Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of apo acyl carrier protein and a proposal to engineer protein crystallization through metal ions

Xiayang Qiu[a,b]* and Cheryl A. Janson[a,c]

[a]GlaxoSmithKline, King of Prussia, Pennsylvania 19406, USA, [b]Department of Exploratory Medicinal Sciences, MS4039, Pfizer Inc., Groton, Connecticut 06340, USA, and [c]Shamrock Structures, 1440 Davey Road, Woodridge, Illinois 60517, USA

Correspondence e-mail: xiayang_qiu@groton.pfizer.com

Received 13 May 2004
Accepted 23 June 2004

PDB References: apo acyl carrier protein, 1t8k, r1t8ksf

A topic of current interest is engineering surface mutations in order to improve the success rate of protein crystallization. This report explores the possibility of using metal-ion-mediated crystal-packing interactions to facilitate rational design. *Escherichia coli* apo acyl carrier protein was chosen as a test case because of its high content of negatively charged carboxylates suitable for metal binding with moderate affinity. The protein was successfully crystallized in the presence of zinc ions. The crystal structure was determined to 1.1 Å resolution with MAD phasing using anomalous signals from the co-crystallized Zn^{2+} ions. The case study suggested an integrated strategy for crystallization and structure solution of proteins *via* engineering surface Asp and Glu mutants, crystallizing them in the presence of metal ions such as Zn^{2+} and solving the structures using anomalous signals.

1. Introduction

Crystal structures of proteins are valuable to basic sciences and pharmaceutical design (Hol, 2000). Scientifically, structure elucidation may suggest functions for the vast number of novel genes from genomic sequencing of various living organisms. In industry, structure-based design has transformed from a fascinating idea into an integral component of drug-discovery technologies. New structures have been reported at a much-increased rate thanks to advances in molecular biology, biochemistry and crystallography, as well as the growth of public and private investments in this field. To crystallize a new protein, the current practice is to generate multiple constructs for protein expression and pass the protein products through a large number of crystallization trials using automation. Unfortunately, the chance of obtaining a crystal of a given protein construct does not correlate linearly with the total number of crystallization trials (Segelke, 2001). The overall success rate from cloned protein to good crystal is as low as ~10% (Chayen & Saridakis, 2002). In structure determination, solving the phase problem can still be a challenge. New and more rational strategies to facilitate protein crystallization and structure solution are clearly desirable.

A large surface area of a protein needs to be in contact with neighboring lattice molecules in order to produce good crystals (Carugo & Argos, 1997; Janin & Rodier, 1995). Unlike biologically relevant protein–protein interactions, the total buried surface area in a crystal is achieved *via* a larger number of protein partners, often 8–10, and a smaller interface area for each protein pair, typically 200–1200 $Å^2$; that is comparable to interface sizes generated from random computer simulations (Janin & Rodier, 1995). The standing perception is that protein crystals are formed through weak and nearly stochastic packing interactions. Most lattice contacts involve

© 2004 International Union of Crystallography
Printed in Denmark – all rights reserved

polar amino acids, as opposed to physiological oligomeric interfaces, which favor hydrophobic interactions (Dasgupta *et al.*, 1997). A large number of hydrophobic residues on a protein surface will limit the solubility of the protein and the non-specific nature of hydrophobic interactions would stipulate the formation of aggregates rather than well ordered lattices (Durbin & Feher, 1996). Biological protein oligomers can exploit hydrophobic interactions for higher specificity because they use larger interfaces and fewer direct partners and do not extend into the repetitive and macroscopic orders of crystals. Very strong interactions such as covalent disulfide-bond formation between free surface cysteines are often vehicles for aggregation rather than crystallization. Therefore, salt bridges and hydrogen bonds are usually the driving forces in crystallization.

Several strategies have emerged over the years for designing protein constructs for crystallization, such as delineating stably folded domains of proteins (Stewart *et al.*, 1998), reducing the size of flexible loops and mutating out post-translational modifications such as glycosylation (Kwong *et al.*, 1998). An area of particular interest is the introduction of surface modifications in order to facilitate crystal growth (Derewenda, 2004). Changes in protein surface are known to have an impact on crystallization (McElroy *et al.*, 1992; D'Arcy *et al.*, 1999; Charron *et al.*, 2002). Eminent successes include the alkylation of lysines in myosin subfragment-1 (Rayment, 1997) and the Phe185Lys mutation in HIV integrase (Dyda *et al.*, 1994). Some rational approaches have also been proposed, *e.g.* the design of β-strand pairing (Wingren *et al.*, 2003) and the mutation of high-entropy surface side chains (Lys, Gln and Glu) into Ala (Derewenda, 2004). The latter approach, which originated from the observation that lysine residues are not preferred in crystal contacts (Dasgupta *et al.*, 1997), has shown recent success. However, it is only suited to highly soluble proteins because a Lys-to-Ala mutation leads to lower protein solubility. Here, we propose a new strategy to mutate protein surface residues in order to introduce metal-mediated crystal lattice contacts. This approach, at its best, will not only improve the rate of success for crystallization but also enable *de novo* structure solutions with MAD phasing.

Metals are important factors for crystallization, usually because they are required for protein function and stability (McPherson, 1991; Durbin & Feher, 1996). H ferritin presents an early example of engineering metal-binding sites to enable lattice contacts (Lawson *et al.*, 1991). Based on homology to L ferritin, which required Cd^{2+} for lattice formation, the Lys86Gln mutant was designed to restore the Cd^{2+} site in H ferritin for crystallization. Engineering metal-binding sites for biochemical studies has focused on the introduction of strong metal-binding sites using cysteines and histidines and is therefore rarely relevant to protein crystallization (Pack *et al.*, 1997; Arnold & Haymore, 1991). For this work, we chose carboxylates (Glu and Asp) because they may bind metals with moderate strength (Auld, 2001) and in multiple configurations (Carrell *et al.*, 1988). Acyl carrier protein is a cytosolic protein of 77 amino acids in *Escherichia coli* (Lambalot & Walsh, 1997). Its holo form (holo-ACP) plays a central role in fatty-acid biosynthesis by acting as a coenzyme to shuttle acyl substrates and products between the biosynthetic enzymes. It is one of the most abundant soluble proteins in *E. coli*, with a cellular concentration of approximately 0.1 m*M* (Cronan & Rock, 1996). Apo acyl carrier protein (apo-ACP), which lacks the 4′-phosphopantetheine prosthetic group on Ser36, has 20 negatively charged carboxylate side chains and only five positively charged residues. Because of its natural abundance and carboxylate content (pI 4.2), apo-ACP was chosen as a test system for our design strategy.

Table 1
Diffraction data statistics.

Values in parentheses are for the last resolution shell.

Data set	Native	Peak	Inflection	Remote
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters				
a (Å)	27.5	27.5	27.5	27.5
b (Å)	41.3	41.2	41.2	41.2
c (Å)	62.5	62.8	62.8	62.8
Wavelength (Å)	0.9184	1.28242	1.28295	1.21940
Resolution (Å)	1.1	1.81	1.82	1.66
Mosaicity (°)	0.2	0.5	0.5	0.5
No. observed reflections	171171	79198	87664	99031
No. unique reflections	46801	12158	12269	15877
Redundancy	3.7	6.6	7.1	6.2
R_{merge}†	0.067 (0.199)	0.083 (0.155)	0.080 (0.194)	0.082 (0.200)
$I/\sigma(I)$	17.1 (2.6)	24.6 (4.5)	25.8 (4.2)	27.8 (4.1)
Completeness (%)	83.9 (36.7)	96.8 (69.2)	98.5 (86.3)	96.8 (70.4)
R_{ano}‡	0.062	0.116	0.115	0.073
f_o'/f_o''§ (e)	−0.1/2.2	−7.5/4.5	−10/3.3	−3.5/3.4

† $R_{merge} = \sum |I - \langle I \rangle| / \sum I$, where I is the intensity of a reflection and $\langle I \rangle$ is the average intensity. ‡ $R_{ano} = \sum |I^+ - I^-| / \sum \langle I \rangle$, where I^+ and I^- are the Friedel mates (counted separately in data statistics) and $\langle I \rangle$ is the average intensity. § f_o' and f_o'' are observed mean scattering factors.

2. Materials and methods

2.1. Protein production and crystallization

E. coli apo-ACP was overexpressed and purified using previously reported procedures (Lambalot & Walsh, 1997). Holo-ACP crystals in space group *C*2 were shown to diffract to 1.6 Å resolution (McRee *et al.*, 1985), but a structure solution has not yet been reported. Crystallization of the *C*2 form of holo-ACP did not include any metal ions as 1 m*M* EDTA was used in the preparation of the protein sample. We crystallized apo-ACP using the standard vapour-diffusion sitting-drop method. The well solution contained 0.1–0.5 *M* zinc acetate, 2–25% PEG (1K, 4K, 6K or 8K) and 0.1 *M* imidazole buffer pH 6.3–8.0. The crystallization drops were made using 2 µl well solution and 2 µl protein sample (10 mg ml^{-1} in 20 m*M* Tris buffer pH 7.5 and 50 m*M* NaCl). Plate-shaped crystals appeared within a couple of days and the presence of zinc was required for crystallization. To our knowledge, this is the first time an apo-ACP has been crystallized, despite many attempts in various laboratories (Roujeinikova *et al.*, 2002).

2.2. Diffraction data collection and MAD phasing

The crystals diffracted to better than 1.1 Å resolution at synchrotron sources. They belong to space group $P2_12_12_1$, with unit-cell parameters $a = 27.5$, $b = 41.3$, $c = 62.5$ Å. There is one molecule per asymmetric unit and an estimated solvent content of 40%. Diffraction data have been collected to 1.1 Å resolution. Owing to physical limits on detector distance and size (2 × 2 CCD), ~100% data completion was only achievable to 1.30 Å resolution at 0.9184 Å wavelength (Table 1). The 1.1 Å resolution native data set is 54.0% complete in the 1.18–1.14 Å shell with an average $I/\sigma(I)$ of 3.9, suggesting an effective resolution of about 1.14 Å. At the time, no crystal structure of ACP was available for structure solution by molecular replacement. NMR structures of ACP were too different to be useful for structure solution, an experience shared by others (Roujeinikova *et al.*, 2002). With the hypothesis that zinc could be bound specifically to bridge crystal lattice interactions, MAD data sets were collected at three wavelengths near the zinc absorption *K* edge (1.28 Å) to 1.8 Å resolution. The high redundancy of the data (Table 1) allowed Friedel mates to be scaled separately during data processing to maximize the observation of anomalous signals. The statistics for anomalous signals indicated good quality for the MAD data (Table 1). Three major zinc sites were quickly identified using the program *SOLVE* (Terwilliger, 2003), resulting in a *Z* score of 15.6. MAD phasing using these three zinc sites gave an overall figure of merit of 0.60 to 1.8 Å resolution (FOM of 0.76 from 20–6.35 Å resolution, 0.59 at ~2.17 Å resolution and 0.34 at ~1.86 Å resolution). Solvent flattening was then used to improve the overall FOM to 0.77 (0.85 to 2.75 Å resolution, 0.75 to 2.00 Å resolution and 0.62 at ~1.8 Å resolution).

2.3. Model building and refinement

Using the electron-density map calculated from MAD phasing and solvent flattening, the automated tracing program *wARP* (Morris *et al.*, 2003) was able to find 35 of the 77 residues. The partial model gave a connectivity of 0.77 and a figure of merit of 0.86. Instead of initiating manual tracing of the excellent electron-density map at 1.8 Å resolution, the coordinates of the 35 residues were used as the starting structure for *wARP* tracing of the 1.1 Å resolution native data, yielding a model of 67 residues and an *R* factor of 0.279. One round of manual fitting allowed the modelling of all 77 residues. Subsequent *REFMAC* (Winn *et al.*, 2003) refinement was carried out using all the data from 20.0 to 1.1 Å resolution (46 056 reflections, weight 4.0). Solvent molecules were identified using $(F_o - F_c)$ difference maps at 3σ, $(2F_o - F_c)$ maps at 1σ and appropriate distances to N and O atoms and were kept only when their $(2F_o - F_c)$ density was greater than 1.5σ after being included in refinement. Zinc occupancies were manually adjusted and refined to minimize difference electron density. Anisotropic *B* factors were introduced in the last stage of refinement; H atoms were not used. There are 607 protein atoms, ten zinc ions, six imidazole ions and 100 water molecules in the final model, giving an *R* factor of 0.133 (last shell, 0.170) and an R_{free} of 0.150 (last shell, 0.150). Alternate conformations were observed for Ser36 and Asp56, while the side chains of Glu57 and Lys61 are partially disordered. All hydrogen-bond donors are involved in hydrogen bonds to protein or solvent atoms. The Thr52 carbonyl O atom is the only acceptor that does not form a hydrogen bond, but is instead buried near the Glu47 side-chain C atoms with a distance of 3.1 Å. The final structure has an average *B* factor of 11 Å^2; all but three water molecules have *B* factors of less than 30 Å^2. The r.m.s. (root-mean-square) deviations are 0.014 Å for bond distances, 1.5° for bond angles, 4.7° for torsion angles, 0.1 Å^3 for chiral centres and 0.009 Å for planar groups. The r.m.s. *B* factors are 1.5 and 3.0 Å^2 for main-chain and side-chain atoms, respectively. The overall coordinate error is 0.03 Å as estimated by *REFMAC*. The Ramachandran plot shows that 94.3% of the residues are in the most favored regions and none are in generously allowed or disallowed regions. The overall geometry of the final model satisfies the criteria for a 1.1 Å resolution crystal structure (PDB code 1t8k).

3. Results

3.1. The structure of apo-ACP

The crystal structure of holo-ACP from *Bacillus subtilis* has been solved in complex with holo-ACP synthase (Parris *et al.*, 2000). The structures of selenomethionine-substituted *Escherichia coli* butyryl-ACP and its I62M mutant have been determined to 2.0 and 1.2 Å resolution, respectively (Roujeinikova *et al.*, 2002). The prosthetic group is disordered in the butyryl-ACP structure, but ordered in the I62M mutant. Here,

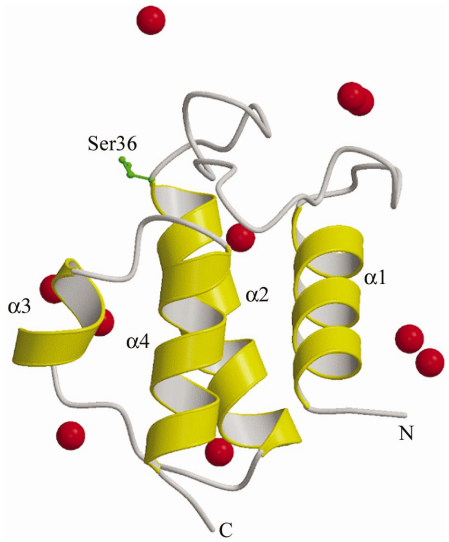

Figure 1
Crystal structure of *E. coli* apo-ACP. The four helices are shown in yellow and labelled. Ser36 is shown in green and labelled. N and C denote the amino- and carboxyl-termini of the protein. Red spheres are the bound zinc ions in the asymmetric unit. Additional zinc sites also exist on the protein surface, but they are crystal symmetry-related sites and have been arbitrarily slotted into a neighboring protein molecule. This figure was generated with the program *MolScript* (Kraulis, 1991).

we report the first crystal structure of an apo-ACP in its native form at 1.1 Å resolution. Similar to the known ACP structures, the overall fold of *E. coli* apo-ACP is a four-helical bundle of up and down topology (Fig. 1). There is a hydrophobic pocket in the core of the four-helical bundle that accommodates the Ser36-phosphopantetheine-linked acyl group. Helices $\alpha1$, $\alpha2$ and $\alpha4$ are largely responsible for the stability of the helical bundle fold, while helix $\alpha3$, the $\alpha1\alpha2$ loop and $\alpha3\alpha4$ loop account for most of the flexibility in this class of proteins (Roujeinikova *et al.*, 2002).

Despite being crystallized in similar crystal lattices, the structures of apo-ACP and butyryl-ACP I62M differ substantially (C^{α} r.m.s.d. of 0.56 Å), especially in the areas surrounding the acyl-binding pocket. Backbone shifts of 1 Å are observed in helix $\alpha3$ and loop $\alpha3\alpha4$ which provide Ala59, Ile62 and Thr63 for the binding of acyl groups. Other regions of notable differences include parts of the $\alpha1\alpha2$ loop and helix $\alpha2$, all of which are elements involved in acyl binding (*e.g.* Phe28 and Thr39). The apo-ACP structure overlays better with that of butyryl-ACP (crystallized in space group $P6_3$), with a C^{α} r.m.s.d. of 0.43 Å. The acyl-binding pockets are unoccupied in both structures and the backbone differences are only about 0.4 Å in the aforementioned acyl-binding regions. This suggests that the conformation of structural elements correlates with ligand binding, rather than random mobility. The most interesting structural difference occurs at the N-terminus of helix $\alpha2$, where the backbone atoms of Ser36 and Leu37 are shifted 1.0 and 1.4 Å, respectively, compared with those in butyryl-ACP and butyryl-ACP I62M. This is likely to reflect the presence or absence of the prosthetic group at Ser36 in the three structures. In our apo-ACP structure, the Ser36 side chain is observed in two alternate conformations (Fig. 2), with the O^{γ} atom pointing towards or away from the N-cap of helix $\alpha2$. This may be relevant for the required flexibility of the prosthetic arm that would swing out to participate in various enzymatic reactions and then swing back to insert the acyl groups into the core hydrophobic pocket for protection. The former conformation would put the prosthetic phosphate group in a position to interact favorably with the $\alpha2$ helical dipole. The latter configuration would lose the interaction with the $\alpha2$ helical dipole, but would permit a prosthetic amide N atom to interact with the helix $\alpha3$ C-terminal dipole (*i.e.* the carbonyl of Glu60) and the acyl chain to be buried in the core hydrophobic pocket. Our structure provides a plausible explanation for how the prosthetic arm of ACP can swing between two distinct conformations with a relatively low energy barrier. Clearly, the helices and conformational changes of ACP are highly evolved and carefully orchestrated for functional reasons.

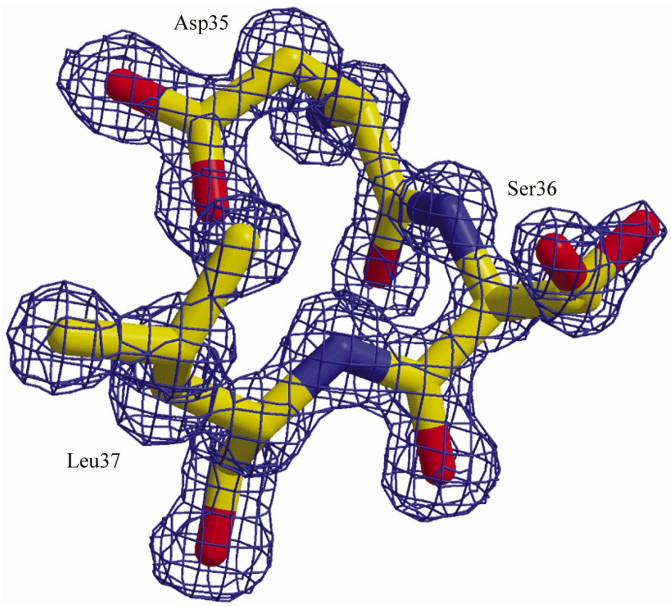

Figure 2
The N-cap of helix $\alpha2$. The $(2F_o - F_c)$ electron density is contoured at 1.5σ. Ser36 is shown in two alternate conformations. Yellow atoms are carbon, blue nitrogen and red oxygen. This and the following figures were generated using the program *XtalView* (McRee, 1993).

Table 2
Anomalous difference Fourier peaks (σ), occupancy, B factors and ligands of zinc ions.

Superscripts denote residues from neighboring molecules (IM, imidazole; W, water).

Zn^{2+}	Peak	Occupancy	B factor (Å^2)	Ligand
Zn1	44.2	1.0	6.6	Glu21, IM1, ^{1}Asp35, ^{1}Asp38
Zn2	41.7	1.0	7.4	Glu58, His75, IM4, ^{2}Asp31
Zn3	24.8	0.75	10.7	Glu5, W49, ^{3}Glu47, ^{3}Glu53
Zn4	23.4	0.75	10.6	Glu13, IM2, W54, ^{4}Asp51
Zn5	21.3	0.70	10.8	Glu48, IM3, W80, ^{5}Glu49
Zn6	14.7	0.40	10.0	Asp56, Glu60, W22, W37
Zn7	14.3	0.50	12.1	Asp56, IM5, IM6, W83
Zn8	9.2	0.35	12.4	W84, W86, W74, W100
Zn9	8.6	0.40	15.2	Glu41, W72, W108, ^{6}IM1
Zn10	8.0	0.35	14.6	Glu20, W57, W64

3.2. Crystal packing

The estimated solvent content of 40% and diffraction limit of better than 1.1 Å resolution suggested that our crystals were tightly packed with well ordered apo-ACP molecules. This was further demonstrated by the low overall B factor (11 Å^2) and the fact that all 77 amino acids were clearly resolved. The structure revealed ten bound Zn^{2+} ions (Fig. 1), which were confirmed by their distinctive peaks in the anomalous difference Fourier maps (Table 2). A total of 22 zinc ions are found to interact with each protein molecule after applying crystal symmetry operations and a distance cutoff of 5.0 Å. These zinc ions are spread broadly over the protein surface (Fig. 3*a*). Each apo-ACP molecule is in direct contact with eight neighboring apo-ACP molecules (Fig. 3*b*), with a buried accessible surface area of 300–600 Å^2 between each protein pair. Remarkably, every pair of the lattice protein–protein interactions involves zinc-bridged contacts (Fig. 3*b*). While metal ions have previously been observed assisting lattice interactions (Durbin & Feher, 1996), our apo-ACP crystal has probably set a new record for the extent that metal ions can be utilized in protein crystal packing.

There are five positively charged side chains in apo-ACP. The single Arg6 is paired with the Glu48 side chain from an

adjacent molecule. The Lys61 side chain is partially disordered, but the other three lysines are ordered and make ion pairs with acidic side chains from neighboring molecules, *i.e.* Lys8 to Asp56, Lys9 to Asp51 and Lys18 to Asp38. Mutating away surface lysine has become a popular strategy in engineering protein for crystallization (Derewenda, 2004). The fact that 75% of the surface lysines in our crystal form good lattice interactions does not invalidate the unfavorable perception of lysine, but rather reaffirms that the strength of salt bridges can overcome the high entropy of lysine side chains in order to form stable interactions. One could argue that lysine is probably less harmful to crystallization for proteins abundant in carboxylates, such as apo-ACP.

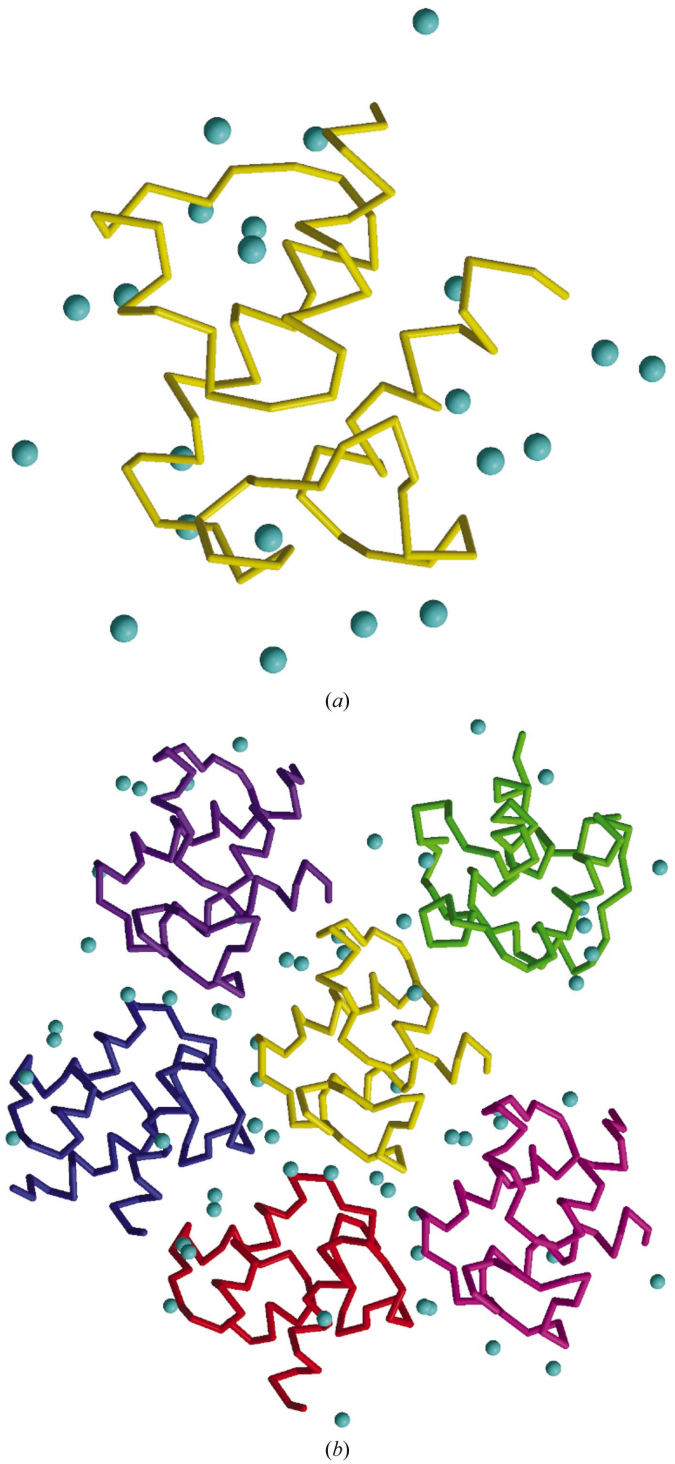

Figure 3
Packing of apo-ACP molecules in the crystal lattice. (*a*) The C^{α} tracing of apo-ACP is shown in yellow. The 22 bound zinc ions after applying symmetry operations are shown as cyan balls. (*b*) The yellow molecule in the middle is that defined by the asymmetric unit of the structure determination. Only five of the neighboring protein molecules in the same viewing plane have been shown in this figure and are coloured red, magenta, green, purple and blue. Three other neighboring protein molecules have been omitted for clarity. The cyan spheres show the bound zinc ions.

3.3. Metal-binding sites

Zn1–Zn5 contribute directly to lattice formation by coordinating with residues from five different neighboring apo-ACP molecules. The five zinc ions appear to bind tightly, as indicated by their high anomalous difference peaks, good occupancies and low *B* factors (Table 2) and are considered to be primary zinc sites. Nearly ideal tetrahedral coordination geometry characterizes Zn^{2+} binding at the primary zinc sites (Fig. 4). Two or three carboxylates are involved in zinc binding, including at least one carboxylate from each of the neighboring proteins. All the carboxylates bind zinc in the *syn* (*Z*) conformation, which is more basic and more stable than the *anti* (*E*) conformation (Carrell *et al.*, 1988). The tetrahedral coordination is completed by at least one solvent molecule in each of the zinc sites (Fig. 4), which agrees with the observation that three protein ligands are sufficient for potent zinc sites (Auld, 2001). In Zn1, Zn2, Zn4 and Zn5, imidazole ions are involved in zinc coordination. The single histidine in apo-ACP, His75, binds Zn2 (Fig. 4). These observations agree with the prior knowledge that Zn^{2+} has a statistical ligand preference for imidazole (Auld, 2001). It is possible, however, to build potent Zn^{2+} sites without involving imidazole moieties, as demonstrated by the Zn3 site in our crystal (Fig. 4).

The other five Zn^{2+} ions, Zn6–Zn10, help to stabilize crystal packing in a less direct manner. They are secondary zinc sites because of their lower occupancies (0.35–0.50). Zn6 and Zn7 are adjacent to each other with a metal-to-metal distance of 3.9 Å (Fig. 5). Because they have to share the same Asp56 side chain in two alternate conformations, the Zn^{2+} occupancies of these two sites have an upper limit of about 0.50. Tetrahedral geometry is preserved in the two zinc sites, but more solvent molecules are involved as zinc ligands (Fig. 5). Glu60 binds Zn6 in the *anti* (*E*) conformation, which is usually a weaker interaction. These features suggest lower affinities for Zn6 and Zn7 compared with those of the primary zinc ions. Zn6 and Zn7 do not bind directly to any neighboring protein, but their ligands are involved in crystal-packing interactions (Fig. 5). Zn8 is incompletely hydrated by at least four water molecules with an irregular coordination geometry. The hydrated Zn^{2+} ion has a much-increased size and is able to interact with the carboxylate side chain of Glu30, Asp70 from an adjacent protein and the carboxyl-terminus of Ala77 from a third neighboring molecule. Zn9 interacts with a symmetry-related

imidazole IM1, one of the Zn1 ligands. Zn10 is bound to Glu20, which interacts with Ser36 of a neighboring protein molecule. All the observed zinc ions contribute to various degrees towards the formation of the crystal lattice, which explains why zinc was essential for the growth of our apo-ACP crystals.

3.4. Structure perturbations by metal ions

Two Zn^{2+} ions were observed in the SeMet butyryl-ACP crystal (PDB code 1l0h). Their *B* factors are 49 and 29 Å^2 for Zn80 and Zn81, respectively, suggesting that they are not tightly bound to the protein. Zn80 is linked to Glu4, W62 and Glu57 and Glu60 of a neighboring protein. This site is different from all the ten sites in apo-ACP. Zn81 interacts with Asp31, W16 and His75 and Glu58 of a symmetry-related molecule. This site resembles our Zn2 site except for the replacement of a water molecule by an imidazole ion. Overlay of this and the apo-ACP structures revealed little backbone difference caused by zinc binding, but several metal ligands show different side-chain rotamers indicative of metal-induced conformational changes.

The butyryl-ACP I62M structure contains seven Zn^{2+} ions (PDB code 1l0h). The crystal is nearly isomorphous to that of apo-ACP, with a notable 3% difference in the *c* unit-cell parameter. All five primary zinc sites of apo-ACP have counterparts in the butyryl-ACP I62M crystal. Zn1002 interacts with Glu13, Asp51 of a neighboring molecule, a cacodylate and a water molecule, similar to the Zn4 site (Fig. 4*d*). Zn1004 is associated with Asp31, His75 of a different protein and two water molecules. This site is related to Zn2 (Fig. 4*b*) but does not use Glu58 as its ligand. Zn1006 is bound to Glu48, Glu49 of an adjacent protein and two water molecules. This site resembles Zn5 (Fig. 4*e*) but lacks the imidazole ligand. Zn1009 coordinates with Glu21, a symmetry-related Asp35 and two water molecules. Its location is equivalent to our Zn1 site, but without the benefit of the Asp38 ligand (Fig. 4*a*). Zn1008 is the only site that uses identical ligands to its counterpart in apo-ACP, Zn3 (Fig. 4*c*). The other two zinc ions do not overlay with the secondary zinc sites of apo-ACP and make little impact on crystal packing. Only three metal ligands show different side-chain conformations between the two structures. Therefore, metal perturbation of proteins is limited to the conformation of side chains rather than main chains.

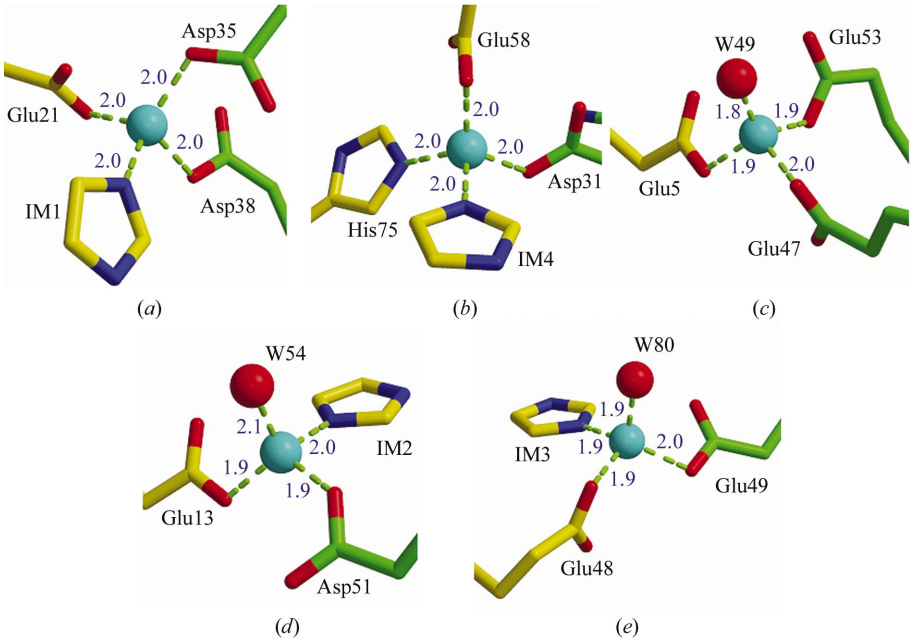

Figure 4
The primary zinc sites. (*a*) Zn1 site. (*b*) Zn2 site. (*c*) Zn3 site. (*d*) Zn4 site. (*e*) Zn5 site. Cyan balls are zinc ions. Carboxylates from neighboring molecules are shown in green. Green broken lines denote metal-to-ligand bonds, with distances labelled in Å. IM and W are imidazole and water molecules. General atom colours are red for oxygen, blue for nitrogen and yellow or green for carbon.

4. Discussion

The central goal of this study is to explore the possibility of engineering protein-surface residues to enable metal-mediated lattice contacts and the use of this method to aid protein crystallization and structure solution. There are many examples of metal-assisted crystal contacts (Durbin & Feher, 1996) and the use of metals as additives to improve crystal diffraction (Chang *et al.*, 1998). However, little attention has been paid to proactively engineering metal-binding sites for protein crystallization, except when prompted by specific prior knowledge (Lawson *et al.*, 1991). We propose the introduction of carboxylates by mutagenesis to spread the protein surface. The mutant proteins would go into crystallization trials in the presence of metal ions such as Zn^{2+}. Metal–ligand interactions can be manipulated with a reasonable range of ion concentrations and parameters such as pH, ionic strength and moderate chelators. Once the mutant protein is crystallized with metal ions, it is possible to solve the crystal structure using the anomalous signals from the metals.

4.1. The rationale for removing surface protrusions

Crystal packing should obey physical and chemical principles for protein–protein interactions, such as a good complement of shape and charge. Shape alterations *via* engineering will be limited to domain delineation and flexible loop removal (Stewart *et al.*, 1998; Kwong *et al.*, 1998) because the authenticity of the resulting structure should not be compromised. Besides shape, charge should be the most pronounced feature of protein surfaces because of the greater strength and reach of its effects. Aside from the N- and C- termini, four types of side chains

are almost always charged in proteins: Arg, Lys, Asp and Glu. A subset of the charged residues will be solvent-exposed and play a critical role in forming a stable lattice (Takahashi *et al.*, 1993), agreeing with the observation that the top residues at crystal contacts are Lys, Glu, Arg, Ser and Asp (Dasgupta *et al.*, 1997). However, lysine becomes the least favored amino acid for lattice contacts after the abundances of surface-residue types are factored in, which led to the hypothesis that higher side-chain entropy is the reason for difficulties in crystallization (Derewenda, 2004). It is counterintuitive to us that the free-energy costs for stabilizing a couple of lysine side chains ($\sim$8 kJ mol^{-1} per residue) would be so critical to crystallization considering the overall size of the lattice interfaces.

The most favorable interaction for a charged residue is formation of a salt bridge with counterions. A neglected property of salt bridges is asymmetry, which requires twice the variables and is thus inherently more difficult to design than symmetrical bindings such as metal-mediated carboxylate interactions. For example, there are 36 (6×6) possibilities for forming a salt bridge using six positive and six negative changes, but 144 (12×12) combinations for creating a metal-mediated symmetrical interaction with 12 carboxylates. The possible number of lattice salt bridges per protein is often not large, *e.g.* five for apo-ACP. This number, along with asymmetry and specificity constraints, suggests that the actual number of salt bridges per protein in crystals can be small and some surface charges may easily be left without a suitable mate in lattices. The unbalanced charges can either be excluded from the lattice interface or be mismatched with less favorable atoms. The former may reduce the total number of packing possibilities, but causes little harm to the stability of an obtainable lattice. The latter is not much of a problem for uncharged atoms because of the higher chances of finding suitable partners among all types of side chains and main chains. In addition, uncharged atoms have limited strength and reach, allowing crystals to tolerate a certain level of mismatches at interfaces (Dasgupta *et al.*, 1997). Mismatching negative charges is less problematic because Glu and Asp can form direct contacts with each other at low pH or be mediated by water or metal ions (Flocco & Mowbray, 1995). A Lys–Lys pair has limited repulsive power owing to high entropy, but arginine is more rigid and will not favor Arg–Arg or Arg–Lys pairs (Dasgupta *et al.*, 1997). When a surface negative charge is available, arginine should find it first because of its protruding size and the superior strength of the resulting ion pair. The side-chain entropy of lysine is a disadvantage here because a weaker bond will be produced. When a positive charge has to be excluded from lattice interfaces, it is easier to hide a lysine than an arginine because the latter is bulkier and more rigid. Therefore, arginines should be more frequently incorporated into crystal lattices, as observed by statistical analyses of crystallized proteins (Dasgupta *et al.*, 1997).

However, the aforementioned analysis does not necessarily imply that arginine is a welcome addition for crystallization. Many proteins may be recalcitrant to crystallization because they lack the means to simultaneously charge-balance or hide all protruding bulkiness and charges of surface arginines and, to a lesser degree, lysines. This explains the lack of success in mutating Lys to Arg for crystallization (Dasgupta *et al.*, 1997). In addition, it provides a different rationale for the successes of Lys-to-Ala mutations (Derewenda, 2004), *i.e.* working mostly through removal of unfavorable shape and charge protrusions, rather than significantly reducing the overall entropy of lattices. This rationale is further supported by the following observations (Derewenda, 2004): (i) the Lys-to-Ala approach is especially effective for Lys clusters, which are more difficult to charge-balance or exclude from interfaces; (ii) the resulting mutant forms tight lattice contacts at the exact site of mutations, suggesting that the removed Lys bulk was not compatible with the observed crystal packing and (iii) Gln-to-Ala experiments have not met with drastic improvements in crystallization. If our reasoning is valid, mutating away Arg for unyielding proteins could be more effective than the Lys-removal approach, especially when the protein of interest has surface arginines that could form clusters of surface positive charges.

Finally, the metal-mediated engineering strategy can be regarded as a new approach to artificially flatten the surface landscapes of proteins and lower the energy barrier between alternative crystal packing states. Because surface charges will not be decreased, the strategy should not compromise the solubility of the mutant proteins. The symmetrical nature of metal-mediated carboxylate interactions should provide a

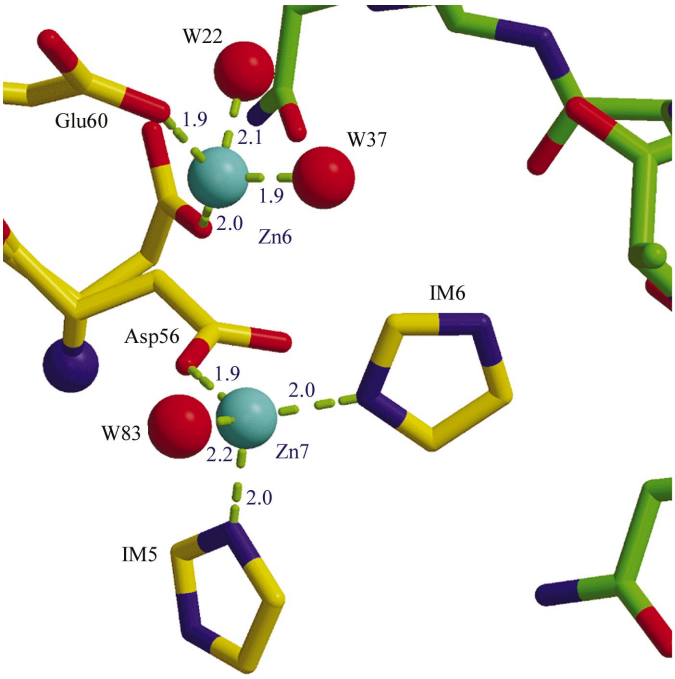

Figure 5
Two of the adjacent secondary zinc sites Zn6 and Zn7. Asp56 is shown in two alternate conformations as it binds each of the zinc ions, drawn as cyan spheres. Amino acids from a neighboring molecule are shown in green. Green broken lines denote metal-to-ligand bonds, with their distances labelled. IM and W are imidazole and water molecules. General atom colours are red for oxygen, blue for nitrogen and yellow or green for carbon.

higher chance of forming lattice contacts than normal salt bridges.

4.2. Case-study implications

Instead of mutating a protein to introduce carboxylate side chains, our test case used apo-ACP, which is naturally abundant in carboxylates, as an initial proof of the concept. No ACP had been crystallized with metal ions at the time and the crystal structure of ACP was unknown. Zn^{2+} was chosen for the study because it is cheap, safe and stable and has a synchrotron-accessible absorption edge at 1.28 Å, which allows structure solution using MAD methods. There are 20 carboxylate side chains in apo-ACP, including 14 glutamates and six aspartates (see supplementary data[1]). 16 of the carboxylates coordinate directly to metal ions, including 11 glutamic acids and five aspartic acids (Table 2). Additionally, Glu30 binds the hydrated Zn8 ion, leaving only three carboxylates (Glu4, Asp57 and Glu70) that are not involved in metal binding. Therefore, the percentage of carboxylates involved in zinc coordination is 85%. 12 carboxylates are direct ligands in primary zinc sites (Fig. 4), giving a ratio of 60% even in the most stringent terms. This suggests high successes rates (60–85%) for achieving metal-mediated interactions using surface Asp and Glu residues.

The rate of success should also be high for using zinc anomalous signals or MAD to solve structures. Great strides have been made in recent years in the areas of data acquisition, locating heavy-atom sites, phasing and phase-improvement algorithms. For proteins of reasonable sizes and crystals of good diffraction, anomalous signals from multiple zinc ions should be sufficient for *de novo* structure determination based on our experiences. Heavier atoms could be soaked in to replace some of the zinc sites for phasing if necessary (Qiu *et al.*, 1995).

4.3. The choice of mutants

Known metal sites are built with a larger number of Asp–Asp and Asp–Glu pairings than Glu–Glu pairings (Flocco & Mowbray, 1995). However, this observation was mostly drawn from biologically relevant metal sites, which may be more potent than those necessary for crystallization. In apo-ACP crystals, Asp and Glu are utilized in similar proportions as metal ligands and two of the primary zinc sites involve only Glu–Glu pairs (Fig. 4). Asp is less flexible than Glu, which may provide stronger bonds but support fewer binding modes. In the absence of the formation of metal-mediated interactions, Glu is less favorable than Asp in lattice interactions (Dasgupta *et al.*, 1997). Further experiments are needed to demonstrate the preference between Asp and Glu, or whether there is one, for our engineering strategy. At this point, Asp mutations are recommended unless there are too few Glu residues in the protein.

The locations for introducing Glu or Asp may be suggested by crude modelling if possible, prompted by knowledge such as exposed hydrophobic residues or undesirable surface Cys residues, or derived from functional or sequence analyses. Histidine mutations are not recommended, but are allowed, especially when they exist naturally. When positively charged surface clusters are predicted to be present, mutating Arg or Lys to Asp or Glu is suggested instead of the Lys-to-Ala approach (Derewenda, 2004). This should alleviate the solubility problem of the Ala mutants. It is desirable to mutate in regions of reasonable stability rather than high mobility. For example, terminal His tags bind metals, but rarely contribute to lattice formation (O'Neill *et al.*, 2001). Finally, the mutants should not hinder protein function, a common prerequisite for all engineering approaches. Under these conditions, metal-induced side-chain movements in the obtained structure are not a cause for concern, because many of these side chains are not native to the protein but are introduced *via* mutagenesis.

To allow 8–10 crystal contacts (Janin & Rodier, 1995), 16–20 metal-binding half-sites are needed; each can be constructed with one or two carboxylates (Fig. 4). Many Asp and Glu residues may exist naturally on protein surfaces (20 in apo-ACP), especially for proteins with low pI values, which are more prevalent in nature than those with high pI values. Surface histidines and the C-terminus may provide additional metal ligands; thus, a small number of carboxylate mutations are probably sufficient for many proteins, *e.g.* zero for apo-ACP. A larger number of carboxylates on protein surfaces could have a compounding effect on the success rate of crystallization, which can be further explored if necessary. Our goal is to improve the rate of crystallization, not to showcase lattices solely built by metal ions. There should be little downside to replacing unfavorable surface residues (Cys, hydrophobic, Lys and Arg clusters) by Asp or Glu, even if the resulting crystals contain a minimal number of metal ions.

4.4. Effectors in crystallization

The innate chemical potential of zinc is not very different from that of other first-row transition metals. However, the lack of redox chemistry because of the filled orbital d^{10} makes Zn^{2+} function as a stable Lewis acid and accept a pair of electrons. There is no inherent energy barrier between different ligand geometries of zinc, although most of the observed zinc sites in proteins have a slightly distorted tetrahedral geometry (Auld, 2001). Zn^{2+} does not have a strong preference for coordinating with nitrogen, sulfur or oxygen, but histidine is the most commonly observed ligand for zinc, followed by cysteine; hence the perception is that histidine is a more potent zinc ligand than carboxylate. Zinc ions have many advantages for our approach, but other metal ions should be included in crystallization trials in order to sample the sizes and coordination geometries (Carrell *et al.*, 1988). Metal ions and their chelators could have a positive effect on the stability of the protein or its mutants (Arnold & Haymore, 1991). In apo-ACP, six imidazole ions were observed as zinc ligands. Chelators such as imidazoles are known to bind zinc (Wolpert

[1] Supplementary data have been deposited in the IUCr electronic archive (Reference: HM5013). Services for accessing these data are described at the back of the journal.

et al., 1977; Cedergren-Zeppezauer, 1983) and enhance zinc affinity (Mock & Wang, 1999). For our purposes, chelators such as imidazole should be tested at a few concentrations for crystallization. Potent chelators such as EDTA are not desirable because they will simply strip metals off proteins. Ionic strength and pH can be varied during crystallization in order to manipulate the affinity of metal–ligand interactions. Because metal binding requires deprotonated carboxylates and imidazoles, pH values much lower than 6.0 are not desirable. When the pH of the solution reaches 8.0 or higher, many transition metals will start to precipitate as hydroxides. Very high ionic strength could weaken metal-mediated interactions and should be generally avoided. Not surprisingly, all the zinc-containing ACP crystals were obtained under low ionic strength conditions, while the zinc-less ACP crystal was grown with 1 m*M* EDTA and 78% saturated ammonium sulfate (McRee *et al.*, 1985).

With our strategy, crystallization requires the screening of fewer conditions, including a narrower pH range (*e.g.* 6–8), low ionic strength (*e.g.* ±50 m*M* NaCl or KCl), a few metal ions (*e.g.* Zn^{2+}, Ni^{2+}, Mn^{2+}, Ca^{2+}, Mg^{2+} at 1–200 m*M*), metal chelators (*e.g.* imidazole at 1–200 m*M*), temperature, organic precipitants (*e.g.* PEG or MPD) and $(NH_4)_2SO_4$ because it is not necessarily harmful for charged interactions and could provide sulfate ions as metal ligands (Qiu *et al.*, 1996). These conditions can be sampled fairly thoroughly within the size of common screens (*e.g.* 96) and used in the initial crystallization trials of the mutant proteins.

4.5. Conclusion

In this report, a new strategy has been proposed for systematically mutating undesirable (Cys, hydrophobic, Arg and Lys clusters) or non-essential protein surface residues into Asp or Glu and carrying out crystallization trials in the presence of metal ions such as Zn^{2+}. Crystals obtained with this method can be used directly for *de novo* structure solution by MAD phasing. While the current apo-ACP case study is of limited scope, this proposal offers an integrated approach for engineering proteins for crystallization and structure solution. It is unlikely that any single strategy will solve the last two hurdles in protein crystallography once and for all. The intention of our study is to stimulate further thinking and research in this area. Hopefully, the combination of this and other strategies would lead to a further breakthrough in our ability to elucidate the three-dimensional structures of proteins.

We thank Drs Howard Kallender and Arun Patel for protein expression and purification. The native data set was collected at the end of 2000 on beamline ID14-1 of ESRF (European Synchrotron Radiation Facility) at Grenoble, France. The MAD data sets were collected at the beginning of 2001 on beamline 17ID of the IMCA-CAT (Industrial Macromolecular Crystallography Association Collaborative Access Team) at the Advanced Photon Source of Argonne National Laboratory, USA. Use of the IMCA-CAT beamline 17-ID at the Advanced Photon Source was supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with Illinois Institute of Technology. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38.

References

Arnold, F. H. & Haymore, B. L. (1991). *Science*, **252**, 1796–1799.
Auld, D. S. (2001). *Biometals*, **14**, 271–313.
Carrell, C. J., Carrell, H. L., Erlebacher, J. & Glusker, J. P. (1988). *J. Am. Chem. Soc.* **110**, 8651–8656.
Carugo, O. & Argos, P. (1997). *Protein Sci.* **6**, 2261–2263.
Cedergren-Zeppezauer, E. (1983). *Biochemistry*, **22**, 5761–5772.
Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T. & Rees, D. C. (1998). *Science*, **282**, 2220–2226.
Charron, C., Kern, D. & Giegé, R. (2002). *Acta Cryst.* D**58**, 1729–1733.
Chayen, N. E. & Saridakis, E. (2002). *Acta Cryst.* D**58**, 921–927.
Cronan, J. E. Jr & Rock, C. O. (1996). *Escherichia coli and Salmonella: Cellular and Molecular Biology*, edited by F. C. Neidhardt, pp. 612–638. Washington, DC: ASM Press.
D'Arcy, A., Stihle, M., Kostrewa, D. & Dale, G. (1999). *Acta Cryst.* D**55**, 1623–1625.
Dasgupta, S., Iyer, G. H., Bryant, S. H., Lawrence, C. E. & Bell, J. A. (1997). *Proteins*, **28**, 494–514.
Derewenda, Z. S. (2004). *Structure*, **12**, 529–535.
Durbin, S. D. & Feher, G. (1996). *Annu. Rev. Phys. Chem.* **47**, 171–204.
Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). *Science*, **266**, 1981–1986.
Flocco, M. M. & Mowbray, S. L. (1995). *J. Mol. Biol.* **254**, 96–105.
Hol, W. G. J. (2000). *Nature Struct. Biol.* **7** *Suppl.*, 964–966.
Janin, J. & Rodier, F. (1995). *Proteins*, **23**, 580–587.
Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946–950.
Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998). *Nature (London)*, **393**, 648–659.
Lambalot, R. H. & Walsh, C. T. (1997). *Methods Enzymol.* **279**, 254–262.
Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M. A., Livingstone, J. C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G., Thomas, C. D., Shaw, W. V. & Harrison, P. M. (1991). *Nature (London)*, **349**, 541–544.
McElroy, H. E., Sisson, G. W., Schoettlin, W. E., Aust, R. M. & Villafranca, J. E. (1992). *J. Cryst. Growth*, **122**, 265–272.
McPherson, A. (1991). *J. Cryst. Growth*, **110**, 1–10.
McRee, D. E. (1993). *Practical Protein Crystallography*. San Diego: Academic Press.
McRee, D. E., Richardson, J. S. & Richardson, D. C. (1985). *J. Mol. Biol.* **182**, 467–468.
Mock, W. L. & Wang, L. (1999). *Biochem. Biophys. Res. Commun.* **257**, 239–243.
Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). *Methods Enzymol.* **374**, 229–244.
O'Neill, J. W., Kim, D. E., Baker, D. & Zhang, K. Y. (2001). *Acta Cryst.* D**57**, 480–487.
Pack, D. W., Chen, G., Maloney, K., Chen, C.-T. & Arnold, F. H. (1997). *J. Am. Chem. Soc.* **119**, 2479–2487.
Parris, K. D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C. C., Seehra, J. & Somers, W. S. (2000). *Structure Fold. Des.* **8**, 883–895.
Qiu, X., Pohl, E., Holmes, R. K. & Hol, W. G. (1996). *Biochemistry*, **35**, 12292–12302.

Qiu, X., Verlinde, C. L., Zhang, S., Schmitt, M. P., Holmes, R. K. & Hol, W. G. (1995). *Structure*, **3**, 87–100.
Rayment, I. (1997). *Methods Enzymol.* **276**, 171–179.
Roujeinikova, A., Baldock, C., Simon, W. J., Gilroy, J., Baker, P. J., Stuitje, A. R., Rice, D. W., Slabas, A. R. & Rafferty, J. B. (2002). *Structure*, **10**, 825–835.
Segelke, B. W. (2001) *J. Cryst. Growth*, **232**, 553–562.
Stewart, L., Redinbo, M. R., Qiu, X., Hol, W. G. & Champoux, J. J. (1998). *Science*, **279**, 1534–1541.
Takahashi, T., Endo, S. & Nagayama, K. (1993). *J. Mol. Biol.* **234**, 421–432.
Terwilliger, T. C. (2003). *Methods Enzymol.* **374**, 22–37.
Wingren, C., Edmundson, A. B. & Borrebaeck, C. A. K. (2003). *Protein Eng.* **16**, 255–264.
Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
Wolpert, H. R., Strader, C. D. & Khalifah, R. G. (1977). *Biochemistry*, **16**, 5717–5721.