Chapter 4.3. Application of protein engineering to enhance crystallizability and improve crystal properties

Z. S. Derewenda

4.3.1. Introduction

Following the dawn of recombinant technology brought about by the groundbreaking overexpression of synthetic genes coding for insulin and somatostatin in *Escherichia coli* (Goeddel *et al.*, 1979; Itakura *et al.*, 1977) and the subsequent discovery of the polymerase chain reaction (PCR; Mullis *et al.*, 1986; Saiki *et al.*, 1985, 1988), macromolecular crystallography was freed of its long-standing dependence on purified native protein samples for crystallization. Heterologous expression made it possible to generate samples of proteins and complexes that are found in only small or trace amounts in living cells and to engineer large and unstable proteins so that isolated domains or modified forms can be made available for crystallization. At the same time, the effort required for protein purification was dramatically reduced by the use of fusion proteins and affinity tags (Brewer *et al.*, 1991; Sassenfeld, 1990; Malhotra, 2009). As a consequence, the overwhelming majority of samples used today for crystallization are recombinantly derived proteins. However, even though material for crystallization is more easily available, the preparation of single, well diffracting crystals of the target macromolecule is still a time-consuming challenge.

Historically, two complementary approaches to protein crystallization were developed in parallel. Firstly, natural variations in the amino-acid sequences of homologues from different species were exploited to identify a target with suitable crystallization properties during the purification procedure (Kendrew *et al.*, 1954; Campbell *et al.*, 1972). The second approach was to extensively screen a specific target protein against a range of diverse precipitating agents, buffers and additives until the right conditions for crystallization were identified (Carter & Carter, 1979; Jancarik & Kim, 1991). These strategies remain the pillars of contemporary macromolecular crystallization. However, as the palette of molecular biology techniques expanded to include site-directed mutagenesis, ligation-independent cloning and other tools, it became possible to modify proteins with relative ease with the specific purpose of enhancing their propensity to crystallize or improving the diffraction quality of the resulting crystals. The early proof-of-principle of these capabilities was the crystallization of an engineered variant of human H-ferritin in which a single-site mutation, K86Q, was introduced to duplicate a crystal contact mediated by Cd^{2+} ions in the crystal structure of the homologous rat L-ferritin (Lawson *et al.*, 1991).

In this review, current progress in the methodologies of protein engineering used to enhance the crystallizability of targets that are recalcitrant to crystallization in their wild-type form is discussed. This burgeoning field is very broad and includes both general strategies that apply to a range of targets and many diverse approaches that only apply to specific proteins or protein families. Thus, owing to space limitations, the focus is on those techniques that have either been demonstrated to be of general utility or are at a point in their development to clearly have the potential to become widely used in the future. Understandably, only representative examples are provided.

4.3.2. Microscopic aspects of protein crystallization

Proteins are inherently dynamic entities, a property that greatly hinders their crystallizability. Not surprisingly, it has been estimated that even for the stable and relatively small single-domain prokaryotic proteins fewer than one in four will yield X-ray-quality crystals when using a routine screening process (Canaves *et al.*, 2004; Price *et al.*, 2009). In order to rationally modify proteins to enhance their crystallizability, it is first necessary to understand the physical properties that make most proteins resistant to crystallization.

Protein crystals are nucleated *ab initio* at supersaturation levels in the 200–1000% range (McPherson, 1982). Nucleation is believed to proceed *via* a two-step process: clusters of solute molecules form first and upon reaching critical size reorganize into three-dimensionally ordered nuclei (Georgalis *et al.*, 1997; Vekilov, 2004; Erdemir *et al.*, 2009). Subsequent transfer of protein molecules from solution onto the growing crystal surface is driven by relatively small negative changes in Gibbs free energy (ΔG°), from approximately -10 to -100 kJ mol^{-1}, at ambient temperature (Vekilov, 2003). Interestingly, enthalpy changes are generally negligible during crystallization (Yau *et al.*, 2000; Petsev *et al.*, 2001; Gliko *et al.*, 2005), so that entropic phenomena dominate (Vekilov *et al.*, 2002; Vekilov, 2003; Derewenda & Vekilov, 2006). The microscopic effects underlying the entropy changes, both favourable and unfavourable, involve the protein itself as well as the solvent. Protein packing, which results in an ordered three-dimensional lattice and loss of translational and rotational degrees of freedom, is unfavourable and produces an energy barrier in the 30–100 kJ mol^{-1} range at room temperature (Finkelstein & Janin, 1989; Tidor & Karplus, 1994). Similarly, incorporation into the growing crystal and ordering of any intrinsically unstructured elements, such as flexible termini or loops and side chains, at the point of crystal contacts further increases the entropic cost. However, the release of ordered solvent molecules from the surfaces involved in crystal contacts, which is estimated to be in the 25–150 kJ mol^{-1} range, may sufficiently compensate for these entropy losses and ultimately provide the driving force for crystal growth (Vekilov *et al.*, 2002; Vekilov, 2003).

Based on these considerations, it is evident that a protein must satisfy certain criteria in order to crystallize. Firstly, it must have a molecular surface that confers adequate solubility under initial conditions to reach the necessary supersaturation level for nucleation. Furthermore, it should have few, if any, intrinsically unstructured fragments such as extended N- or C-termini or long and solvent-exposed loops which may impede crystallization. Finally, the protein should have distinct 'sticky' patches on the surface with a structured layer of solvent molecules, allowing the ordering of nascent nuclei by mediating thermodynamically viable specific crystal contacts.

The notion that protein crystallization involves specific and anisotropic intermolecular interactions, as opposed to random contacts, is relatively new. Early analyses of intermolecular contacts in protein crystals concluded that crystallization is a

references