*1.1 with manual intervention for linking and extending the fragments and fitting using the graphics program *O* between rounds of refinement. Here, we present the results of application to PH1061.

PH1061 is a good test sample as it consists of only a single molecule of 100 residues in an asymmetric unit. The crystals of protein PH1061 belong to space group $P4_32_12$, with unit-cell parameters $a = b = 45.2$, $c = 96.2$ Å. The phases of PH1061 were calculated using *SHARP* (de La Fortelle & Bricogne, 1997) with solvent flattening using *SOLOMON* (Abrahams &

Leslie, 1996). The initial model of PH1061 was built to 92% with three fragments (residues 1–26, 29–65 and 72–100) at 2.05 Å resolution using *ARP*/*wARP*. The missing fragment of residues 27–28 was part of an α-helix, while the fragment consisting of residues 66–71 was a turn between two strands. As shown in Fig. 3, the electron density was poor around residue 68. Auto-refinement was performed using *LAFIRE* with the program *CNS*1.1 in about 6 h (or using *REFMAC*5 in about 3 h) using Octane2 (R14000 CPU; 400 MHz blocks). All residues were built and the structure, which included 52 water molecules, was refined to an $R_{\rm free}$ and R factor of 24.5 and 21.5%, respectively, which were very close to the final values of 24.2 and 21.6% achieved by manual refinement (Table 1).

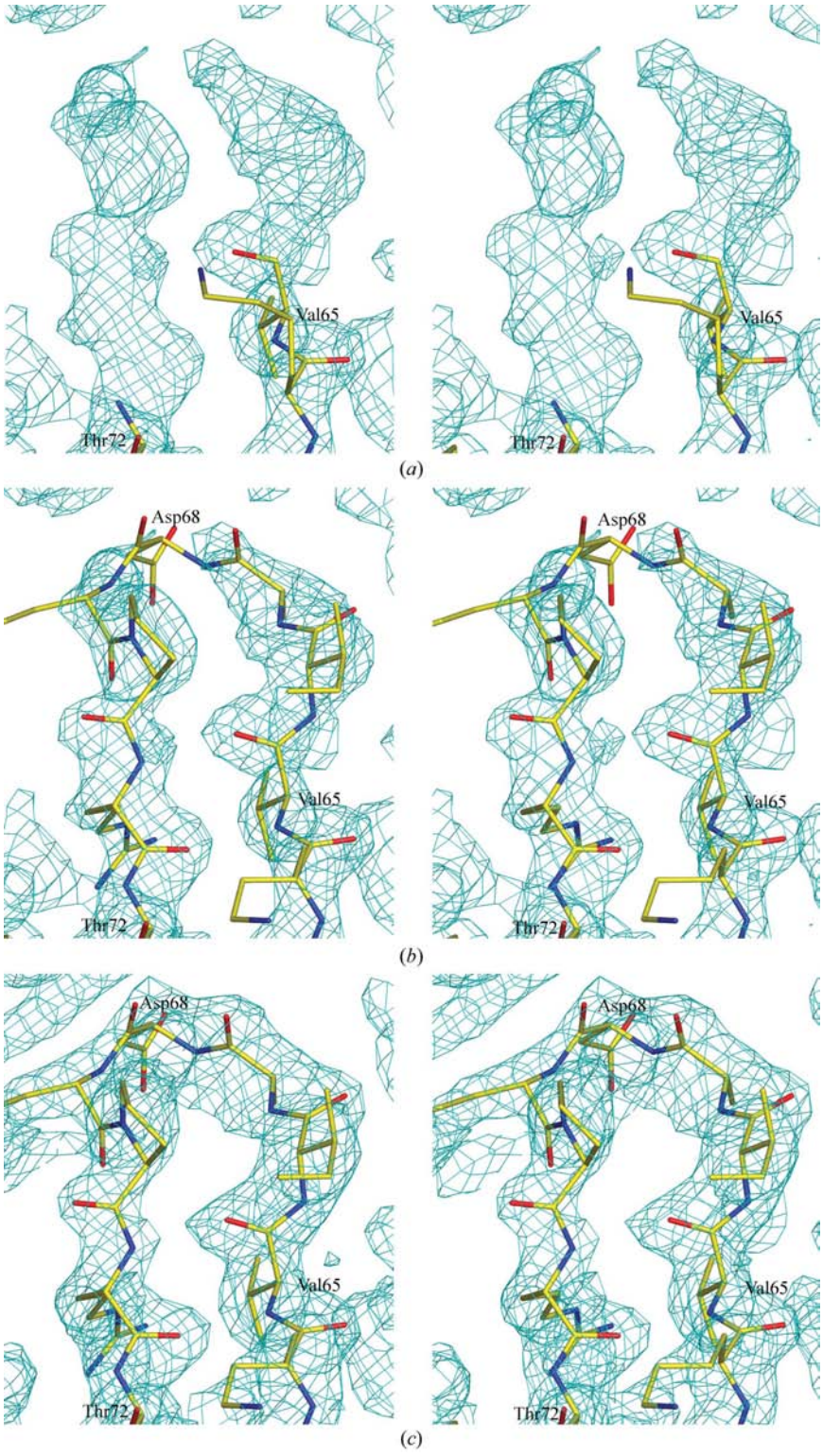

Figure 3
Stereo diagrams showing the initial and refined model of the loop region (residues 64–72) of protein PH1061. (*a*) The initial model of PH1601 obtained from *ARP*/*wARP* with the experimental density map contoured at 0.8σ. (*b*) The model refined automatically by *LAFIRE* with the *CNS*1.1 program. The density map is the same as that in (*a*). (*c*) The model is the same as that in (*b*) and the density map is $2F_{\rm o} - F_{\rm c}$ (contoured at 0.8σ) calculated after automatic refinement.

3.2. Applications

The protein ST2072 was crystallized in space group $P2_13$, with unit-cell parameters $a = b = c = 71.7$ Å and one molecule in an asymmetric unit (Tanaka *et al.*, 2005). The structure was solved by Se-MAD phasing and the initial model was built to 82 of 134 residues (61.2%) in six fragments using *SOLVE*/*RESOLVE* at 2.0 Å resolution (Figs. 4*a* and 4*c*). The missing fragments were linker loops and connected regions (β-strands and short α-helices). Refinement was carried out automatically from this initial model by *LAFIRE* with the *CNS*1.1 program. All residues were built and the structure was refined to an $R_{\rm free}$ and R factor of 24.6 and 20.2%, respectively, after the location of 63 water molecules without human intervention (Figs. 4*a*, 4*b*, 4*d* and Table 1). The Ramachandran plot showed that 88.1% of the non-glycine and non-proline residues were in the favoured region and 11.9% of residues were in the allowed region.

The protein PH0099 was assigned as a 2′-5′ RNA ligase with 184 amino-acid residues. The crystals of protein PH0099 belong to space group $P2_12_12_1$, with unit-cell parameters $a = 41.5$, $b = 45.7$, $c = 97.6$ Å. The crystal contained one molecule in an asymmetric unit. The structure was solved by the MR method at 1.85 Å resolution using 2′-5′ RNA ligase from *Thermus thermophilus* as a search model (32.6% identity; PDB code 1ihu). The initial model of PH0099 was rebuilt automatically to 78.8% with a poly-Ala/Ser/Gly model in five fragments using *ARP*/*wARP* (Figs. 5*a* and 5*b*). The missing fragments were linker loops and a C-terminal long β-sheet (residues 167–184).

Semi-automatic refinement was performed by *LAFIRE*. As the density map around residue 170 was very poor, the structure was built to only 91.8% (residues 1–169; Figs. 5*a* and 5*c*). All residues except the C-terminal residue 184 were finally built with the aid of manual intervention for residue 170 (Figs. 5*a* and 5*d*) and the structure was automatically refined to an $R_{\rm free}$ and R factor of 27.2 and 23.6%, respectively, after the location of 94 water molecules. The percentage of non-glycine and non-proline resides that fell in the favoured region of the Ramachandran plot was 91%; 8.4% of residues were in the allowed region, with only one residue in the generously allowed region.

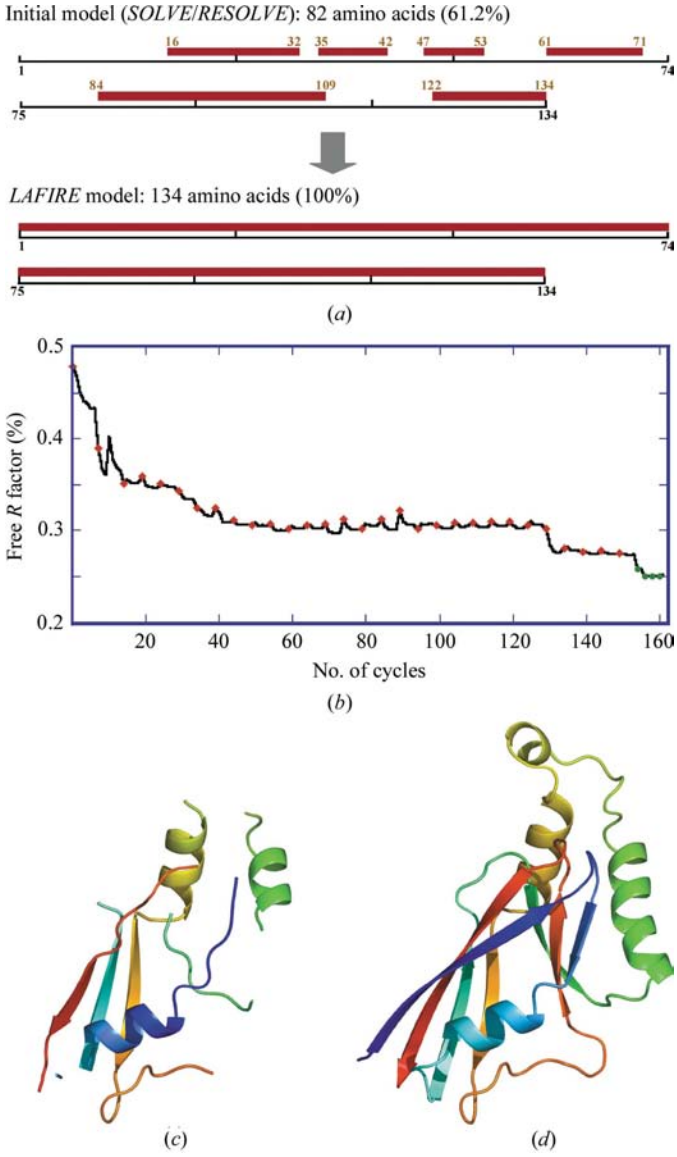

Figure 4
Refinement by *LAFIRE* with *CNS* for protein ST2072. (*a*) Status of building before and after *LAFIRE*. The red bar represents built residues. (*b*) $R_{\rm free}$ plot of the whole refinement process. The red triangles indicate steps of fitting and building, the black line shows refinement cycles of *CNS*1.1 and the green points show the steps for picking up water molecules. (*c*) Ribbon diagram of the initial model built with *RESOLVE*. (*d*) Ribbon diagram of the model after automatic refinement by *LAFIRE*.

PF0475 was crystallized in space group $P2_12_12_1$, with unit-cell parameters a = 78.1, b = 126.9, c = 140.7 Å. The crystal contained four molecules in an asymmetric unit. The structure was solved by the MR method at 2.9 Å resolution. As PF0475 has high sequence identity (60%) with the search model (PDB code 1vb5), the search model (residues 2–275) was used directly as the initial model for refinement. The residues were replaced automatically in reference to the sequence file and the structure was refined in two steps by *LAFIRE*. PF0475 was first refined without NCS restraints to an $R_{\rm free}$ and R factor of 34.5 and 27.3%, respectively. Two types of NCS restraints were then applied to four copies (*A*, *B*, *C* and *D*) in an asymmetric unit and *LAFIRE* was run again, once for only the main chain and once for all atoms. The NCS restraints for all atoms showed better results, with an $R_{\rm free}$ and R factor of 27.6 and 24.8%, respectively (29.5 and 25.0%, respectively, in the case of NCS restraints for the main chain) and the model of PF0475 was built to 100%. The Ramachandran plot showed that 89.9% of the non-glycine and non-proline residues fell in the favoured region and 10% of residues were in the allowed region, with one remaining residue in the generously allowed region.

4. Discussion

LAFIRE was designed essentially as a simulation program that automatically performs the refinement process usually performed by experienced crystallographers. In most cases, all refinement rounds can be performed automatically and the whole process takes only a few hours or days depending on the size of the protein, the resolution and the computer power. The resulting structure obtained by *LAFIRE* requires only a final inspection using a graphics program. Therefore, the program is very useful for high-throughput crystallography. Furthermore, by fully automating the whole refinement process, *LAFIRE* provides a suitable tool to optimize refinement conditions, such as NCS-restraint conditions and choice of data set from among a number of different diffraction data sets.

Compared with other building programs, the model-building routine in *LAFIRE* is designed to build less regular

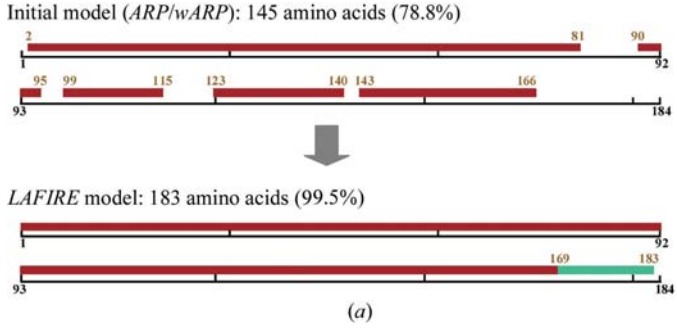

Figure 5
Refinement by *LAFIRE* with *CNS*1.1 for protein PH0099. (*a*) Status of building before and after processing with *LAFIRE*. The red bar represents built residues. The green bar represents the region built with manual involvement at residue 170.

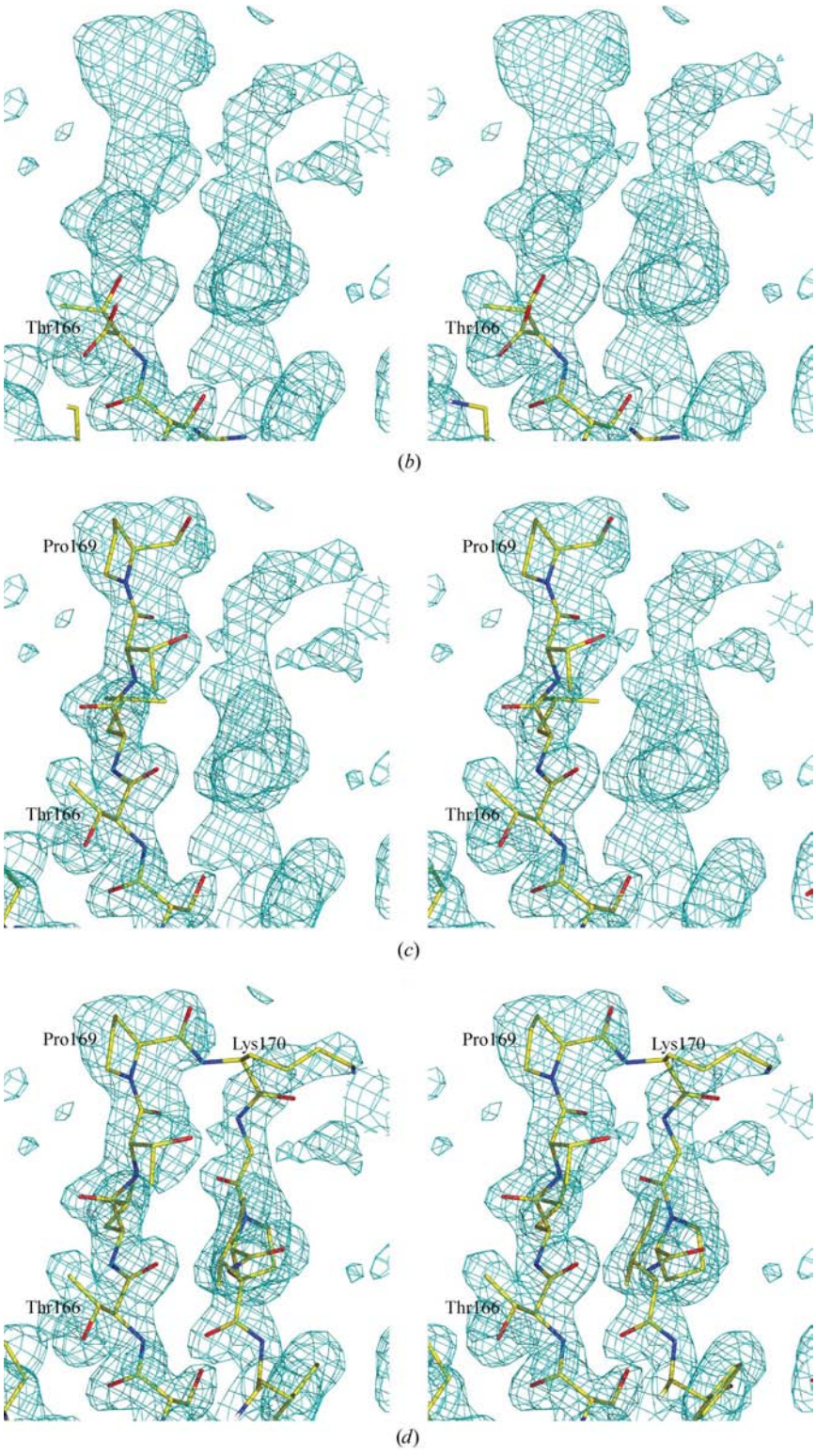

Figure 5 (continued)
(*b*) Stereo diagram of the electron-density region of the missing fragment consisting of residues 165–174. The density map is an σ-weighted $2F_o - F_c$ map contoured at 0.8σ. (*c*) Stereo diagram of the model consisting of residues 165–169, which were built by *LAFIRE* without manual intervention. The density map is an σ-weighted $2F_o - F_c$ map contoured at 0.8σ. (*d*) Stereo diagram of the refined fragment model of resides 165–174 with an σ-weighted $2F_o - F_c$ map contoured at 0.8σ. The fragment of residues 170–174 was built with *LAFIRE* after manual intervention around residue 170.

(and thus more difficult) parts such as loops and terminal regions and thus the program can be used during refinement. Very recently, a model completion program *Xpleo* has also been published which uses an inverse-kinematics algorithm with a real-space torsion-angle refinement (van den Bedem *et al.*, 2005). As shown in the refinement of protein ST2072, *LAFIRE* can accurately connect fragments and build to nearly 100% from a 61% main-chain model. In the most favoured case, it was possible to build a long missing fragment over 40 residues (data not shown). However, to perform this, *LAFIRE* requires correct sequence assignment of the fragments. Therefore, to make a fully automatic structure analysis program by combining *LAFIRE* with automatic phasing and building programs, automatic residue assignment of the fragments is required. Another difficult function that remains to be developed is the detection of an ill-built chain, including the insertion and deletion of residues in the initial model. In the current version of *LAFIRE*, only GLCC is used to detect an ill-built chain. In the next stage of development of our program, a method for detection by combining GLCC with fragment analysis including amino-acid assignment will be included.

Although it has been demonstrated that *LAFIRE* is able to work on data in a resolution range between 1.65 and 3.0 Å, refinement with low-resolution data is still difficult. The map calculated from low-resolution data often leads to ambiguous chain-tracing. To cope with this problem, users can apply secondary-structure restraints during main-chain tracing in *LAFIRE* to obtain a more reliable model. Future versions of *LAFIRE* will include automatic detection of the secondary structure.

Three (RelEB, PH0066, and PH0099) of the 14 new crystals listed in Table 1 could not be refined fully automatically. The initial model of RelEB contained serious insertions and deletions, while that of PH0066 was first built at 2.6 Å resolution with 12 fragments where the longest missing fragment had 30 residues. *LAFIRE* was interrupted several times for manual intervention to remove wrongly built residues for RelEB and PH0066. Such manual intervention will be reduced in the subsequent version of the program by including the fragment analysis described above. Using

the current version of *LAFIRE*, it is difficult to build the missing parts at the C- or N-terminal fragments where the electron density is very poor and the map is segmented, such as that of the fragment consisting of residues 170–184 in PH0099 (Fig. 5). In such cases, human intervention is still necessary to build two or three residues to connect separated map segments. A building algorithm to overcome these problems is currently under development.

As described in §2, building and fitting are carried out in the current version of *LAFIRE* based on the experimental electron-density map or σ_A-weighted $2F_o - F_c$ maps and the $F_o - F_c$ map is used only to check side-chain fitting. However, the building ability of *LAFIRE* is relatively low in the MR method compared with the MIR, MAD or SAD methods because there is no experimental electron-density map in the MR method. Use of $F_o - F_c$ maps will be included in the building strategy in the next version of *LAFIRE*.

With regard to the job-control system, off-line parallel processing using multiple computers at the laboratory level will be included in future versions of *LAFIRE* to automate refinement of large proteins. In parallel processing, building and fitting will be performed as several independent jobs that can be executed on different computers simultaneously in the laboratory.

5. Caution

Although *LAFIRE* is designed to perform the whole refinement process automatically as described, users should always check refined coordinates based on both $2F_o - F_c$ and $F_o - F_c$ maps with care through computer graphics. These checks include whether refined parameters are acceptable with respect to stereochemistry, whether water molecules are reliably located and whether any more ligands exist in the electron density (unknown ligands are not built in the current version of *LAFIRE*).

6. Availability

LAFIRE is available from http://altair.sci.hokudai.ac.jp/g6/Research/Lafire_English.html. For help, contact lafire@castor.sci.hokudai.ac.jp.

We would like to thank all those who provided their data for program development. We also thank Mr T. Matsumoto for testing and debugging our program. This work was supported by a research grant from the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1956.
Bedem, H. van den, Lotan, I., Latombe, J.-C. & Deacon, A. M. (2005). *Acta Cryst.* D**61**, 2–13.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
Brunzelle, J. S., Shafaee, P., Yang, X., Weigand, S., Ren, Z. & Anderson, W. F. (2003). *Acta Cryst.* D**59**, 1138–1144.
Fu, Z.-Q., Rose, J. & Wang, B.-C. (2005). *Acta Cryst.* D**61**, 951–959.
Hendrickson, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
Holton, J. & Alber, T. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 1537–1542.
Ioerger, T. R. & Sacchettini, J. C. (2002). *Acta Cryst.* D**58**, 2043–2054.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Levitt, D. G. (2001). *Acta Cryst.* D**57**, 1013–1019.
Luo, Y., Jiang, X. L., Lai, L. H., Qu, C. X., Xu, X. J. & Tang, Y. Q. (1992). *Protein Eng.* **5**, 147–150.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Morris, R. J., Perrakis, A. & Lamzin, V. S. (2002). *Acta Cryst.* D**58**, 968–975.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Ness, S. R., de Graaff, R. A. G., Abrahams, J. P. & Pannu, N. S. (2004). *Structure*, **12**, 1753–1761.
Numata, T., Suzuki, A., Kakuta, Y., Kimura, K., Yao, M., Tanaka, I., Yoshida, Y., Ueda, T. & Kimura, M. (2003). *Biochemistry*, **42**, 5270–5278.
Pavelcik, F., Zelinka, J. & Otwinowski, Z. (2002). *Acta Cryst.* D**58**, 275–283.
Perrakis, A., Morris, R. J. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. J. (1963). *J. Mol. Biol.* **7**, 95–99.
Tanaka, Y., Tsumoto, K., Tanabe, E., Yasutake, Y., Sakai, N., Yao, M., Tanaka, I. & Kumagai, I. (2005). *Proteins*, **61**, 1127–1131.
Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1937–1940.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.
Turk, K. & Guncar, G. (1999). Am. Crystallogr. Assoc. Abstracts.
Yao, M., Ohasawa, A., Kikukawa, S., Tanaka, I. & Kimura, M. (2003). *J. Biochem.* **133**, 75–81.
Yasutake, Y., Yao, M., Sakai, N., Kirita, T. & Tanaka, I. (2004). *J. Mol. Biol.* **344**, 325–333.
Zhou, Y., Yao, M. & Tanaka, I. (2006). *J. Appl. Cryst.* **39**, 57–63.