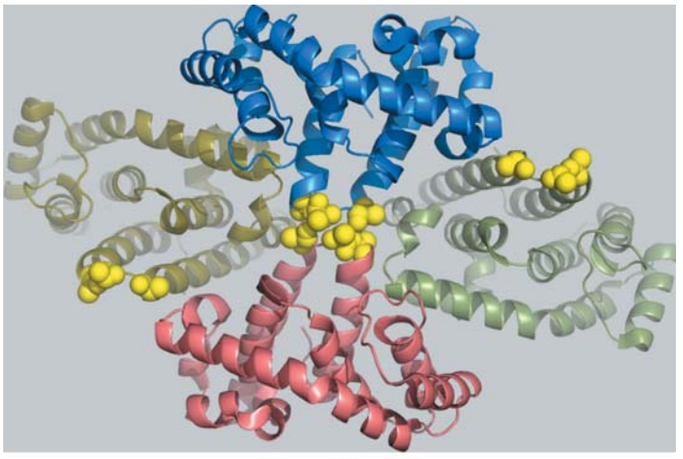
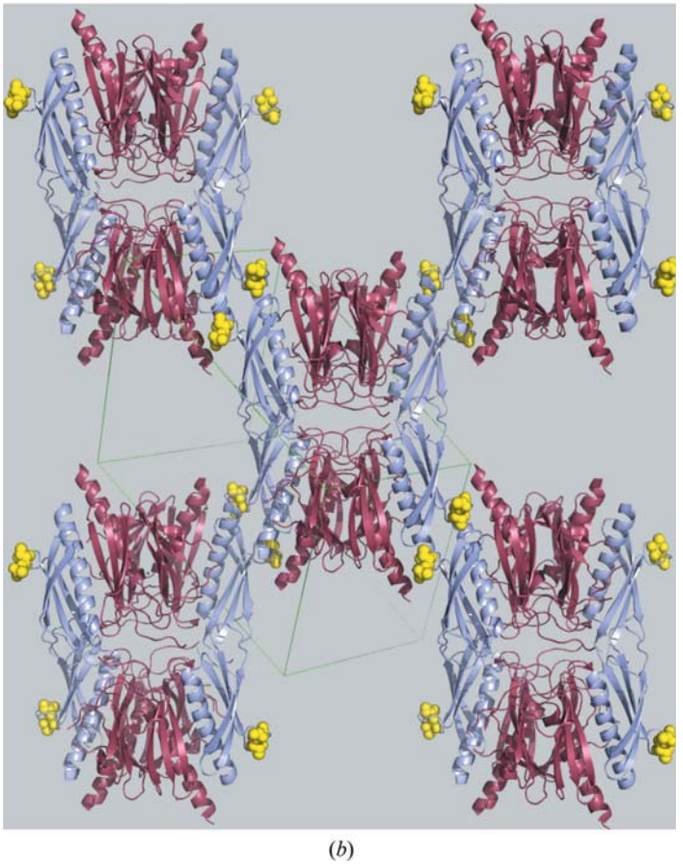


(*a*)



(*b*)

Figure 4.3.9.1
Two examples of proteins crystallized by the surface-entropy reduction (SER) method. (*a*) The RGSL domain of the PDZRhoGEF nucleotide-exchange factor (PDB code 1htj; Longenecker, Lewis *et al.*, 2001); the yellow spheres show the alanines introduced by mutagenesis, which mediate an isologous crystal contact across a crystallographic twofold axis. (*b*) The crystal structure of EpsI complexed with EpsJ (PDB code 2ret; Yanez *et al.*, 2008); the EpsI protein (pale blue) contains two surface mutations, shown by yellow spheres, which mediate heterologous crystal contacts.

protein only crystallized after an entropy-reducing mutation (E104Y) was introduced into the protein (Pornillos *et al.*, 2009). The protein–protein complexes solved to date underscore the utility of the method, which extends beyond individual proteins because high-entropy patches occur outside complex interfaces. For example, the complex of c-Src and its inactivator Csk was crystallized using a variant of Csk carrying K361A and K362A mutations (Levinson *et al.*, 2008). Similarly, it was possible to crystallize the complex of two pseudopilins EpsI and EpsJ from the type 2 secretion system of *Vibrio vulnificus* when a variant of EpsI carrying two mutations (E128T and K129T) was used (Yanez *et al.*, 2008; Fig. 4.3.9.1).

An interesting variation of the SER method was used in the investigation of the RACK1 protein, which was crystallized as an in-line fusion with an MBP variant carrying D82A, K83A and K239A mutations (Ullah *et al.*, 2008). This is the first example of the application of the surface-entropy reduction strategy to a carrier protein and not the crystallization target itself.

The SER strategy is attractive not only because of its efficacy but also because of its simplicity: once an expression construct for a target protein is available several rounds of mutagenesis can easily create variants with systematically enhanced crystallizability. To assist in the design of crystallizable variants, a server has been developed that uses the amino-acid sequence of the target to identify suitable mutation sites (Goldschmidt *et al.*, 2007).

4.3.10. Improvement of crystal quality

In most cases, protein engineering is used as a tool of last resort to obtain variants for proteins for which no crystals can be grown using the wild-type form. However, it may sometimes be necessary to obtain a new, different crystal form even when the wild-type protein does crystallize. Such a need may arise, for example, in drug-design investigations, where high-resolution structures are particularly critical for evaluation of the interactions between lead compounds and the target protein and may not always be possible using wild-type crystals. A novel crystal form may also be necessary if the wild-type crystals contain the target protein in an orientation in which the active site is obscured by crystal contacts, making it impossible to soak in drug lead compounds and screen small-molecule libraries by high-throughput crystallography (Blundell & Patel, 2004).

One possible strategy for obtaining a new crystal form is to modify the existing crystal contacts by replacing some of the participating amino acids. While this approach occasionally leads to improvement of the X-ray data resolution (Liu *et al.*, 2007; Mizutani *et al.*, 2008), modification of crystal contacts is typically counterproductive as it abolishes the propensity of the target to crystallize in one form but does not necessarily induce another (Charron *et al.*, 2002). A more successful strategy is to generate a novel crystal form by engineering new crystal contacts through SER. For example, a novel crystal form of the insulin-like growth factor 1 receptor kinase domain, a putative drug target, was obtained using a double mutant (E1067A and E1069A); the new form diffracted to 1.5 Å resolution, whereas the wild-type crystals only diffracted to 2.7 Å resolution (Munshi *et al.*, 2003). In the case of the catalytic domain of activated factor XI, a key enzyme in the blood coagulation cascade and another potential drug target, a single K437A mutation allowed the preparation of a crystal form that diffracted to 2.0 Å resolution (Jin *et al.*, 2005). Entropy-reducing mutations were also key in the preparation of a crystal form of HIV-1 reverse transcriptase for structure-based drug design that diffracted to 1.8 Å resolution, in contrast to the typical 2.5–3.0 Å range observed for the wild-type protein crystals (Bauman *et al.*, 2008).

4.3.11. Conclusions

Protein engineering has become a routine tool that is used to generate crystallizable macromolecules and their complexes. While some approaches may only apply to very specific targets, a

number of strategies offer general applicability. Among these, gene-construct optimization or surface-entropy reduction are quickly gaining popularity as methods of choice. However, it should be stressed that none of the methods described here offer a guarantee that the target protein can be coerced to crystallize. To maximize the chances of success, one must frequently attack the problem on multiple fronts based on an understanding of the chemical and physical properties of a specific protein. This is particularly true of technically difficult targets such as membrane proteins. A classic example illustrating this principle is the study of the HIV gp120 envelope glycoprotein (Kwong *et al.*, 1998, 1999). The construct that was ultimately used in successful crystallization screens had deletions of 52 and 19 residues from the N- and C-termini and two flexible loops replaced by Gly-Ala-Gly linkages; additionally, the protein was 90% deglycosylated compared with the wild type. Moreover, this engineered gp120 was only crystallized in the form of a ternary complex with the CD4 receptor and an Fab fragment from a neutralizing antibody. In the recent case of the ATP-gated P2X4 ion channel, a crystallizable variant was obtained after a series of N- and C-terminal deletions were screened to identify the smallest functional unit and the introduction of three mutations (C51F/N78K/N187R) to eliminate both aggregation arising from oxidation and N-glycosylation (Kawate *et al.*, 2009).

The rapidly expanding database of macromolecular structures greatly enhances our understanding of the physical chemistry of proteins, ultimately enhancing our ability to predict the behaviour of a protein in solution from its sequence. It is therefore increasingly possible to rely on such theoretical predictions *in lieu* of tedious experimental screens. A number of online tools have been developed for this purpose. The propensity of a protein target to crystallize can be evaluated using the *XtalPred* server (http://ffas.burnham.org/XtalPred-cgi/xtal.pl), which offers insights into potential sources of problems arising from sequence features (Slabinski *et al.*, 2007). Automated design of optimally truncated constructs for structural analysis has been made possible by the *ProteinCCD* meta-server (http://xtal.nki.nl/ccd), which uses the cDNA sequence of the target (Mooij *et al.*, 2009). This server collects information about secondary structure, disorder, putative coiled coils, transmembrane segments, domains and domain linkers, and suggests constructs so that the user can interactively choose suitable options and obtain sequences of oligonucleotides needed for appropriate PCR amplification (Mooij *et al.*, 2009). For proteins recalcitrant to crystallization in their wild-type form, surface mutations enhancing crystallizability can be designed using the surface-entropy reduction server (http://nihserver.mbi.ucla.edu/SER/; Goldschmidt *et al.*, 2007).

As the focus of macromolecular crystallography shifts from the principles of protein architecture to increasingly complex biological questions, the approach to crystallization is also undergoing dramatic evolution. As we gain better understanding of the microscopic nature of protein crystallization, we will be able to develop rational protein-engineering strategies that systematically and significantly improve the success rate of crystallization.

The research on protein crystallization in my laboratory is supported by the National Institute for General Medical Sciences. I am greatly indebted to Drs Urszula Derewenda, Anthony Kossiakoff, Alexander McPherson, Geoff Waldo and David Cooper for discussions and comments on the early version of the manuscript.

This chapter is based on a recent article published in *Acta Crystallographica Section D* (Derewenda, 2010).

References

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. & Iwata, S. (2003). *Structure and mechanism of the lactose permease of Escherichia coli*. *Science*, **301**, 610–615.

Al-Ayyoubi, M., Gettins, P. G. & Volz, K. (2004). *Crystal structure of human maspin, a serpin with antitumor properties: reactive center loop of maspin is exposed but constrained*. *J. Biol. Chem.* **279**, 55540–55544.

Aleshin, A. E., Hoffman, C., Firsov, L. M. & Honzatko, R. B. (1994). *Refined crystal structures of glucoamylase from Aspergillus awamori var. X100*. *J. Mol. Biol.* **238**, 575–591.

Bandeiras, T. M. *et al.* (2008). *Structure of wild-type Plk-1 kinase domain in complex with a selective DARPin*. *Acta Cryst.* D**64**, 339–353.

Bauman, J. D., Das, K., Ho, W. C., Baweja, M., Himmel, D. M., Clark, A. D. Jr, Oren, D. A., Boyer, P. L., Hughes, S. H., Shatkin, A. J. & Arnold, E. (2008). *Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design*. *Nucleic Acids Res.* **36**, 5083–5092.

Berman, H., Henrick, K., Nakamura, H. & Markley, J. L. (2007). *The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data*. *Nucleic Acids Res.* **35**, D301–D303.

Birtley, J. R. & Curry, S. (2005). *Crystallization of foot-and-mouth disease virus 3C protease: surface mutagenesis and a novel crystal-optimization strategy*. *Acta Cryst.* D**61**, 646–650.

Blundell, T. L. & Patel, S. (2004). *High-throughput X-ray crystallography for drug discovery*. *Curr. Opin. Pharmacol.* **4**, 490–496.

Brewer, S. J., Haymore, B. L., Hopp, T. P. & Sassenfeld, H. M. (1991). *Engineering proteins to enable their isolation in a biologically active form*. *Bioprocess Technol.* **12**, 239–266.

Brooun, A., Foster, S. A., Chrencik, J. E., Chien, E. Y., Kolatkar, A. R., Streiff, M., Ramage, P., Widmer, H., Weckbecker, G. & Kuhn, P. (2007). *Remedial strategies in structural proteomics: expression, purification, and crystallization of the Vav1/Rac1 complex*. *Protein Expr. Purif.* **53**, 51–62.

Cabantous, S., Pedelacq, J. D., Mark, B. L., Naranjo, C., Terwilliger, T. C. & Waldo, G. S. (2005). *Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from Mycobacterium tuberculosis*. *J. Struct. Funct. Genomics*, **6**, 113–119.

Cabantous, S. & Waldo, G. S. (2006). *In vivo and in vitro protein solubility assays using split GFP*. *Nat. Methods*, **3**, 845–854.

Campbell, J. W., Duée, E., Hodgson, G., Mercer, W. D., Stammers, D. K., Wendell, P. L., Muirhead, H. & Watson, H. C. (1972). *X-ray diffraction studies on enzymes in the glycolytic pathway*. *Cold Spring Harb. Symp. Quant. Biol.* **36**, 165–170.

Canaves, J. M., Page, R., Wilson, I. A. & Stevens, R. C. (2004). *Protein biophysical properties that correlate with crystallization success in Thermotoga maritima: maximum clustering strategy for structural genomics*. *J. Mol. Biol.* **344**, 977–991.

Carter, C. W. Jr & Carter, C. W. (1979). *Protein crystallization using incomplete factorial experiments*. *J. Biol. Chem.* **254**, 12219–12223.

Carugo, O. & Argos, P. (1997). *Protein–protein crystal-packing contacts*. *Protein Sci.* **6**, 2261–2263.

Center, R. J., Kobe, B., Wilson, K. A., Teh, T., Howlett, G. J., Kemp, B. E. & Poumbourios, P. (1998). *Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein*. *Protein Sci.* **7**, 1612–1619.

Charron, C., Kern, D. & Giegé, R. (2002). *Crystal contacts engineering of aspartyl-tRNA synthetase from Thermus thermophilus: effects on crystallizability*. *Acta Cryst.* D**58**, 1729–1733.

Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). *High-resolution crystal structure of an engineered human beta(2)-adrenergic G protein-coupled receptor*. *Science*, **318**, 1258–1265.

Cieślik, M. & Derewenda, Z. S. (2009). *The role of entropy and polarity in intermolecular contacts in protein crystals*. *Acta Cryst.* D**65**, 500–509.

Cohen, S. L., Ferré-D'Amaré, A. R., Burley, S. K. & Chait, B. T. (1995). *Probing the solution structure of the DNA-binding protein max by a combination of proteolysis and mass-spectrometry*. *Protein Sci.* **4**, 1088–1099.

Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M. & Derewenda, Z. (2007). *Protein crystallization by*