Entropy and surface engineering in protein crystallization

Zygmunt S. Derewenda[a]* and Peter G. Vekilov[b]

[a]Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine, Charlottesville, Virginia 229008-0736, USA, and [b]Department of Chemical Engineering, University of Houston, 4800 Calhoun Avenue, Houston, TX 77204-4004, USA

Correspondence e-mail: zsd4n@virginia.edu

Protein crystallization remains a key limiting step in the characterization of the atomic structures of proteins and their complexes by X-ray diffraction methods. Current data indicate that standard screening procedures applied to soluble well folded prokaryotic proteins yield X-ray diffraction crystals with an ~20% success rate and for eukaryotic proteins this figure may be significantly lower. Protein crystallization is predominantly dependent on entropic effects and the driving force appears to be the release of ordered water from the sites of crystal contacts. This is countered by the entropic cost of ordering of protein molecules and by the loss of conformational freedom of side chains involved in the crystal contacts. Mutational surface engineering designed to create patches with low conformational entropy and thereby conducive to formation of crystal contacts promises to be an effective tool allowing direct enhancement of the success rate of macromolecular crystallization.

1. Introduction

The limiting step in macromolecular crystallography (assuming the availability of large amounts of the purified target protein) is the preparation of X-ray diffraction-quality crystals. For decades there has only been anecdotal data regarding how difficult it is to crystallize a protein and until recently the success rates of crystallization were never rigorously evaluated, because virtually no one reported negative results. The advent of structural genomics (Burley & Bonanno, 2003) finally allowed a more detailed evaluation, since statistics for all projects and attrition rates are routinely recorded and are available. These data point to an ~20% average crystallization success rate (*i.e.* preparation of X-ray quality crystals) for proteins that express in soluble form (Hui & Edwards, 2003). General extrapolation of these statistics is difficult because in most cases the data relate to prokaryotic single-domain proteins selected for ease of expression and crystallization, so that the database is biased and the statistics are over optimistic. On the other hand, the structural genomics pipelines often neglect unique aspects of the particular protein's chemistry such as potentially stabilizing ligands, inhibitors *etc.*, which might critically facilitate crystallization. Thus, the estimates of success rates, while suggestive, should be taken with caution. Suffice it to say that only a fraction of proteins will succumb to crystallization efforts.

Ironically, detailed knowledge of the physics and thermodynamics of crystallization has essentially no predictive value and crystallization of new proteins is always performed by screening methods. The sample is screened against a multitude

of commercially prepared solutions, often with little attention to protein's chemistry or function. Is it possible to design a more rational approach that would increase the success rate of protein crystallization and put the endeavor on a more rational foundation?

2. The thermodynamics of crystallization

As any process in nature at constant temperature and pressure, the transfer of protein molecules from solution to the crystal is governed by the change of Gibbs free energy, $\Delta G^{\circ}_{\text{cryst}}$. According to the Gibbs–Helmholtz equation, the change in $\Delta G^{\circ}_{\text{cryst}}$ at constant temperature T can be expressed as the net effect of the opposing contributions of the enthalpy (or latent heat) $\Delta H^{\circ}_{\text{cryst}}$ and entropy $\Delta S^{\circ}_{\text{cryst}}$ as

$$\Delta G^{\circ}_{\text{cryst}} = \Delta H^{\circ}_{\text{cryst}} - T\Delta S^{\circ}_{\text{cryst}}. \qquad (1)$$

If $\Delta G^{\circ}_{\text{cryst}}$ is negative, the process becomes thermodynamically favored and the associated crystallization equilibrium constant can be expressed as

$$K_{\text{cryst}} = \exp(-\Delta G^{\circ}_{\text{cryst}}/RT), \qquad (2)$$

where R (the universal gas constant) is 8.314 J mol^{-1} K^{-1} and T is the absolute temperature.

Recent analyses of protein crystallization thermodynamics have shown that $\Delta G^{\circ}_{\text{cryst}}$ is only moderately negative, i.e. within the range −10 to −100 kJ mol^{-1} (Vekilov, 2003). This is in contrast to the crystallization of inorganic salts, e.g. NaCl, where the absolute free-energy change can be of significantly greater magnitude. In the case of protein crystallization $\Delta G^{\circ}_{\text{cryst}}$ is small and can easily be shifted to positive values by concurrent solution phenomena, making crystallization thermodynamically impossible. This explains why protein crystallization is so sensitive to even the slightest changes in conditions.

For further insight, we can evaluate the individual contributions of $\Delta H^{\circ}_{\text{cryst}}$ and $\Delta S^{\circ}_{\text{cryst}}$ in (1) to determine whether crystallization is enthalpy or entropy driven. In those few cases where experimental enthalpy determinations are available, $\Delta H^{\circ}_{\text{cryst}}$ was either only moderately negative, e.g. −70 kJ mol^{-1} for lysozyme (Schall et al., 1996), or insignificant, i.e. ∼0 kJ mol^{-1} as measured for ferritin, apoferritin and lumazine synthase (Yau et al., 2000; Petsev et al., 2001; Gliko et al., 2005). This is consistent with the notion that protein crystallization does not involve formation of strong intermolecular bonds in the crystal lattice. In the unique case of haemoglobin C, a surprisingly large positive enthalpy ($\Delta H^{\circ}_{\text{cryst}}$) of 155 kJ mol^{-1} indicated that the enthalpy effects strongly disfavour crystallization of this protein (Vekilov, Feeling-Taylor, Petsev et al., 2002; Vekilov, Feeling-Taylor, Yau et al., 2002). Thus, enthalpy changes are unlikely to rationalize crystallization in a general sense and at least in some cases may be very unfavourable.

If enthalpy is not the defining factor, then it must be entropy. This may seem unexpected, because crystallization is prohibitively disfavored by a massive negative change in entropy as three-dimensional order is imposed on the mole-cules in the crystal lattice. Indeed, this entropy cost consists of a loss of six translational and rotational degrees of freedom per protein molecule and is only marginally compensated for by the newly created vibrational degrees of freedom (Finkelstein & Janin, 1989; Tidor & Karplus, 1994). Theoretical estimates suggest that this results in an average change of $\Delta S^{\circ}_{\text{cryst}}$ in the range of −100 to −300 J mol^{-1} K^{-1} (Finkelstein & Janin, 1989; Tidor & Karplus, 1994; Fersht, 1999). At room temperature this leads to an unfavorable energy barrier of 30–100 kJ mol^{-1}. So unless this entropy loss is compensated for by gains arising from other factors, no crystallization will occur. Where does one find the compensating factors?

To better answer this question, let us consider the microscopic structure of the solvent (Vekilov, 2003, 2004). The above-mentioned studies of protein crystallization thermodynamics allowed an estimation of the enthalpy, entropy and the standard free-energy changes as functions of the temperature and of the composition of the respective solutions (Yau et al., 2000; Vekilov, Feeling-Taylor, Petsev et al., 2002; Vekilov, Feeling-Taylor, Yau et al., 2002; Bergeron et al., 2003). Additional data derived from the investigations of the interactions between protein molecules in solution yielded intermolecular interaction potentials for these proteins (Petsev et al., 2000, 2001; Petsev, Chen et al., 2003; Petsev, Wu et al., 2003) consistent with many other biophysical data (Leckband & Israelachvili, 2001; Israelachvili, 1995). Taken together, these data show that water, either by itself or with other small solvent molecules, is clearly structured around both the hydrophobic and hydrophilic patches on the surface of the protein molecules. The thickness of the structured layer is two to three water molecules deep, ∼7 Å in thickness (Ball, 2003; Pal & Zewail, 2004). The microscopic structure of this layer is probably very similar to that visualized by many protein crystal structures (Madhusudan et al., 1993). Within this 'biological' solvent layer (Pal & Zewail, 2004; Bhattacharyya et al., 2003), the water molecules are either firmly attached to the protein surface or relatively 'free'. The two states are in equilibrium with one another. An equilibrium also exits between the biological layer and the bulk solution water (Pal & Zewail, 2004). It has been observed that enthalpy and entropy contributions from the biological solvent layer largely determine the thermodynamics of molecular recognition during crystallization in a way similar to phenomena associated with enzyme–substrate and DNA–drug binding (Chalikian et al., 1994, 1999).

Upon incorporation into a crystal lattice, some of the structured water/solvent molecules are released or conversely additional water/solvent molecules may be trapped. Both phenomena would have a significant entropy effect: the transfer of water from clathrate, crystal hydrate or other ice-like structures leads to an entropy gain of ∼22 J mol^{-1} K^{-1} (Fersht, 1999; Tanford, 1980; Eisenberg & Kauzmann, 1969; Eisenberg & Crothers, 1979; Dunitz, 1994) and it has been suggested that the entropy effect of the water structured around protein molecules would be similar. Considering the complexity and importance of the solvent entropy effects, we

should explicitly distinguish between solvent and protein entropy changes during crystallization as follows,

$$\Delta G_{\mathrm{cryst}}^{\circ} = \Delta H_{\mathrm{cryst}}^{\circ} - T(\Delta S_{\mathrm{protein}}^{\circ} + \Delta S_{\mathrm{solvent}}^{\circ})_{\mathrm{cryst}}. \quad (3)$$

To deconvolute $\Delta S_{\mathrm{protein}}^{\circ}$ and $\Delta S_{\mathrm{solvent}}^{\circ}$, *in situ* atomic force microscopy with molecular resolution (Yip & Ward, 1996; Kuznetsov *et al.*, 2000; Yau *et al.*, 2000; Chen & Vekilov, 2002; Reviakine *et al.*, 2003) images of growing crystal surfaces were employed (Yau *et al.*, 2000; Vekilov, Feeling-Taylot, Yau *et al.*, 2002; Burton *et al.*, 1951; Swartzentruber *et al.*, 1990; Giesen *et al.*, 1996; Kuipers *et al.*, 1993; Vekilov & Chernov, 2002). The resulting $\Delta S_{\mathrm{protein}}^{\circ}$ values reach, as expected, up to -100 J mol^{-1} K^{-1}.

Thus, to overcome the previously discussed enthalpic and entropic energy barriers, $\Delta S_{\mathrm{solvent}}^{\circ}$ must be significantly positive. Indeed, the estimated values of $\Delta S_{\mathrm{solvent}}^{\circ}$ range from 100 J mol^{-1} K^{-1} to more than 600 J mol^{-1} K^{-1}, corresponding to the release of $\sim$5 to 30 water or solvent molecules upon the incorporation of a protein molecule into a crystal (Vekilov, 2003, 2004; Vekilov & Chernov, 2002). It therefore appears that in a general case the release of structured water from the protein's surface is the main thermodynamic driving force for crystallization (Vekilov, 2003; Vekilov, Feeling-Taylor, Yau *et al.*, 2002).

Although the above-described thermodynamic approach yields considerable insight into the mechanisms and the driving forces of macromolecular crystallization, it suffers from one significant oversimplification: while assigning microscopic properties to the solvent, it does not consider the microscopic properties of the protein surface structure: the protein molecules are viewed as rigid spheres, ignoring the structure of their surfaces. A better model for the protein molecules is a rigid body enveloped by the sheath of mostly conformationally labile high-entropy side chains, whose chemical nature affects the water structure. Small residues with short side chains (*e.g.* Ala, Thr *etc.*) are more likely to allow substantial ordering of water and solvent, while the more common large residues (*e.g.* Lys, Glu) have the opposite effect. Further, upon the formation of intermolecular contacts in the lattice of the crystal, the residues directly involved in those contacts lose conformational entropy, increasing the thermodynamic cost of crystallization.

Thus, $\Delta S_{\mathrm{protein}}^{\circ}$ has contributions both from the loss of the molecular degrees of freedom as well as from the loss of conformational freedom of those amino acids that make up the crystal contact interfaces.

Clearly, the protein's microscopic surface properties have a critical impact on the thermodynamics and kinetics of crystallization. It follows then that some proteins will crystallize more easily than others and that the amino-acid composition and sequence are more informative with respect to possible crystallization outcome than is normally believed.

3. Homolog screening: a relic of the past?

The notions that some proteins crystallize better than others is not new *per se*. More than 50 years ago John Kendrew

recognized this phenomenon and screened myoglobins from several animal species before deciding to pursue the crystal structure of the sperm whale protein (Kendrew *et al.*, 1954). Two decades later, Campbell *et al.* (1972) explicitly formulated the principle of homolog screening by stating that if a particular enzyme resists crystallization efforts, a homologous one from a related species should be tried. This approach allowed the crystallization and solution of the structures of most of the enzymes in the glycolytic pathway.

Shortly after the first successful expression of recombinant proteins in *Escherichia coli*, including insulin (Goeddel *et al.*, 1979) and somatostatin (Itakura *et al.*, 1997), protein crystallographers turned to recombinant methods as the means to obtain samples for crystallization. Today, with the notable exception of many membrane proteins, most samples for crystallization are obtained by recombinant methods and mutational modification of the target molecule is easy. In the meantime, homolog screening continued to enjoy support (Yee *et al.*, 2003) and has even gained in popularity owing to the role it now plays in membrane-protein crystallization (Wiener, 2004).

It is arguable that homolog screening is a relic of the bygone era when it was the only means of diversifying the protein sample. In reality, it suffers from a major limitation: the crystallizability of any given homolog is as unpredictable as that of the original target. Why, then, are we not using a more rational approach to protein modification as a means to enhance protein crystallizability?

4. Mutational enhancement of protein crystallizability

One of the first examples of the application of protein engineering to crystallization was the work of Lawson and colleagues, who a decade ago mutated human ferritin H-chain to generate the same crystal contacts as those in the rat L ferritin (Lawson *et al.*, 1991). A replacement of Lys86 found in the human sequence with Glu which occurs in rat recreated a Cd^{2+}-binding bridge which mediates crystal contacts in the rat homolog. This was followed in 1992 by a seminal study by Villafranca and coworkers (McElroy *et al.*, 1992), who showed that even single-site amino-acid substitutions in thymidylate synthase substantially affected the crystallization of the protein, although unlike the ferritin work their results were not easily rationalized in structural terms. Since then, several examples of successful crystallization of mutants in lieu of wild-type proteins have been reported, most by serendipity rather than by design. The obese protein leptin E100 (Zhang *et al.*, 1997) was only crystallized after a single-site mutant Trp100→Glu was used. Schwede and coworkers crystallized histidine ammonia-lyase by exchange of a surface cysteine residue (Schwede *et al.*, 1999). GroEL was crystallized using samples with two mutations accidentally introduced by PCR (Braig *et al.*, 1994; Horwich, 2000). Very recently, D'Arcy and coworkers at Hoffman–La Roche used systematic mutagenesis of surface residues in DNA gyrase B subunit to study the effect of surface substitutions on the ability of the protein to

crystallize and obtained new crystal forms (D'Arcy *et al.*, 1999).

Considering the breadth of macromolecular crystallography, the efforts to improve crystallizability by protein engineering are conspicuously few in number and very limited in scope. Why did the concept not find wider use in protein crystallography? There are a number of possible reasons. It is generally believed that for a protein of unknown structure it is not trivial to identify with confidence potential target sites for point mutagenesis to enhance crystallization. Perhaps more importantly, no rational strategy has ever been formulated to predict what kind of mutation would directly improve crystallization (as opposed to altering solubility). It has also been argued that mutations might alter the native structure of a protein, whereas the natural variants, *i.e.* homologs, typically retain both structure and function. Last, but not least, additional labor and direct costs required to generate mutant proteins for crystallization were until recently often prohibitive in academic laboratories. Contemporary advances in molecular biology removed the last impediment by introducing a plethora of user-friendly kits which allow efficient preparation of multiple protein variants, including single- and multiple-site mutants at the cDNA level, within hours. This is definitely simpler than subcloning of novel homologs and optimization *de novo* of expression and purification protocols. What is still needed is a strategy that would clearly enhance protein crystallizability, rather than simply introduce yet another variable into the crystallization screening process.

5. The concept of surface-entropy reduction

Recently, a method was proposed to enhance protein crystallizability by generating 'low-entropy' surface patches through site-directed mutagenesis (Derewenda, 2004*b*; Longenecker, Garrard *et al.*, 2001). This relatively simple concept is firmly rooted in the thermodynamic principles of crystallization (see above) and specifically stems from the assumption that the entropic cost of burying surface residues at crystal contact regions may seriously impede the crystallization process. Looking back at our thermodynamic arguments, this would significantly contribute to the already negative $\Delta S^{\circ}_{\text{protein}}$ in (3).

It is becoming recognized that most globular proteins have evolved a 'surface-entropy shield' created mostly by Lys and Glu residues, which prevents non-specific aggregation and precipitation (Doye, 2004). In most cases, spontaneous crystallization of proteins *in vivo* actually leads to serious diseases (Doye, 2004). The surface-entropy reduction concept states that mutating residues with high conformational energy, such as Lys and/or Glu, to alanines or other small amino acids should be an effective way to reduce surface conformational entropy. Statistically, lysines constitute 5.8% of the total amino-acid content in proteins, ranging from zero in rare and small acidic proteins to over 10% in some microtubule-associated proteins. These residues are predominantly located on the surface, with 68% exposed, 26% partly exposed and only 6% buried (Baud & Karlin, 1999), making them the most solvent-exposed residues in proteins. Even in a protein with a unique sequence, lysines provide a better than 90% chance of targeting a solvent-accessible site. The high conformational entropy of a solvent-exposed side chain, $T\Delta S \simeq 8$ kJ mol^{-1} at room temperature (Avbelj & Fele, 1998), is likely to impair the formation of protein–protein contacts involving these residues. It is important in this context to realise that even in a protein with average lysine content (5.8%), these amino acids will constitute a large fraction of the surface area because most are solvent-accessible. In fact, it has been calculated that lysines account on average for 12–15% of the solvent-accessible surface (DoConte *et al.*, 1999). As shown by DoConte *et al.* (1999), lysines constitute only 5.4% of interface surface in oligomeric proteins, compared with 14.9% of the total solvent-accessible surface. This clearly indicates that protein–protein interactions disfavor lysines. Glutamic acid (Glu) also occurs most frequently at the surface, with only 12% buried. On average Glu constitutes 6.1% of amino-acid content in proteins (Baud & Karlin, 1999), although proteins with a much higher content are quite common. For glutamate, conformational $T\Delta S$ under normal conditions is estimated to be ~7 kJ mol^{-1}, depending on the secondary-structure context (Avbelj & Fele, 1998). Protein interfaces discriminate against Glu almost as much they do against Lys. For example, in a studied set of oligomeric proteins, the percentage of interface surface attributed to Glu was 4.1%, in contrast to 10.3% of the exposed surface (DoConte *et al.*, 1999). The cumulative content of Lys and Glu in many proteins reaches the range 15–20%. This, in turn, is likely to translate into 30–50% of the total solvent-exposed surface and over 400 kJ mol^{-1} of conformational entropy. Quite often proteins contain surface sites where Glu and Lys are found in proximity and this offers an opportunity of double or triple mutations, which we thought might be particularly useful. The proximity of Glu and Lys can also be used as a suitable landmark for a solvent-exposed site, allowing higher confidence in the design of the mutants.

In principle, glutamine (Gln) may also serve as a target for crystal engineering. The obvious advantage is that it carries no charge and the Gln→Ala mutation should not be destabilizing, yet should reduce the entropic cost of crystallization. However, glutamines occur more commonly as buried residues (17%) and the probability of disturbing the integrity of the protein structure with a Gln→Ala mutation is more significant when either Lys or Glu are targeted. It should also be noted that glutamines occur with significantly lower frequency in proteins (3.7%) than either Lys or Glu. Thus, glutamines can be considered when in close proximity to a Lys or a Glu, but perhaps not on their own.

Interestingly, arginines do not appear to be good targets in spite of their equally high conformational entropy. They are not discriminated against at interfaces and constitute a higher percentage of buried surfaces (9.9%) than exposed surfaces (8.4%). For similar reasons, it has been suggested in the past that replacing lysines with arginines should also facilitate crystallization (Dasgupta *et al.*, 1997).

research papers

6. Proof of principle

To test the concept of surface-entropy reduction, a series of mutational experiments were conducted using a model system of the globular domain of the human regulatory protein RhoGDI. This molecule, in spite of its relatively small size, is difficult to crystallize and its surface is rich in Lys and Glu, which constitute nearly 20% of the combined amino-acid content (Fig. 1). Lysines and glutamates occurring singly and in clusters of 2–3 such residues within a short stretch of sequence were replaced with alanines (in one case with Arg) and the mutants were screened for crystallization. The results were dramatic (Longenecker, Garrard *et al.*, 2001; Mateja *et al.*, 2002; Czepas *et al.*, 2004). A vast majority of the mutations resulted in an enhanced tendency of the protein to crystallize. Interestingly, the double and triple cluster mutants involving neighboring Lys and Glu residues typically showed the highest potential to yield new crystal forms. What was perhaps most interesting was that the crystal contacts for multiple mutants were in most cases directly mediated by the mutated epitopes, validating the expected mechanism of a causal relationship between mutagenesis and crystallization. Finally, some of the novel crystal forms of mutant protein exhibited superior diffraction, despite the lower stability of the mutant or significantly higher solvent content compared with wild-type crystals (Mateja *et al.*, 2002). This observation is critically important: it shows that the nature of crystal contacts is the primary determinant of the physical quality of the crystal.

The pilot studies using RhoGDI paved the way for the application of the surface-entropy reduction protocol to novel proteins recalcitrant to crystallization in their wild-type form.
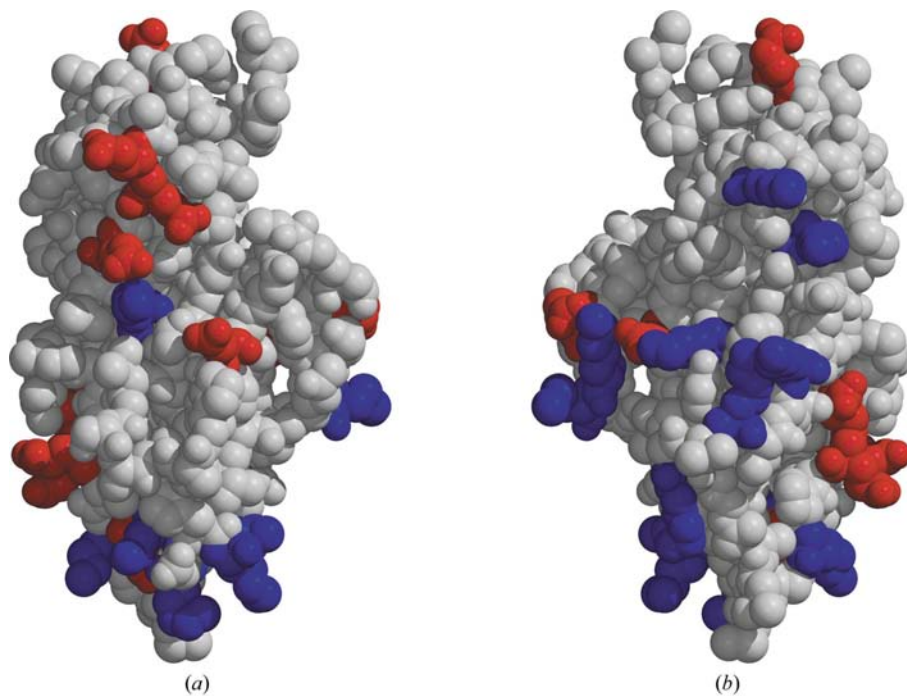
7. Case studies

The first of several proteins crystallized to date by surface-entropy reduction were the RGSL domain from the nucleotide-exchange factor PDZRhoGEF (Garrard *et al.*, 2001; Longenecker, Lewis *et al.*, 2001) and the *Yersinia pestis* (plague pathogen) antigen LcrV (Derewenda *et al.*, 2004*a*). In both cases the wild-type proteins did not crystallize despite extensive efforts. In particular, the *Y. pestis* LcrV antigen had been the target of numerous attempts at structural investigation, all of which had been in vain (Nilles, 2004). More recently, a number of *Bacillus subtilis* proteins that had been shown by others to be recalcitrant to crystallization in the wild-type form were also crystallized. In five cases, crystallization was accomplished successfully and crystal structures of the respective proteins were easily solved (Table 1). These include the *B. subtilis* homolog of an Hsp33 chaperone (Janda, Devedjiev, Cooper *et al.*, 2004), a hypothetical protein YdeN, a member of a new family of α/β hydrolases (Janda, Devedjiev, Derewenda *et al.*, 2004), the *YkoF* gene product (Devedjiev *et al.*, 2004), the *B. subtilis* YkuD protein, a member of the LysM domain-containing superfamily which include a variety of enzymes involved in bacterial cell-wall degradation (Bielnicki *et al.*, in preparation), and the *B. subtilis* homolog of the hydroperoxide-resistance protein Ohr (Cooper *et al.*, in preparation). In addition, two other groups used the surface-entropy reduction approach to crystallize the CUE domain of Vps9p with ubiquitin (Prag *et al.*, 2003) and to obtain high-quality crystals of the kinase domain of the insulin-like growth factor receptor (Munshi *et al.*, 2003). Some of the details of the structures solved using crystals obtained by surface engineering are shown in Table 1. While the details of each crystallization and crystal contacts are described in the respective publications, here we provide an overview of the results.

One of the most important questions is whether the premise of surface-entropy reduction is validated by the experimental data. It is arguable that enhanced crystallizability *per se* does not constitute proof that entropic effects are behind the method's success. Replacement of large polar residues with alanines affects the protein's pI, changes the electrostatic potential surface and decreases the solubility. In principle, any one of these effects might account for the enhanced crystallizability. However, there are compelling reasons to believe that the data are consistent with the theoretical premise. The high success rate clearly shows that the surface-engineering strategy is directly responsible for the observed crystallizability and the causal relationship is reaffirmed by the fact that in

Figure 1
A van der Waals surface representation of human RhoGDI with glutamate and lysine side chains visualized: (*a*) and (*b*) correspond to two views of the molecule after a rotation of 180° around the vertical axis.

virtually all cases the crystal contacts are formed by the mutated patches (Figs. 2 and 3). Solubility, pI and electro-statics are macroscopic in nature. This means that reduced solubility does not rationalize why specific contacts are made. The same applies to electrostatics. On the other hand, the premise of low-entropy crystal contacts exploits the micro-
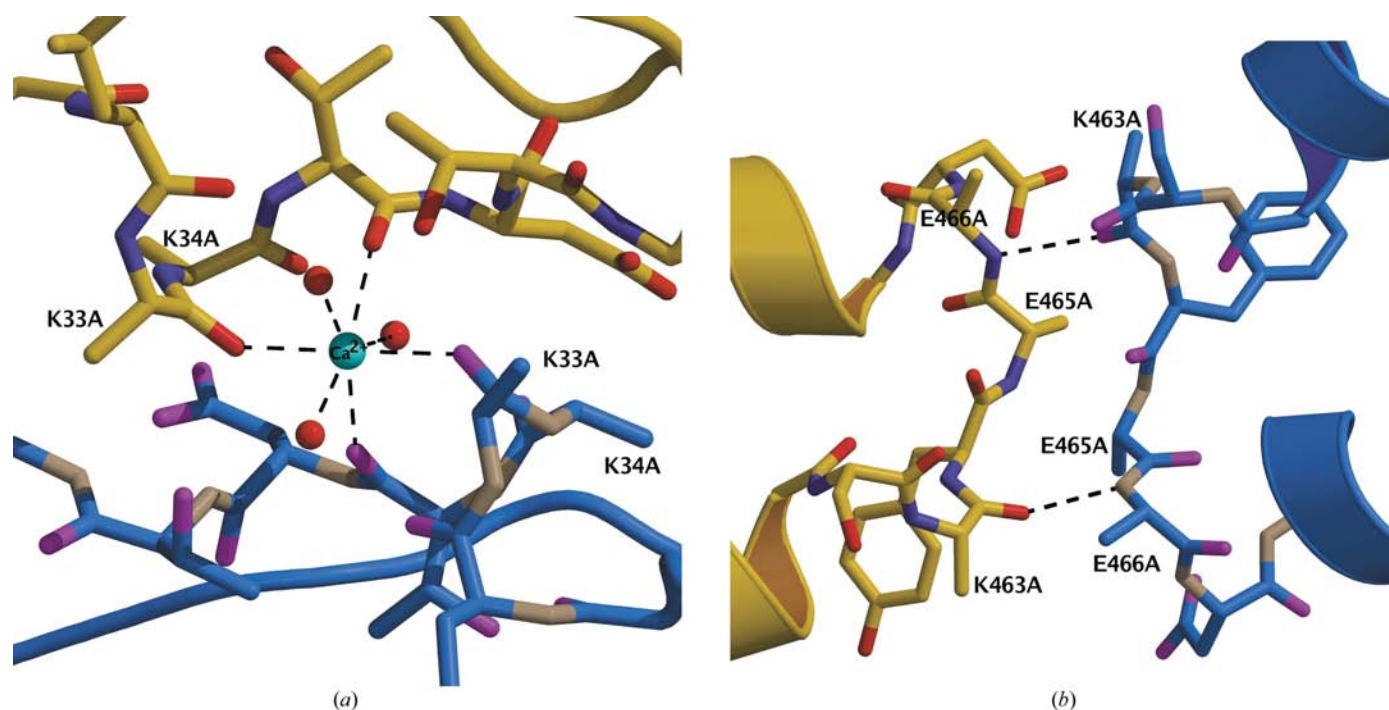

Figure 2
The homotypic crystal contacts mediated by mutated surface patches. (*a*) Short antiparallel β-structure formed by two mutated surface loops in adjacent RGSL molecules related by a crystallographic twofold axis. (*b*) A Ca²⁺-binding site formed by two adjacent mutated surface patches of the YkoF protein. Mutated residues are identified. For further details see text.
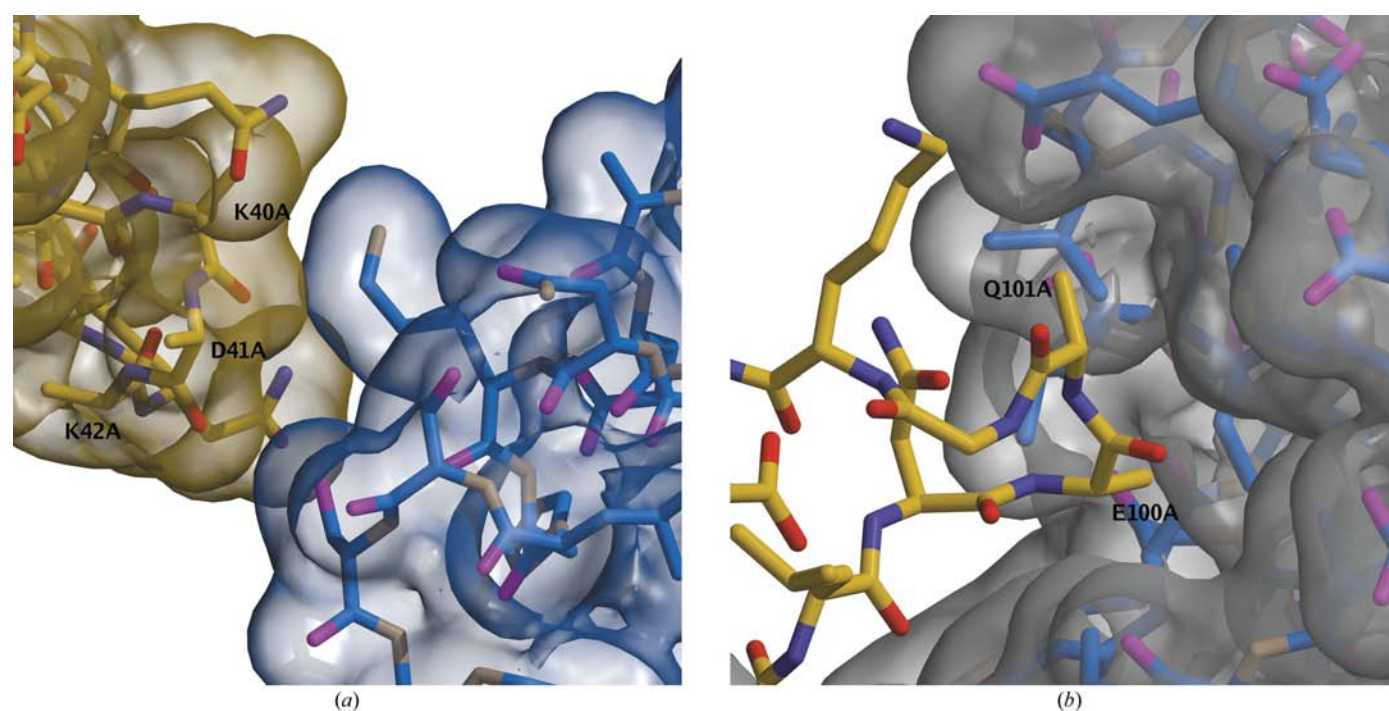

Figure 3
The heterotypic crystal contacts mediated by mutated surface patches. (*a*) Crystal contacts between adjacent molecules of LcrV in the *P*1 crystal form. (*b*) A non-crystallographic crystal contact between monomers of the Hsp33 chaperone. Mutated residues are labeled. For further details see text.

Table 1
Proteins with structures solved using crystals prepared with the use of the surface-entropy reduction strategy.

Protein	Reference, PDB code and resolution	Space group, unit-cell parameters ($\text{Å}, °$)	Mutants screened (the crystallized mutant is highlighted in bold)	Crystal contacts mediated by the mutated patch/ molecules per ASU
Y. pestis LcrV antigen	Derewenda *et al.* (2004), 1rf6, 2.2 Å	$P1$, $a = 35.9$, $b = 45.1$, $c = 46.9$, $\alpha = 76.2$, $\beta = 78.4$, $\gamma = 77.1$	K54A/D55A/E57A, K72A/K73A, E155A/E156A/E159A, K214A/E217A/K218A, **K40A/D41A/K42A**	Heterotypic/ crystallographic, 1
RGSL domain of human PDZRhoGEF	Longenecker *et al.* (2001), 1htj, 2.2 Å	$P6_122$, $a = b = 61.6$, $c = 201.9$	E90A/K91A, E123A/E126A, E131E/E134A, E171A/E172A, **K183A/E185A/E186A**	Homotypic/ crystallographic, 1
B. subtilis heat-shock protein Hsp33	Janda, Devedjiev, Cooper *et al.*, 2004, 1vzy, 2.0 Å	$P3_121$, $a = b = 115.5$, $c = 106.4$	**E100A/Q101A**	Heterotypic/ crystallographic, 2
B. subtilis peroxidase-resistance protein Ohr	Cooper *et al.* (in preparation), 2.1 Å	$P1$, $a = 35.3$, $b = 41.4$, $c = 44.6$, $\alpha = 84.4$, $\beta = 91.5$, $\gamma = 73.9$	E106A/K107A, **K34A/K35A/E36A**	Homotypic/ crystallographic, 2
B. subtilis YdeN gene product (serine hydrolase)	Janda, Devedjiev, Derewenda *et al.*, 2004, 1uxo, 1.8 Å	$P2_12_12_1$, $a = 36.2$, $b = 54.1$, $c = 93.2$	E124A/K127A, E167A/E169A, **K88A/Q89A**	Heterotypic/ crystallographic, 1
B. subtilis thiamin-binding protein YkoF	Devedjiev *et al.* (2004), 1s99, 1.6 Å	$P2_12_12_1$, $a = 60.9$, $b = 80.3$, $c = 85.9$	K112A/E114A, **K33A/K34A**	Homotypic/ noncrystallographic, 2
B. subtilis YkuD hydrolase	Bielnicki *et al.* (in preparation), 2 Å	$P2_12_12_1$, $a = 56.3$, $b = 63.9$, $c = 93.7$	**K117A/Q118A**	Homotypic/ crystallographic, 2
CUE–ubiquitin complex	Prag *et al.* (2003), 1p3q, 1.7 Å	$C2$, $a = 101.6$, $b = 45.9$, $c = 57.8$, $\beta = 96.5$	**K435A/K436A**	Homotypic/ crystallographic, 2
Tyrosine kinase domain of IGF-1	Munshi *et al.* (2003), 1p4o, 1.5 Å	$P2_1$, $a = 52.9$, $b = 85.6$, $c = 78.9$, $\beta = 99.1$	K1025A/K1026A, K1237A/E1238A, **E1067A/E1069A**	Homotypic/ crystallographic, 2

scopic structure of the protein's surface and predicts the nature of the potential patches allowing a relative reduction of the entropic barrier to crystallization.

Another important question is to what extent the success rate of crystallization is really being enhanced by surface engineering. To date, ~70% of targets attempted in one of our laboratories (ZSD) yielded X-ray quality crystals allowing complete structure determination to a resolution of 2.2 Å or better. We believe that this success rate is valid for target proteins that are soluble and stably folded. What is of particular importance is that this high success rate did not require extensive screening of large numbers of mutants. With the first two targets, several mutants were screened in parallel to maximize the chances of success. However, with subsequent targets one mutant was prepared at a time to minimize labor and cost. Out of those five proteins, only one required screening of three mutants, whereas four required only two or one mutant to crystallize (Ohr, YkoF, Hsp33, YkuD). The two other groups who have published structural results based on the crystal engineering approach also report that using few mutants, *i.e.* one or three, respectively, was enough to bring the projects to fruition. The question now is whether surface engineering is more potent then homolog screening. Assuming that the crystallization events for proteins derived from different species are statistically independent, the probability P that at least one will crystallize can be defined as

$$P = 1 - \prod_i (1 - P_i), \qquad (4)$$

where P_i is the overall probability of obtaining crystals from organism i. Assuming that the protein from the organism of choice did not crystallize, the chance of obtaining a crystal from another three organisms (with an average success of each estimated at 0.25) is still less than 0.60. Thus, our average of ~70% success with approximately 1.5 variants screened per project demonstrates that surface mutagenesis is significantly more effective than homolog screening.

What mutations should one use in surface engineering? The studies of RhoGDI as a model system indicated that mutating two or three closely spaced Lys or Glu residues to alanines constitutes the most productive strategy. In the actual case applications, the mutations were not limited exclusively to Lys or Glu, but instead sites containing a mix of these residues were used. If such sites were difficult to find, then Gln was also targeted provided that it was in the vicinity of Lys or Glu. In most cases the sites were chosen based either on known tertiary models or using secondary-structure prediction, so as to place the mutations on short and well ordered solvent-exposed loops. Although this strategy yielded very good results, it is possible that other types of mutations and other sites might also be useful. One of such alternatives, particularly attractive in the case of poorly soluble proteins, is the replacement of lysines by arginines. The results of a small feasibility study using RhoGDI yielded modest results, but did identify an interesting crystal form in which the crystal contacts were mediated by the arginine-rich patches with sulfate ions sequestered between them (Czepas *et al.*, 2004). More work is required to establish whether this approach has a more general utility.

The quality of crystals obtained by surface engineering is also of importance. The crystals of novel proteins obtained so far diffract to no worse than 2.2 Å resolution and the best crystals diffract to 1.5 Å (Table 1). Thus, the mean resolution appears to exceed the average in the Protein Data Bank by a significant margin. It should also be noted that in most projects the screening was stopped when the first diffracting crystal form was identified and no attempts were made to generate more mutants which might have produced even better diffracting crystals. In the case of the model system of RhoGDI, the best crystals obtained by surface engineering diffracted to the nearly atomic resolution of 1.3 Å (Mateja et al., 2002), while the wild-type crystals which are irreproducible diffract to no better than 2.5 Å (Keep et al., 1997). These data indicate that the nature of the crystal contacts engineered by surface engineering leads to high order and high diffraction quality. Two important consequences should be noted. Firstly, the method is useful to improve the precision of the atomic models of drug targets to facilitate detailed analysis of inhibitor–enzyme interactions. In fact, this was the rationale used by the Merck group in their study of the insulin-like growth hormone receptor kinase domain (Munshi et al., 2003). Secondly, the data obtained from crystals grown from surface-engineered variants are more likely to translate into the atomic models by automated methods. Any additional effort invested at the point of mutagenesis and protein production will lead to returns at the point of model building and refinement.

What is the nature of the crystal contacts in crystals of proteins modified for crystallization by surface engineering? Looking at the structures of the novel proteins, we note interesting patterns. In many cases, the mutated patches form homotypic, i.e. head-to-head, contacts creating symmetrical dimers, as is well represented by the structure of the RGSL domain of PDZRhoGEF (Longenecker, Lewis et al., 2001). The dimer of the engineered RGSL molecules is crystallographic (Fig. 2a), but this need not be the norm. For example, a similar homotypic association occurs in YkuD with the two molecules related by non-crystallographic symmetry (unpublished data). An interesting variation of homotypic interactions can be observed in YkoF (Devedjiev et al., 2004). This protein forms a tight homodimer in solution and the mutated patch is distal with respect to the monomer–monomer interface. The patches mediate non-crystallographic homotypic contacts, but interestingly these contacts are only possible owing to coordination of Ca^{2+} (Fig. 2b). The elimination of two Lys side chains exposed the main-chain carbonyl groups, but no unpaired amides are found within this loop, thus creating a negatively charged crevice ideal for divalent cation binding.

While homotypic crystallographic or noncrystallographic contacts seem to be most frequent, we also note a class of heterotypic interactions in which the mutated patch from one molecule fits like a joint in a socket of a surface crevice on an adjacent molecule (Fig. 3). At least in one case such an interaction is associated with lower symmetry, i.e. the P1 LcrV (Derewenda et al., 2004). It is also seen

in the Hsp33 structure, which is homodimeric in solution.

8. Conclusion

Rational surface engineering based on the concept of conformational entropy reduction is at its infancy as a tool for protein crystallization. However, the available examples demonstrate that this approach holds great promise and is likely to have a significant impact on the field. It can be used most effectively when a soluble and stable protein in its wild-type form remains recalcitrant to crystallization despite extensive screening. The method may also be a powerful method for the optimization of crystal quality, once it has been established that wild-type crystals diffract to medium or low resolution only. It is also expected that as the database of proteins crystallized by surface engineering expands, the results, combined with thermodynamic analyses of protein crystallization, will allow us to learn more about the crystallization process itself.

References

Avbelj, F. & Fele, L. (1998). J. Mol. Biol. 279, 665–684.
Ball, P. (2003). Nature (London), 423, 25–26.
Baud, F. & Karlin, S. (1999). Proc. Natl Acad. Sci. USA, 96, 12494–12499.
Bergeron, L., Filobelo, L. F., Galkin, O. & Vekilov, P. G. (2003). Biophys. J. 85, 3935–3942.
Bhattacharyya, S. M., Wang, Z.-G. & Zewail, A. H. (2003). J. Phys. Chem. B, 107, 13218–13228.
Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994). Nature (London), 371, 578–586.
Burley, S. K. & Bonanno, J. B. (2003). Methods Biochem. Anal. 44, 591–612.
Burton, W. K., Cabrera, N. & Frank, F. C. (1951). Philos. Trans. R. Soc. London Ser. A, 243, 299–360.
Campbell, J. W., Duee, E., Hodgson, G., Mercer, W. D., Stammers, D. K., Wendell, P. L., Muirhead, H. & Watson, H. C. (1972). Cold Spring Harbor Symp. Quant. Biol. 36, 165–170.
Chalikian, T. V., Plum, G. E., Sarvazyan, A. P. & Breslauer, K. J. (1994). Biochemistry, 33, 8629–8640.
Chalikian, T. V., Volker, J., Srinivasan, A. R., Olson, W. K. & Breslauer, K. J (1999). Biopolymers, 50, 459–471.
Chen, K. & Vekilov, P. G. (2002). Phys. Rev. E, 66, 021606.
Czepas, J., Devedjiev, Y., Krowarsch, D., Derewenda, U., Otlewski, J. & Derewenda, Z. S. (2004). Acta Cryst. D60, 275–280.
D'Arcy, A., Stihle, M., Kostrewa, D. & Dale, G. (1999). Acta Cryst. D55, 1623–1625.
Dasgupta, S., Iyer, G. H., Lawrence, C. E. & Bell, J. A. (1997). Proteins, 28, 494–514.
Derewenda, Z. S. (2004a). Structure, 12, 529–535.
Derewenda, Z. S. (2004b). Methods, 34, 354–363.
Derewenda, U., Mateja, A., Devedjiev, Y., Routzahn, K. M., Evdokimov, A. G., Derewenda, Z. S. & Waugh, D. S. (2004). Structure, 12, 301–306.

research papers

Devedjiev, Y., Surendranath, Y., Derewenda, U., Gabrys, A., Cooper, D. R., Zhang, R. G., Lezondra, L., Joachimiak, A. & Derewenda, Z. S. (2004). *J. Mol. Biol.* **343**, 395–406.

DoConte, L. L., Chothia, C. & Janin, J. (1999). *J. Mol. Biol.* **285**, 2177–2198.

Doye, J. P. K. (2004). *Phys. Biol.* **1**, P9–P13.

Dunitz, J. D. (1994). *Science*, **264**, 670.

Eisenberg, D. & Crothers, D. (1979). *Physical Chemistry with Applications to Life Sciences*. Menlo Park: Benjamin/Cummins.

Eisenberg, D. & Kauzmann, W. (1969). *The Structure and Properties of Water*. Oxford University Press.

Fersht, A. (1999). *Structure and Mechanism in Protein Science*. New York: W. H. Freeman.

Finkelstein, A. V. & Janin, J. (1989). *Protein Eng.* **3**, 1–3.

Garrard, S. M., Longenecker, K. L., Lewis, M. E., Sheffield, P. J. & Derewenda, Z. S. (2001). *Protein Expr. Purif.* **21**, 412–416.

Giesen, M., IckingKonert, G. S., Stapel, D. & Ibach, H (1996). *Surf. Sci.* **366**, 229–238.

Gliko, O., Neumaier, N., Fischer, M., Haase, I., Bacher, A., Weinkauf, S. & Vekilov, P. G. (2005). *J. Cryst. Growth*, **275**, e1409–e1416.

Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. & Riggs, A. D. (1979). *Proc. Natl Acad. Sci. USA*, **76**, 106–110.

Horwich, A. (2000). *Nature Struct. Biol.* **7**, 269–270.

Hui, R. & Edwards, A. (2003). *J. Struct. Biol.* **142**, 154–161.

Israelachvili, J. N. (1995). *Intermolecular and Surface Forces*. New York: Academic Press.

Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. & Boyer, H. W. (1997). *Science*, **198**, 1056–1063.

Janda, I., Devedjiev, Y., Cooper, D., Chruszcz, M., Derewenda, U., Gabrys, A., Minor, W., Joachimiak, A. & Derewenda, Z. S. (2004). *Acta Cryst.* D**60**, 1101–1107.

Janda, I., Devedjiev, Y., Derewenda, U., Dauter, Z., Bielnick, J., Cooper, D., Graf, P. C., Joachimiak, A., Jakob, U. & Derewenda, Z. S. (2004). *Structure*, **12**, 1901–1907.

Keep, N. H., Barnes, M., Barsukov, I., Badii, R., Lian, L. Y., Segal, A. W., Moody, P. C. & Roberts, G. C. (1997). *Structure*, **5**, 623–633.

Kendrew, J. C., Parrish, R. G., Marrack, J. R. & Orlans, E. S. (1954). *Nature (London)*, **174**, 946–949.

Kuipers, L., Hoogeman, M. & Frenken, J. (1993). *Phys. Rev. Lett.* **71**, 3517–3520.

Kuznetsov, Y. G., Malkin, A. J., Lucas, R. W. & McPherson, A. (2000). *Colloids Surf. B Biointerfaces*, **19**, 333–346.

Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M., Livingstone, J. C., Treffry, A., Luzzago, A., Levi, S., Arosio, P. & Cesareni, G. (1991). *Nature (London)*, **349**, 541–544.

Leckband, D. & Israelachvili, J. (2001). *Quart. Rev. Biophys.* **34**, 105–267.

Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. (2001). *Acta Cryst.* D**57**, 679–688.

Longenecker, K. L., Lewis, M. E., Chikumi, H., Gutkind, J. S. & Derewenda, Z. S. (2001). *Structure*, **9**, 559–569.

McElroy, H. H, Sisson, G. W., Schottlin, W. E., Aust, R. M. & Villafranca, J. E. (1992). *J. Cryst. Growth*, **122**, 265–272.

Madhusudan, Kodandapani, R. & Vijayan, M. (1993). *Acta Cryst.* D**49**, 234–245.

Mateja, A., Devedjiev, Y., Krowarsch, D., Longenecker, K., Dauter, Z., Otlewski, J. & Derewenda, Z. S. (2002). *Acta Cryst.* D**58**, 1983–1991.

Munshi, S., Munshi, S., Hall, D. L., Kornienko, M., Darke, P. L. & Kuo, L. C. (2003). *Acta Cryst.* D**59**, 1725–1730.

Nilles, M. L. (2004). *Structure*, **12**, 357–358.

Pal, S. K. & Zewail, A. H. (2004). *Chem. Rev.* **104**, 2099–2124.

Petsev, D. N., Chen, K., Gliko, O. & Vekilov, P. G. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 792–796.

Petsev, D. N., Thomas, B. R., Yau, S. & Vekilov, P. G. (2000). *Biophys. J.* **78**, 2060–2069.

Petsev, D. N., Thomas, B. R., Yau, S.-T., Tsekova, D., Nanev, C. N., Wilson, W. W. & Vekilov, P. G. (2001). *J. Cryst. Growth*, **232**, 21–29.

Petsev, D. N., Wu, X., Galkin, O. & Vekilov, P. G. (2003). *J. Phys. Chem. B*, **107**, 3921–3926.

Prag, G., Misra, S., Jones, E. A., Ghirlando, R., Davies, B. A., Horazdovsky, B. F. & Hurley, J. H. (2003). *Cell*, **113**, 609–620.

Reviakine, I., Georgiou, D. K. & Vekilov, P. G. (2003). *J. Am. Chem. Soc.* **125**, 11684–11693.

Schall, C., Arnold, E. & Wiencek, J. M. (1996). *J. Cryst. Growth*, **165**, 293–302.

Schwede, T. F., Badeker, M., Langer, M., Retey, J. & Schulz, G. E. (1999). *Protein Eng.* **12**, 151–153.

Swartzentruber, B. S., Mo, Y., Kariotis, R., Lagally, M. G. & Webb, M. B. (1990). *Phys. Rev. Lett.* **65**, 1913–1916.

Tanford, C. (1980). *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*. New York: John Wiley & Sons.

Tidor, B. & Karplus, M. (1994). *J. Mol. Biol.* **238**, 405–414.

Vekilov, P. G. (2003). *Methods Enzymol.* **368**, 84–105.

Vekilov, P. G. (2004). *Nanoscale Structure and Assembly at Solid–Fluid Interfaces*, Vol. II, *Assembly in Hybrid and Biological Systems*, edited by X. Y. Lui, pp. 145–200. Dordrecht: Kluewer Academic Publishers.

Vekilov, P. G. & Chernov, A. A. (2002). *Solid State Physics*, edited by F. Spaepen, pp. 1–147. New York: Academic Press.

Vekilov, P. G., Feeling-Taylor, A. R., Petsev, D. N., Galkin, O., Nagel, R. L. & Hirsch, R. E. (2002). *Biophys. J.* **83**, 1147–1156.

Vekilov, P. G., Feeling-Taylor, A. R., Yau, S. T. & Petsev, D. (2002). *Acta Cryst.* D**58**, 1611–1616.

Wiener, M. C. (2004). *Methods*, **34**, 364–372.

Yau, S.-T., Petsev, D. N., Thomas, B. R. & Vekilov, P. G. (2000). *J. Mol. Biol.* **303**, 667–678.

Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A. M. & Arrowsmith, C. H. (2003). *Acc. Chem. Res.* **36**, 183–189.

Yip, C. M. & Ward, M. D. (1996). *Biophys. J.* **71**, 1071–1078.

Zhang, F., Basinski, M. B., Beals, J. M., Briggs, S. L., Churgay, L. M., Clawson, D. K., DiMarchi, R. D., Furman, T. C., Hale, J. E., Hsiung, H. M., Schoner, B. E., Smith, D. P., Zhang, X. Y., Wery, J. P. & Schevitz, R. W. (1997). *Nature (London)*, **387**, 206–209.