Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A novel strategy for the crystallization of proteins: X-ray diffraction validation

Steven B. Larson,[a] John S. Day,[a] Robert Cudney[b] and Alexander McPherson[a]*

[a]Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697, USA, and [b]Hampton Research, 34 Journey Road, Aliso Viejo, California 92656, USA

Correspondence e-mail: amcphers@uci.edu

Received 1 October 2006
Accepted 8 December 2006

Recently, the hypothesis was advanced that protein crystallization could be driven by the inclusion of small molecules rich in hydrogen-bonding, hydrophobic and electrostatic bonding possibilities. Conventional organic and biologically active molecules would promote lattice formation by their mediation of intermolecular interactions in crystals. The results of an extensive series of crystallization experiments strongly supported the idea. Here, difference Fourier X-ray diffraction analyses of nine crystals grown in the experiments are presented, which convincingly demonstrate the validity of the hypothesis and illustrate some of the ways in which small molecules can participate in lattice interactions.

1. Introduction

A consensus has emerged from the structural genomics initiative (Chandonia & Brenner, 2006; Norvell & Machalek, 2000) that accelerated progress in structural biology at the molecular level will be increasingly dependent on the development of novel and innovative approaches to crystallizing biological macromolecules for X-ray diffraction analyses. To take full advantage of this powerful technology, imaginative new ideas are therefore essential. Current practices no longer appear to be sufficient. Simply manipulating further the variables and reagents that are now common currency is not likely to produce the major breakthroughs that are clearly needed.

We recently proposed an alternative strategy for crystallizing proteins and viruses that is based on the inclusion of a wide range of conventional and bioactive small molecules in the mother liquors (McPherson & Cudney, 2006). The underlying hypothesis of this approach was that the small molecules would form reversible cross-links in the crystal lattice through intermolecular electrostatic, hydrogen-bonding and possibly hydrophobic interactions. Such interactions, by promoting and stabilizing lattice contacts and by filling voids between protein molecules in the crystals, would in turn promote crystal nucleation and growth.

Such auxiliary small molecules have traditionally been referred to as additives, suggesting a secondary role in the crystallization process. Our alternative strategy made low-molecular-weight compounds a primary factor in promoting crystallization, as opposed to the usual variables of precipitant, concentration and pH. Thus, the approach is essentially orthogonal to those of common practice and offers a new means of overcoming a persistent but increasingly acute problem.

In the original paper, we presented statistical data based on the crystallization of 81 different proteins and viruses in the presence of about 200 small molecules, almost all of which

Table 1
Summary of crystal structures: starting models, crystal data, data statistics and refinement statistics.

	Thaumatin	RNase A	Lysozyme	Lysozyme	Lysozyme	Bovine trypsin	Porcine trypsin	Porcine trypsin	Porcine trypsin
Experiment	I-48	III-18	I-4	II-9	I-22	II-33	I-34	I-3	I-18
PDB model	1thw	1rno	1hel	1hel	1hel	1qcp	1ldt	1s81	1v6d
Space group of model	$P4_12_12$	$P3_221$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P3_121$	$P4_32_12$	$P2_12_12_1$	$P2_12_12_1$
Model unit-cell parameters									
a (Å)	58.60	65.10	79.10	79.10	79.10	54.43	63.40	76.42	46.93
b (Å)	58.60	65.10	79.10	79.10	79.10	54.43	63.40	53.47	53.72
c (Å)	151.8	65.52	37.90	37.90	37.90	107.76	131.20	46.61	77.47
Space group, this study	$P4_12_12$	$P3_1$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P3_121$	$P4_12_12$	$P4_12_12$	$P2_12_12_1$
Unit-cell parameters									
a (Å)	58.54	55.08	78.93	78.90	79.11	138.39	58.82	59.00	47.04
b (Å)	58.54	55.08	78.93	78.90	79.11	138.39	58.82	59.00	54.01
c (Å)	152.00	39.20	37.89	37.93	37.91	150.54	135.34	139.66	77.43
Data statistics									
Resolution (Å)	27.31–1.47	27.54–1.33	27.38–1.31	34.51–1.54	34.19–1.31	38.81–1.80	27.07–1.41	25.93–1.28	26.15–1.24
High-resolution shell	1.52–1.47	1.38–1.33	1.36–1.31	1.59–1.54	1.36–1.31	1.86–1.80	1.46–1.41	1.33–1.28	1.29–1.24
No. of reflections	40563	27155	26140	15575	26946	153457	36756	58890	44379
Redundancy	4.90 (1.26)	3.33 (1.49)	6.18 (2.01)	10.69 (3.01)	7.87 (2.34)	15.25 (14.94)	3.41 (1.32)	5.78 (1.62)	3.19 (1.45)
$\langle I/\sigma(I)\rangle$	17.7 (1.9)	13.7 (0.8)	12.1 (1.8)	13.3 (1.9)	17.2 (1.8)	3.9 (2.1)	13.5 (2.0)	16.5 (1.8)	15.9 (1.6)
Completeness (%)	88.0 (24.6)	89.3 (24.9)	89.0 (29.8)	83.1 (23.5)	91.9 (32.7)	99.8 (100.0)	78.7 (14.9)	91.5 (32.0)	78.6 (19.9)
$R_{\rm merge}$†	0.051 (0.334)	0.049 (0.491)	0.065 (0.476)	0.082 (0.459)	0.053 (0.471)	0.168 (0.641)	0.053 (0.270)	0.054 (0.375)	0.043 (0.410)
Structure solution and refinement									
Solution method‡	Direct	MR	Direct	Direct	Direct	MR	MR	MR	Direct
Resolution (Å)	1.47	1.33	1.50	1.54	1.35	1.80	1.52	1.28	1.29
Data cutoff [$\sigma(F)$]	0.0	0.5	0.0	0.0	4.0	4.0	1.0	0.0	0.0
R§	0.189	0.205	0.198	0.199	0.203	0.225	0.195	0.188	0.190
$R_{\rm free}$§	0.206	0.224	0.247	0.240	0.229	0.283	0.227	0.208	0.200
Test set (%)	5.6	8.7	6.7	8.6	8.6	5.4	6.4	9.2	5.5
Ligands found	Tartrate, histidine	5′-dGMP	PAB¶	Trimesic acid	None	Mellitic acid, benzamidine	Benzamidine, D-sorbitol, glycerol	ABS¶, benzamidine, oxamate	Benzamidine, malonate

† $R_{\rm merge} = \sum_h \sum_i |I_{hi} - \langle I_h\rangle| / \sum_h \sum_i \langle I_h\rangle$, where I_{hi} is the ith used observation for unique hkl h and $\langle I_h\rangle$ is the mean intensity for unique hkl h. ‡ Direct implies that the space group and unit cell were the same as the PDB model and that refinement proceeded directly from model coordinates; MR implies a new space group and unit cell and that the structure had to be solved using molecular-replacement techniques employed in *CNS*. § $R = \sum_h |F_o - F_c| / \sum_h F_o$, where F_o and F_c are the observed and calculated structure-factor amplitudes, respectively. ¶ PAB, *p*-aminobenzoic acid; ABS, sulfanilic acid (*p*-aminosulfonic acid).

were potential protein ligands. These results strongly supported our underlying hypothesis regarding the positive influence and potential value of the compounds. Two features of those experiments were conspicuous. Firstly, only two fundamental crystallization conditions were employed, one based on PEG and the other on a salt mixture, both at a single pH, and secondly, multicomponent 'cocktails' of five or more small molecules were added to the crystallization trials rather than individual compounds. The latter took advantage of the fact that crystallization is a selection process. The developing crystal selects from the solution what it needs to grow and rejects what it does not. Both served to make screening of a broad range of macromolecules and small molecules practical.

At the same time, we accepted the risk that one component of a cocktail might have a deleterious effect on crystal growth and thereby mask the positive contribution of another. We in fact saw evidence of this for at least one cocktail in the crystallization experiments, where the beneficial effects of cobalt hexamine (as shown by a subsequent experiment) were obscured by other components which caused rapid precipitation of nearly all of the test proteins. This problem is largely overcome experimentally, however, by including every component in two or more cocktails.

The crystallization data provided persuasive evidence that for many macromolecules incorporation of one or more small molecules could be crucial to obtaining crystals of specific proteins. The results further indicated that certain classes of small molecules, such as dicarboxylic acids and diamino compounds of various sizes and geometries, promoted the crystallization of proteins in a general sense. In addition, different crystal polymorphs were produced of some proteins in the presence of various small molecules and effects on the diffraction resolutions of some crystals were also observed.

To determine whether the underlying hypothesis that the small ligands did indeed serve to tether macromolecules to one another and thereby encourage lattice formation was valid, X-ray analysis of at least a sample of the crystals grown in the experiments was necessary. Only by this technique could detailed interactions within lattices be directly visualized and the original idea rigorously evaluated. In the following, we briefly describe the analyses of nine crystalline proteins obtained in the original experiments (McPherson & Cudney, 2006), each of which was obtained in the presence of a 'cocktail' of low-molecular-weight compounds.

2. Experimental procedures

X-ray diffraction data were collected to the highest resolution allowed by the crystals or by the experimental apparatus on an R-AXIS detector system with 1.54 Å radiation generated by a Rigaku RU-200 rotating-anode source fitted with Osmic mirrors (Rigaku USA, Woodlands, TX, USA). Rotation

angles varied from 0.5 to 1.0°, with recording times of 5–10 min per frame. Intensities were integrated, merged and converted to structure amplitudes using the program *d*TREK* (Pflugrath, 1999). All data were collected at 295 K from crystals conventionally mounted in quartz capillaries. We chose to collect data at 295 K in order to obtain a fair appraisal of the diffraction limits of the crystals in the absence of any damage that might result from cooling to 100 K. Although we did not thoroughly investigate the matter, we did carry out some cursory cooling experiments and these gave no indications that the presence of the small-molecule mixtures had any effect in terms of cryocrystallography. Except for the bovine trypsin data set, which was obtained from three crystals, all data sets were obtained from a single crystal.

Models for the unliganded protein molecules, identified in Table 1, were obtained from the Protein Data Bank (PDB; Berman *et al.*, 2000) and stripped of all nonprotein atoms that might obscure the presence of any small molecule incorporated in the lattice. For crystals having space groups and unit-cell parameters identical to models in the PDB, the protein coordinates obtained from the PDB were refined against the structure amplitudes recorded from the liganded protein crystals (McPherson & Cudney, 2006). For crystals with previously unreported space groups and unit-cell parameters, the structures were first solved by molecular replacement with *CNS* (Brünger *et al.*, 1998) using the PDB protein models and then refined. Difference Fourier maps were calculated from the refined models at the maximum resolutions of the refinement (Table 1) and the ligands or structural changes (lysozyme experiment I-22) were identified. The protein models were rebuilt and the ligands were added to the model. Alternate cycles of refinement and water addition were performed to obtain the final models, the statistics of which are shown in Table 1.

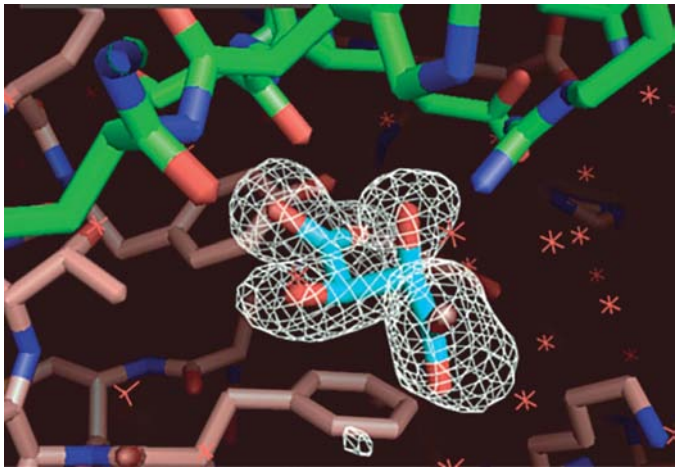

Figure 1
A molecule of tartrate is bound at the junction of three protein molecules in a tetragonal crystal of thaumatin. Two of the three protein molecules are shown in green and pink and the tartrate is shown in blue. The tartrate serves as the center of a complex network of intermolecular hydrogen bonds. Density from a $2F_o - F_c$ tartrate-omitted map, calculated at 1.47 Å and contoured at 1.4σ, is superimposed on the model. Red asterisks in this figure (and the other figures) represent water molecules.

Refinement and other computing operations were carried out using the program *CNS* (Brünger *et al.*, 1998). The refined models and $F_o - F_c$ difference densities were superimposed and inspected using the program *PyMOL* (DeLano, 2002), which was also used for the preparation of figures. Models were rebuilt and ligands were placed using the program *O* (Jones & Kjeldgaard, 1994).

3. Results

We present here nine examples where bound small molecules from the various reagent mixes were clearly seen in difference Fourier maps and, in general, where we could identify which component of the relevant reagent mix the ligand represented. These cases are most frequently those where the small molecule contained an aromatic ring, such as sulfanilic acid, or a larger ligand such as a nucleotide. We could not be certain that some small ligands, such as formate, would be readily identifiable in difference maps, particularly at resolutions poorer than 1.8 Å. A summary of the crystals here is presented in Table 1. X-ray diffraction analyses, in which data were recorded from crystals grown in the presence of various reagent mixes and then used to calculate difference Fourier syntheses, generally revealed the presence of ligands in the lattices at interfaces between protein molecules, consistent with the motivating hypothesis.

3.1. Thaumatin

Thaumatin crystals were grown from a mixture of PEG and Tacsimate in the presence of a relatively crude preparation of protamine in experiment I-48 (numbering according to the previous paper; McPherson & Cudney, 2006). Crystals belonging to the tetragonal space group $P4_12_12$ (see Table 1) can generally be obtained from about 1 *M* sodium/potassium tartrate. This is an important point, because previous analyses have shown tartrate to be an essential factor in growing the tetragonal form. Tacsimate, however, contains 0.16 *M* tartrate as a component and the crystallization sample, after equilibration by vapor diffusion, would have had a tartrate concentration of about 0.08 *M*.

As illustrated in Fig. 1, even though the tartrate concentration was relatively low, a single tartrate molecule was incorporated into the crystal lattice just as reported previously (Ko *et al.*, 1994), at the juncture of three crystallographically related thaumatin molecules. The three thaumatin molecules are drawn together and maintained by the single tartrate molecule, which lies at the center of a network of hydrogen bonds, all of which involve the carboxylate groups of the small molecule.

An important point is that the tartrate molecule was selected by the growing crystal from the array of small organic acids present in Tacsimate. Some (for example, succinate) are not only structurally similar to tartrate, but were present at significantly higher concentrations. The appearance of tartrate alone in the lattice illustrates a fundamental principle of the experiments: that crystal growth is indeed a selection process

whereby the growing lattice incorporates those molecules and ions which aid its development and rejects those that serve no useful purpose. This confirms the idea that in practice 'cocktails' of several or even many potential ligands may be used in screening experiments rather than individual compounds.

The thaumatin crystals from experiment I-48 show one other significant feature in difference Fourier maps. This arises from the protamine included in the mother liquor. While no intact protamine polypeptide is evident, at one juncture of the protein molecules a difference electron-density mass is observed. We cannot unambiguously identify the source of this mass, but have modeled it as an amino-acid residue (a histidine). This is not unreasonable given the complex nature of the protamine preparation.

3.2. Bovine ribonuclease A

Ribonuclease A (RNase A) crystallized in experiment III-18 from PEG in the presence of a reagent mix comprised of dGMP, cholesterol, thymine and oxamic acid. It appeared in a previously unreported unit cell with space-group symmetry $P3_1$ (Table 1). Thus, the structure had to be solved by molecular replacement and refined before difference Fourier maps could be calculated. A more detailed crystallographic analysis of this somewhat complicated structure will be presented elsewhere, but some of its key features are pertinent. Two dGMP molecules (labeled GMP200 and GMP300) are clearly bound in the active-site cleft of the enzyme. The electron density superimposed on the two molecules of dGMP is shown in Fig. 2. GMP300 is in the canonical 'purine-binding site' as has been seen in previous nucleotide and oligonucleotide complexes (McPherson *et al.*, 1986; Wlodawer, 1985). The base, deoxyribose and phosphate all appear to be properly positioned and to be oriented as one would predict.

The guanine ring of GMP200, however, occupies what would normally be considered to be the 'pyrimidine binding site'. The orientation, however, is completely unconventional and represents some abortive mode of binding. The position of the base and sugar are reversed from the positions that would be occupied by a pyimidine ribonucleotide and the guanine moiety makes two good hydrogen bonds to the phosphate group of the GMP300 at the catalytic center through its amino group and a ring nitrogen. The phosphate of the abortively bound dGMP extends out of the 5′ end of the binding cleft and makes contact with another RNase A molecule in the lattice related by the 3_1 axis. Thus, the order of bonded components in the active-site cleft 3′ to 5′ is base–sugar–phosphate–base–sugar–phosphate.

The intermolecular interactions mediated by the two dGMP molecules, as shown in Fig. 2(*b*), are intriguing. At the 5′ end of the catalytic cleft, the phosphate of the abortively bound GMP200 is associated directly with amino-acid residues Asn24 and Gln28 of a 3_1 screw axis-related RNase A molecule, thus linking them firmly together. At the 3′ end of the cleft, the guanine ring in the 'purine-binding site' simultaneously makes an unambiguous stacking interaction with Tyr115 of another RNase A molecule related by a second 3_1 screw axis. Thus, the RNase A molecules are drawn into a helical configuration

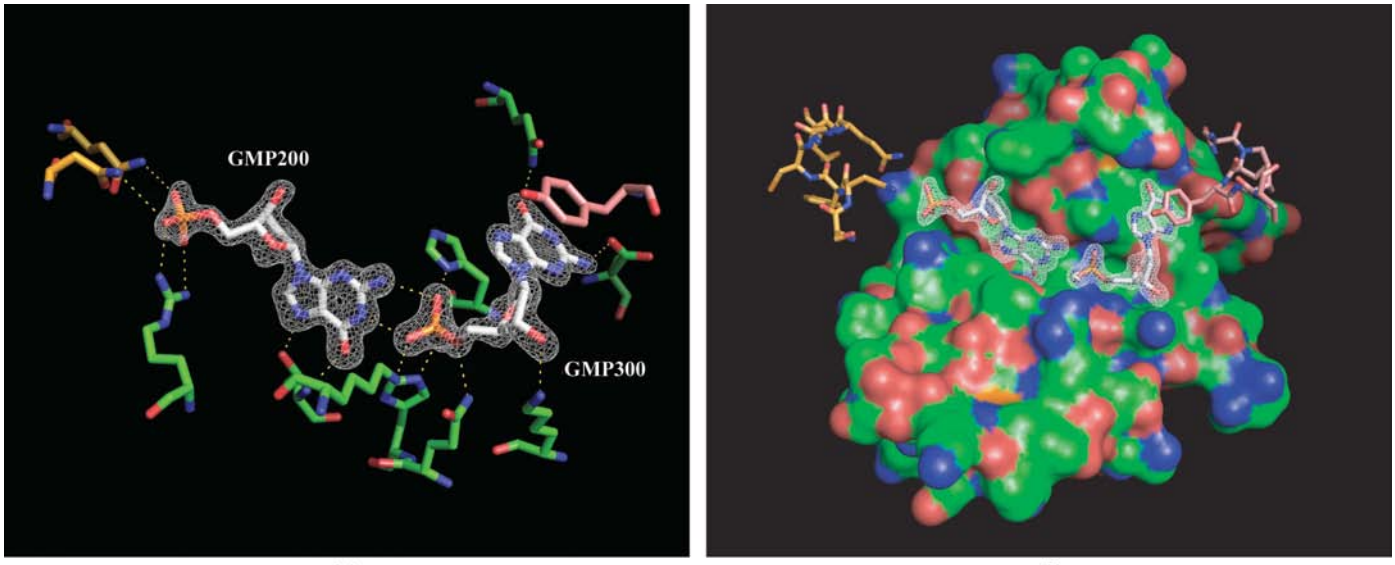

Figure 2
In (*a*) the electron density calculated at 1.33 Å resolution from the RNase A crystal grown in experiment III-18 is shown superimposed on two molecules of dGMP. The mononucleotides were one of four components present in the reagent mix used in the experiment, but were the only component to be incorporated into the crystal. The residues to which dGMP molecules make direct hydrogen bonds, indicated by broken yellow lines and ranging in length from 2.59 to 3.52 Å, are also shown in three different colors (orange, green and pink) representing the three molecules with which dGMP interacts. In (*b*) the two dGMP molecules, bound principally in the active-site groove of the RNase A represented by the van der Waals surface, also make significant intermolecular interactions with two other crystallographically equivalent RNase A molecules related by 3_1 screw axes and shown in pink and orange. The crystal is held together to a large extent by the nucleotide linkages across the protein–protein interfaces. The electron density displayed here is from $2F_o - F_c$ maps in which the dGMP molecules were omitted from the model and is contoured around the dGMP molecules at 2.0σ for GMP300 and 3.0σ for GMP200.

around the 3_1 screw axis by a continuous chain of interactions arising from the bound nucleotides. For the most part, structurally determinant intermolecular interactions are principally provided by the bound dGMP molecules.

It cannot be argued that the arrangement of protein molecules arises from the physiologically consistent binding of natural ligands alone or pseudo-substrates, since RNase A normally acts only on ribonucleotides. While GMP300 bound in the purine-binding site reflects the accepted binding of a nucleotide phosphate, the GMP200 most certainly does not. The fact that it is a nucleotide is essentially irrelevant. The important features are its ability to bind to the protein by improper insertion of the guanine ring into the pyrimidine pocket, the ability of the guanosine base to hydrogen bond to the 5′ phosphate of the GMP300 and the ability of its 5′ phosphate to make good hydrogen-bonding interactions with a 3_1 symmetry-related RNase A molecule.

3.3. Hen egg lysozyme

Three experiments involving lysozyme illustrate how small molecules can assist in the establishment of crystal lattices. The tetragonal form of hen egg lysozyme is readily obtained from NaCl solutions of 0.6–1.6 *M*, but not from PEG solutions. In experiment I-4, tetragonal lysozyme crystals were grown from PEG in the presence of a reagent mix containing six components. Of these, one in particular, *p*-aminobenzoic acid, was readily identified in difference Fourier maps. The *p*-aminobenzoic acid, as shown in Fig. 3, is bound at the leading edge of the known binding cleft of one lysozyme molecule, but also makes extensive contacts with a second protein molecule in the lattice. The amino-acid residues involved are Gly22, Ser24 and Thr118 of one molecule and Arg114 of another. Some hydrogen bonds are through water molecules to the proteins, but the intermolecular nature of the *p*-aminobenzoic acid binding is evident.

The tetragonal crystal form of lysozyme was also grown in experiment II-9 from PEG, but in the presence of a 'cocktail' containing trimesic acid, phloroglucinol, mellitic acid, *myo*-inositol and phytic acid. Difference Fourier maps showed two prominent features. The first was a chain of density partly filling and emerging from the active-site cleft of one molecule and terminating at the surface of a second. This was very near the position of the *p*-aminobenzoic acid in the experiment above. It appeared to be involved in intermolecular interactions. The chain of density, however, did not have a shape that correlated well with any one of the components of the reagent mix. We concluded that multiple components of the 'cocktail' are likely to bind within the active-site cleft, but that they are largely disordered and hence unrecognizable in electron-density maps.

The other major feature in the same difference Fourier map is a single ring-shaped mass of density with three protrusions separated by about 120° around the ring; this fitted well with trimesic acid. The density, with a trimesic acid molecule superimposed, along with associated protein model is shown in Fig. 4. The trimesic acid molecule is tightly sandwiched at the interface of three lysozyme molecules and is in contact with residues Asn27 and Ser24 of one protein molecule, Arg114 and Lys33 of a second molecule and Arg73 of a third. It clearly provides a bridging function in the lattice and bears a major responsibility for its maintenance.

In experiment I-22, a crystal was again obtained from PEG in the presence of a reagent mix containing *trans*-aconitic acid, glyceric acid, indolbutyric acid and hydroxyphenylacetic acid. The most distinctive feature appearing in the difference Fourier map was an unambiguous alteration in the conformation of a polypeptide loop lying immediately below the active site. Specifically, this involved a significant change in the dispositions of Ser100 and Asp101, with small movements of the amino acids on each side. The change in the loop conformation, shown in Fig. 5, was not observed in any of the other (nearly 15) difference Fourier analyses that we carried out on tetragonal lysozyme crystals. Indeed, the conformational change appeared to be unique to this particular reagent mix.

Further examination of the difference Fourier map did not indicate the presence of either indolbutyric acid or hydroxyphenylacetic acid in the crystal, both of which contain prominent ring structures, nor was *trans*-aconitic acid obviously present. In the active site and probably involved with the loop with altered conformation, however, was density that could be ascribed to glyceric acid. Alhough the mechanism is not at all clear, binding of glyceric acid would seem to be responsible for the observed conformational change. The importance of this last observation is that in at least some cases the small-molecule ligands can induce conformational changes in a protein. Thus, they may act to alter intermolecular contacts even when they themselves are

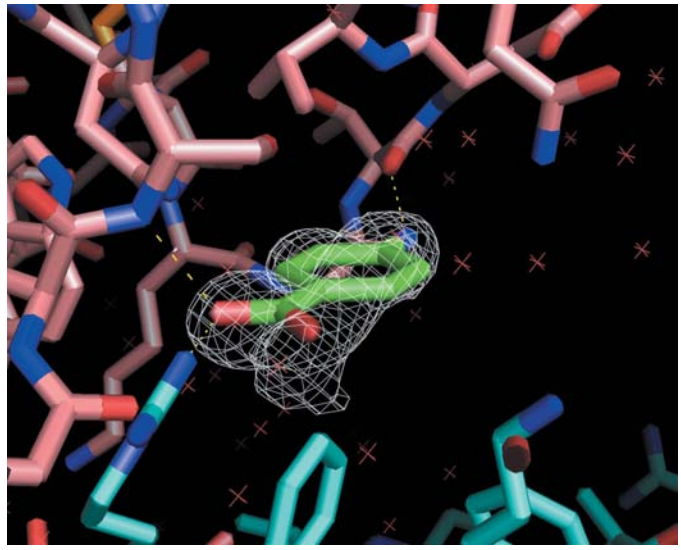

Figure 3
A molecule of *p*-aminobenzoic acid (green) lies at the interface between two protein molecules (pink and cyan) in a lysozyme crystal. The electron density from a $2F_o - F_c$ OMIT map calculated at 1.50 Å resolution and contoured at 0.7σ is superimposed on the model. Direct hydrogen bonds between the ligand and the protein molecules, ranging from 2.59 to 3.45 Å in length, are shown as broken yellow lines.

not directly involved. This suggests yet another mechanism for their influence on crystal growth.

3.4. Bovine trypsin

The most unusual and striking crystal structure that we analyzed (really a superstructure) was of bovine trypsin crystallized in the presence of a mixture of trimesic acid, mellitic acid, pyromellitic acid, terepthalic acid and benzamidine. The unit cell was previously unreported (space group $P3_121$, unit-cell parameters $a = b = 138.39$, $c = 150.94$ Å) and the structure therefore had to be solved by molecular replacement and refined prior to calculation of difference Fourier maps. Cycles of manual rebuilding and refinement were required as the model was developed. A complete description of the structure determination and analysis will be presented elsewhere, but features of the structure particularly relevant to the questions at hand are appropriate here.

The asymmetric unit consists of eight independent molecules of bovine trypsin, which form a multihelical arrangement about the crystallographic threefold screw axis of the unit cell. Every trypsin molecule has a distinctive tightly bound benzamidine molecule in the S1 pocket of the active site (Ascenzi *et al.*, 1984). The disposition of the benzamidine and its interaction with the active-site residues is exactly as observed previously in other crystal forms of the inhibited enzyme. The benzamidines are entirely associated with individual protein molecules and make no intermolecular contacts. The same is seen for the benzamidine molecules that occupy the active sites of the porcine trypsin structures described below.

The trypsin superstructure is maintained almost entirely by mellitic acid molecules, as shown in Figs. 6(*a*) and 6(*b*), which are disposed at the contact points between trypsin molecules. These were readily identified by their distinctive electron-density masses, which appeared eight independent times in the asymmetric unit. As seen in Fig. 6(*b*), each mellitic acid is heavily engaged in hydrogen bonds with two trypsin molecules. Two carboxylate groups on one side of the ring interact with one trypsin molecule, while two carboxylate groups on the other side of the ring interact with a neighbor. Two opposing carboxylate groups of each mellitic acid are not utilized in the intermolecular interactions and form hydrogen bonds with water.

The mellitic acid molecules and their arrangement in this crystal provide an exquisite illustration of the underlying hypothesis that intermolecular cross-links introduced by small molecules can drive lattice formation. The mellitic acid is homologous to the tartrate in tetragonal thaumatin crystals. Once again, the unique crystal lattice has selected from an array of structurally and chemically similar small molecules exactly as needed to self-assemble.

3.5. Porcine trypsin

Three different crystal forms of porcine trypsin grown in the experiments were analyzed in detail using X-ray diffraction. These were from experiments I-34, I-3 and I-18. The first reagent mix contained Tacsimate, benzamidine, glycerol, sorbitol and sucrose, while the second reagent mix contained PEG 3350, benzamidine, sulfanilic acid, oxamic acid and *p*-aminobenzoic acid. The third was grown from Tacsimate and

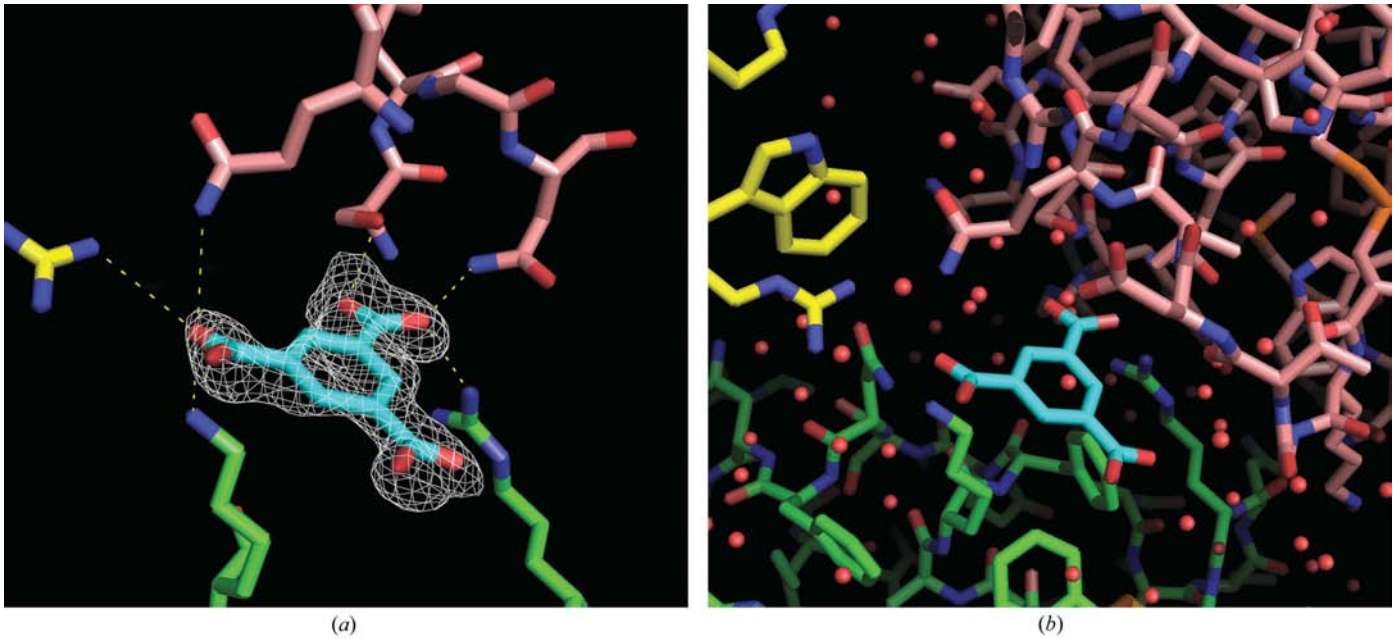

Figure 4
In (*a*) a molecule of trimesic acid (cyan) lying at the interface of three protein molecules (green, yellow and pink) in a lysozyme crystal grown in experiment II-9 is shown with electron density from a $2F_o - F_c$ map superimposed. The map was calculated at 1.54 Å resolution and is contoured at 0.7σ. All three of the carboxylate groups are clearly apparent. Six potential hydrogen bonds between the trimesic acid molecule and the lysozyme molecules, with lengths in the range 2.72–3.65 Å, are indicated by broken yellow lines. In (*b*) the molecule of trimesic acid is seen at the interface of three lysozyme molecules (green, pink and yellow), where it engages in an extensive hydrogen-bonding network that includes water molecules (shown as red spheres).

PEG 3350 in the absence of any reagent mix, but in the presence of benzamidine.

The crystals grown from Tacsimate in experiment I-18 had a unit cell that was virtually identical to a previous PDB entry (1v6d), which also belonged to the orthorhombic space group $P2_12_12_1$. As with the bovine trypsin, a benzamidine was clearly present at the active site and again made contacts with only a single trypsin molecule. A second difference electron-density mass was also seen somewhat away from the active site, at the surface of the protein, and lying at the interface between two crystallographically related trypsin molecules. Tacsimate is a mixture of seven organic acids and while there may be some ambiguity in the assignment, we found that a single malonate molecule best fitted the density. The location of the malonate appears significant in terms of lattice contacts in light of the other crystal forms of porcine trypsin described below.

Crystals grown in both experiments I-3 and I-34, as seen in Table 1, belong to space group $P4_12_12$. Because there was no PDB entry with the same space group, the structures first had to be solved by molecular replacement and refined. Also present in the PDB, however, were two entries for benzamidine-inhibited porcine trypsin having the same unit-cell parameters (1ldt and 1c9p) and also similar to the crystals investigated here, but belonging to space group $P4_32_12$. Thus, the small molecules present in the reagent mixes of experiments I-3 and I-34 induced a change in hand of the unique axis and resulted in a new crystal of the alternative enantiomorph.

The crystals grown in experiments I-3 from PEG and I-34 from Tacsimate, although belonging to the same space group, $P4_12_12$, were not otherwise identical. The unit-cell parameters of the crystals from experiment I-3 were rather close to those of the PDB entry 1ldt belonging to the enantiomorphic space group $P4_32_12$. As seen in Table 1, however, the unit-cell parameters of the crystals grown in experiment I-34 were significantly different. Thus, the components of reagent mix I-3 and I-34 both produced an enantiomorphic space group, but the particular components of the two reagent mixes determined the unit-cell parameters.

The contents of the asymmetric units of the $P4_12_12$ crystals from both experiments I-3 and I-34 are instructive. As before, there is one porcine trypsin molecule in the asymmetric unit with a benzamidine tightly sequestered at the active site in both cases. In the crystals from experiment I-3 we could in addition readily identify a molecule of oxamic acid and a molecule of sulfanilic acid from their electron-density masses. As seen in Fig. 7, both the oxamic acid and sulfanilic acid molecules reside on the exterior surface of the protein, both lie at the interfaces between two protein molecules in the lattice and both serve to link multiple protein molecules together through networks of hydrogen bonds. The oxamic acid occupies the same site that malonate occupied in the crystals of experiment I-18. It is almost certainly the sulfanilic acid that promotes the reversal of the hand of the molecular packing and the changes in the unit-cell parameters in these porcine trypsin crystals. Virtually every hydrogen-bond donor and acceptor of the sulfanilic acid is utilized in forming the intermolecular cross-links.

The asymmetric units of the crystals grown in experiment I-34, which also belong to the enantiomorphic space group $P4_12_12$ but have similar unit-cell parameters to those of PDB entry 1ldt, are different. In place of the sulfanilic acid, we found a molecule of sorbitol occupying virtually the same location. It thus appears that the presence or absence of a small molecule at this crucial lattice site determines the enantiomorph of the molecular packing. The specific nature of the small molecule located there, on the other hand, strongly influences the unit-cell parameters. At the site occupied by the oxamic acid in the crystals from experiment I-3, a small electron-density mass was also observed in the crystals from experiment I-34. We were unable to definitively identify the source of this density, but it would be consistent with either a formate or acetate ion present in the Tacsimate or a glycerol from the reagent mix.

4. Discussion

The results we obtained illustrate a number of possibilities for the interaction of the small molecules in the crystallization mixtures with proteins and how these may affect crystallization propensity, crystal form or crystal properties such as inherent order. Using X-ray diffraction, we analyzed about 50 different crystals grown in the three screening experiments that we described previously (McPherson & Cudney, 2006). In some of these, the contents of the unit cell appeared to be no different than for the same crystal grown in the absence of the

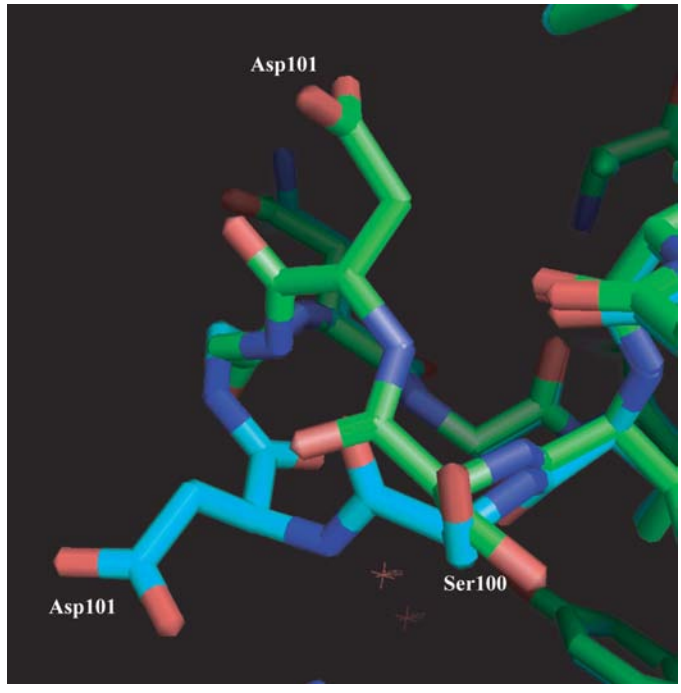

Figure 5
As a consequence of small-molecule binding at the active site, a small conformational change in a loop underlying the active site, a loop which is otherwise engaged in intermolecular interactions, altered the conformation. The normal conformation of Ser100 and Asn101 is shown in green and the conformation in the crystals grown in experiment I-22 is shown in blue.

multi-ligand mixes. This could be due to a number of reasons. The most obvious is that no ligand in a mix was incorporated into the crystal lattice and none contributed to the crystallization process either by making intermolecular interactions, altering protein conformation or surface properties or by ensuring a higher degree of stability. This was undoubtedly true for some crystals.

An alternate explanation for the apparent absence of small-molecule electron density is that the small molecules were in some way disordered. This might be particularly true for the linear dicarboxylic acids and polyamines that we investigated. This might also pertain in those cases, such as the active-site cleft of lysozyme in the presence of numerous ligands, where density was observed which did not seem to correspond to any

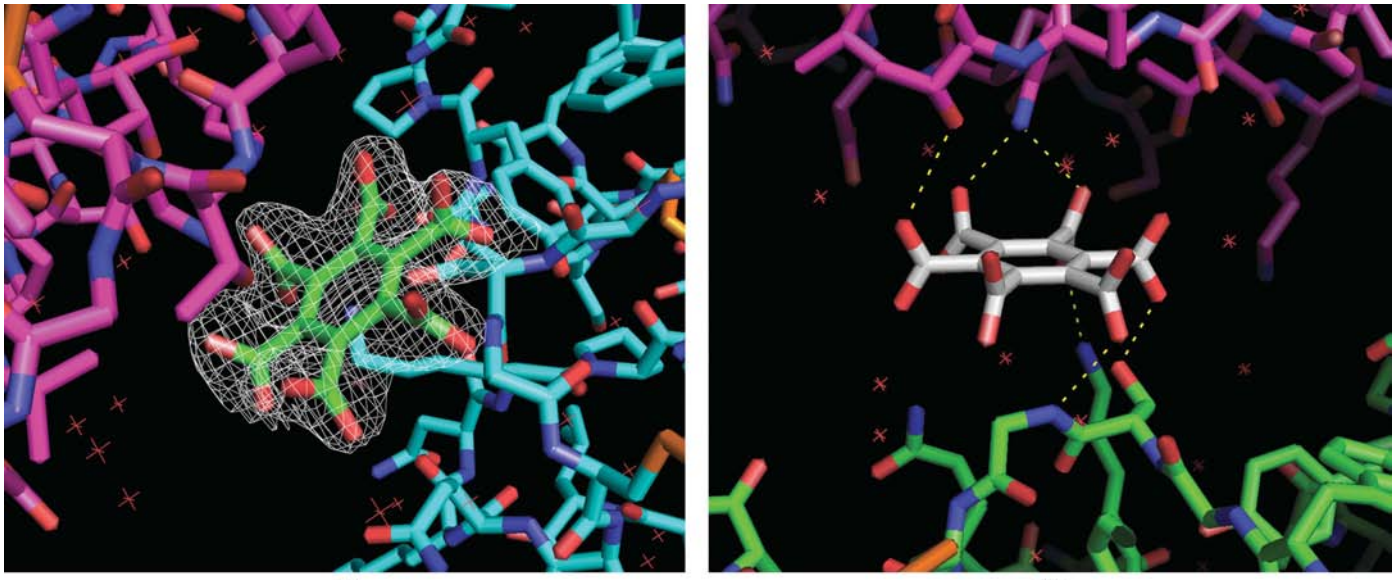

Figure 6
In (*a*) one of the eight crystallographically independent mellitic acid molecules maintaining the lattice of the bovine trypsin crystals grown in experiment II-33 is shown here with its electron density superimposed. The six carboxylate groups extending from the central ring are clearly represented in the electron density, calculated as a $2F_o - F_c$ map in which the mellitic acid molecule was omitted from the model. The resolution of the map is 1.80 Å and is contoured at 1.0σ. In (*b*) a mellitic acid molecule is shown at the interface between two molecules of bovine trypsin (green and pink), where it serves as the principal link between protein molecules in the lattice. The hydrogen-bonding network between the two protein molecules through the mellitic acid is illustrated by the dashed yellow lines. The hydrogen bond distances have a range of 2.68–3.26 Å.

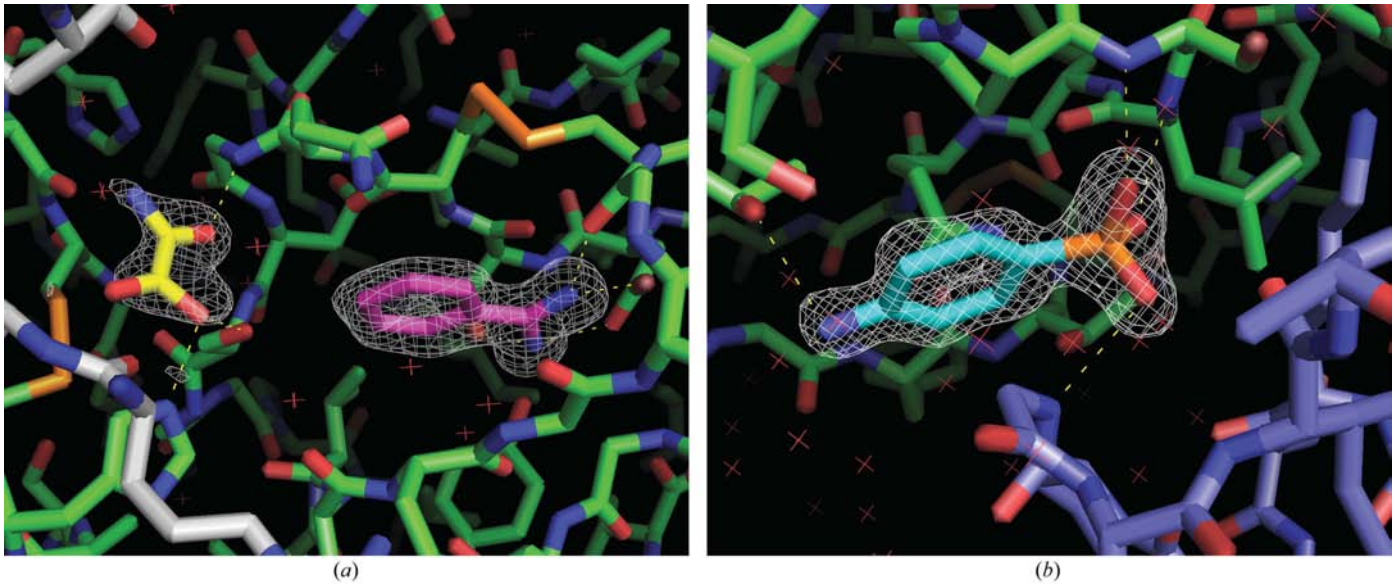

Figure 7
In experiment I-3, porcine trypsin binds benzamidine at the active site of the enzyme and both oxamic acid and sulfanilic acid at interfaces with crystallographically related molecules. The hydrogen bonds between ligands and protein, with distances from 2.43 to 2.94 Å, are shown as yellow dashed lines. In (*a*) the benzamidine (purple) and oxamate (yellow) molecules are superimposed on their electron densities. In (*b*) the same superposition is shown for the sulfanilic acid molecule (cyan) lying between molecules of porcine trypsin (green and blue). All electron densities are from $2F_o - F_c$ maps in which the small molecule was omitted from the model. The resolution is 1.28 Å and the contour levels are 1.0σ.

molecules in the mother liquor but nonetheless persisted as the model was refined. A final reason for failure to identify any small molecules in some crystal lattices is that of technical shortcomings. That is, the small molecules were simply too small in terms of number of atoms (*e.g.* acetate, formate) to provide unambiguous electron density and could not be discriminated from clusters of water or the X-ray data were not of sufficient quality or of high enough resolution.

On the other hand, as illustrated by the examples presented above, the small molecules often do interact in defined ways with protein molecules and do affect the way in which the proteins crystallize. In some cases, as with benzamidine in trypsin, a ligand binds tightly at the active site and does not form intermolecular bonds, yet enhances the crystallization of the protein. Without benzamidine present, neither bovine nor porcine trypsin crystallizes under the same conditions. In this case, the positive effect of the small molecule must be exerted through stabilization of the protein structure or by its promotion of imperceptibly small alterations in conformation or disposition of surface groups.

As illustrated by the lysozyme crystals grown in experiment I-22, significant conformational changes may be induced by interaction with small molecules. Although the ligand may not itself be directly involved in intermolecular interactions, the loop, the conformation of which is altered, probably is. Thus, a ligand indirectly influences lattice interactions by modifying the surface of the protein and we know from a long history of protein crystallography that minor changes in surface structure can be all-important in crystallization.

Also presented here are examples that clearly illustrate the validity of the hypothesis underlying the original crystallization experiments; that is, the formation of intermolecular lattice interactions through the incorporation of small molecules at protein–protein interfaces. In lysozyme (experiment I-4) and porcine trypsin (experiment I-3), *p*-aminobenzoic acid and sulfanilic acid, respectively, were clearly observed to be present in the lattice and in both crystals served to establish an interface between protein molecules. This kind of binding was exactly what we might have anticipated. The same was true of trimesic acid in another lysozyme crystal (experiment II-9), malonate, oxamic acid and sorbitol in porcine trypsin crystals and tartrate in crystals of thaumatin (experiment I-48). In all of these cases, virtually every hydrogen-bonding possibility inherent to the ligand was utilized in making the intermolecular interactions.

We noted that in some of the cases we analyzed, notably thaumatin and lysozyme, crystals were grown in the presence of certain reagent mixes from precipitant systems (*e.g.* PEG 3350) that otherwise would not produce crystals. We also noted that in some cases, perhaps more than we unambiguously identified, crystals contained more than one of the ligands present in the reagent mixture. This was true for some crystals of lysozyme, porcine trypsin and bovine trypsin. In the case of RNase A, two molecules of dGMP per protein molecule were seen.

The most impressive confirmations of the hypothesis to emerge from these X-ray analyses were, of course, RNase A in experiment III with reagent mix 18 and bovine trypsin in experiment II with reagent mix 33. Both of these crystal forms were unique in that they had not been previously reported and both represented unusual and unexpected superstructures in their respective crystals. In these crystals, components of the reagent mixes were almost entirely responsible for creating and maintaining the lattice. Intermolecular interactions involving dGMP with RNase A and mellitic acid with bovine trypsin were extensive and dominant. Although interactions were principally hydrogen bonds, hydrophobic interactions were also important, as in the RNase A structure. Thus, in at least the nine examples described here, the observations conformed well to our expectations. All of the results support the idea that 'cocktails' of small molecules may provide a highly useful alternative for promoting the crystallization of macromolecules.

This work was supported by NIH grant GM074899 for the establishment of the Center for High Throughput Structural Biology.

References

Ascenzi, P., Bolognesi, M., Guarneri, M., Menegatti, E. & Amiconi, G. (1984). *Mol. Cell. Biochem.* **64**, 139–144.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
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.
Chandonia, J. M. & Brenner, S. E. (2006). *Science*, **311**, 347–351.
DeLano, W. L. (2002). *The PyMOL Molecular Visualization System.* DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.
Jones, T. A. & Kjeldgaard, M. (1994). *O – The Manual*, version 5.10. Uppsala, Sweden: Uppsala University Press.
Ko, T.-P., Day, J., Greenwood, A. & McPherson, A. (1994). *Acta Cryst.* D**50**, 813–825.
McPherson, A., Brayer, G. D. & Morrison, R. D. (1986). *J. Mol. Biol.* **189**, 305–327.
McPherson, A. & Cudney, R. (2006). *J. Struct. Biol.* **156**, 387–406.
Norvell, J. C. & Machalek, A. Z. (2000). *Nature Struct. Biol.* **7**, 931.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Wlodawer, A. (1985). *Biological Macromolecules and Assemblies*, Vol. 2, edited by F. A. Jurnak & A. McPherson, pp. 393–439. New York: John Wiley & Sons.