A synergistic approach to protein crystallization: Combination of a fixed-arm carrier with surface entropy reduction

Andrea F. Moon, Geoffrey A. Mueller, Xuejun Zhong, and Lars C. Pedersen*

Laboratory of Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

Received 15 January 2010; Revised 5 February 2010; Accepted 8 February 2010
DOI: 10.1002/pro.368
Published online 1 March 2010 proteinscience.org

Abstract: Protein crystallographers are often confronted with recalcitrant proteins not readily crystallizable, or which crystallize in problematic forms. A variety of techniques have been used to surmount such obstacles: crystallization using carrier proteins or antibody complexes, chemical modification, surface entropy reduction, proteolytic digestion, and additive screening. Here we present a synergistic approach for successful crystallization of proteins that do not form diffraction quality crystals using conventional methods. This approach combines favorable aspects of carrier-driven crystallization with surface entropy reduction. We have generated a series of maltose binding protein (MBP) fusion constructs containing different surface mutations designed to reduce surface entropy and encourage crystal lattice formation. The MBP advantageous increases protein expression and solubility, and provides a streamlined purification protocol. Using this technique, we have successfully solved the structures of three unrelated proteins that were previously unattainable. This crystallization technique represents a valuable rescue strategy for protein structure solution when conventional methods fail.

Keywords: protein crystallography; surface entropy reduction; carrier-driven crystallization; MBP; fixed-arm; rescue strategy

Introduction
Solving the structure of a protein using X-ray crystallography is accomplished via a multi-stage process which requires successful completion of each stage before proceeding to the next. One of the largest obstacles to obtaining crystal structures is the need for milligram quantities of highly purified, soluble protein. Once soluble protein is obtained and purified, it must be crystallized in a form that yields diffraction quality crystals, which can also represent a serious hurdle. According to recent statistics from the Joint Center for Structural Genomics (http://www.jcsg.org), only 9.8% of expressed proteins generate crystals of sufficient size and quality for X-ray screening. Of these crystals, 74% demonstrate adequate diffraction for data collection. Despite having diffraction-quality crystals, only 50% of these are solved.
Historically, when faced with a problem along the crystallography pipeline, many crystallographers have taken advantage of specific rescue strategies. Several methods of chemical modification have been developed, including reductive lysine methylation and carboxymethylation. In addition, crystallization of orthologous and/or isomeric proteins is also a respectable alternative. An in situ proteolysis method utilizing batch protease digestion has also been established. However, if these strategies are insufficient to produce diffraction quality crystals, what other options are available? In recent years, several different rescue techniques for managing intractable proteins have been utilized.

One such approach involves crystallization of the target protein in complex with its endogenous binding partners. However, expression and purification of such partners are not always possible, nor do all target proteins have known binding partners. An alternative to crystallization of endogenous multiprotein assemblies is complexation to an antibody. Many proteins (membrane proteins and viral capsid proteins in particular) have been crystallized in complex with high affinity antibody fragments (Supporting Information Table 1). Antibody-mediated crystallization is thought to increase the likelihood of crystallization by providing a large, hydrophilic interaction surface for initiating crystal lattice contacts and by effectively limiting the conformational flexibility of solvent exposed loop regions. There are some drawbacks that also must be considered. Generating the antibody fragments has become a routine process, but is still costly, time-consuming and labor-intensive. To create an antibody complex suitable for crystallization, large quantities of soluble protein must be generated for both the antibody fragment and the target protein, which can be problematic in many cases. The antibodies must display high affinity binding to the native conformation of the target protein, and must be soluble/stable under the same conditions. A major consideration for this technique is that soluble, stable, and specific antibodies must be produced individually for each target protein to be studied.

Another rescue strategy uses crystallization of a large carrier protein fused to the protein of interest. Many different carrier fusion proteins have been used in this manner (Table I), including maltose binding protein (MBP), glutathione-S-transferase (GST), thioredoxin (TRX), and lysozyme. In many cases, the presence of the carrier increases the expression level and solubility of the protein of interest. Positioning the carrier on the N-terminus of the polypeptide can have a “chaperone-like” effect, aiding in protein folding and increasing the yield of active product. Many such carrier proteins are affinity tags, which also streamlines the purification protocol, making this method highly adaptable for high-throughput screening.

From a crystallographic standpoint, addition of the carrier protein to the crystallizable unit can be both advantageous and detrimental. Because the structures of the carrier proteins used in this process have been solved to high resolution, they can be used as search models to solve the phase problem via molecular replacement. In addition, the carrier proteins used in these studies often crystallize readily, which suggests that they can provide molecular surfaces that are conducive for crystal lattice formation, similar to the addition of an antibody. However, the unavoidable linker region connecting the carrier to the protein of interest becomes a possible source for structural heterogeneity, a less pertinent issue in antibody-mediated crystallization. In the hopes of reducing conformational flexibility, minimizing the length of the linker region appears to be of paramount importance.

In some cases, the increased expression and solubility of a carrier protein alone is not sufficient to produce diffraction quality crystals. There are several reported instances where expression of a recalcitrant target protein fused downstream of MBP has produced a high yield of soluble, active protein in adequate quantities for crystallization, yet diffraction quality crystals were not obtained.

Another rescue strategy now used with problematic systems is the powerful surface entropy reduction (SER) technique. Protein crystallization is an entropy-driven process—a delicate thermodynamic interplay between the ordering of protein molecules within a crystal lattice and the release of water molecules ordinarily bound to the surface of those protein molecules. SER attempts to compensate for the entropic behavior of large hydrophilic side chains on the protein surface. Random motions by these side chains can subtly increase the entropy of the system to the point that forming positive lattice contacts is inhibited. Therefore, by mutating patches of surface-exposed large, charged residues to small nonpolar amino acids, the surface entropy is reduced, allowing formation of crystal contacts featuring backbone amide and carbonyl groups.

Though the advantages of SER lie in crystal lattice formation, there are some drawbacks to this technique. Altering the electrostatic properties of a protein can alter its biochemical behavior, and its interaction with other protein binding partners. If the structure of the protein target is unknown, determining the appropriate amino acid sequence to mutate—without disrupting the protein structure or physiology—can be difficult. In addition, replacing solvent-exposed hydrophilic residues with nonpolar amino acids can decrease the solubility and stability.
Therefore, relatively unstable or less soluble proteins are not optimal candidates for SER.

To increase the probability of obtaining diffraction-quality crystals for problematic proteins, we have devised a flexible technique that harnesses the advantages of carrier-driven crystallization and surface entropy reduction. This approach works synergistically to increase the likelihood of crystallization. We have selected MBP for use in this system due to its high solubility, published success with this affinity tag, and its streamlined purification protocol. MBP is fused to the N-terminus of the target protein. The linker joining the two proteins has been truncated as published in Center et al. (1998) to create as little conformational heterogeneity as possible. To expand the capabilities of this system and improve the probability of crystallization, we have generated entropy reducing mutations on solvent exposed loops on the surface of the MBP based on the known crystal structures. These mutations maximize the potential for crystallization of the resulting fusion protein. Five easily interchangeable “cassettes,” carrying different combinations of the MBP/SER mutations, were created for use in this system (Table II).

By producing a set of vectors with different SER mutations on the MBP rather than on the target protein, we negate the need to create SER mutations on each target, providing for a straightforward and efficient cloning protocol requiring minimal cost and effort. Using the SER technique in conjunction with the MBP carrier protein also addresses the issue of decreased protein solubility and stability. This

<table>
<thead>
<tr>
<th>Carrier Protein</th>
<th>Target Protein</th>
<th>Size (AA)</th>
<th>Resolution (Å)</th>
<th>Linker sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBP (~366AA)</td>
<td>Hepatitis B surface antigen</td>
<td>14</td>
<td>2.7-2.9</td>
<td>GS</td>
</tr>
<tr>
<td></td>
<td>IAPP</td>
<td>22</td>
<td>1.75-1.86</td>
<td>AAA</td>
</tr>
<tr>
<td></td>
<td>MutS (C-term)</td>
<td>34</td>
<td>2.0</td>
<td>HM</td>
</tr>
<tr>
<td></td>
<td>MATa</td>
<td>50</td>
<td>2.1-2.3</td>
<td>AAAA</td>
</tr>
<tr>
<td></td>
<td>Huntingtin (N-term)</td>
<td>66</td>
<td>3.5-3.7</td>
<td>AAA</td>
</tr>
<tr>
<td></td>
<td>TIM40/MIA46</td>
<td>82</td>
<td>3.0</td>
<td>SSSVPGRGSIEGRPEF</td>
</tr>
<tr>
<td></td>
<td>gp21</td>
<td>88</td>
<td>2.5</td>
<td>AAA</td>
</tr>
<tr>
<td></td>
<td>Monobody YSX1</td>
<td>91</td>
<td>2.0</td>
<td>GSSGSS</td>
</tr>
<tr>
<td></td>
<td>Monobody YS1</td>
<td>93</td>
<td>1.8</td>
<td>GSS</td>
</tr>
<tr>
<td></td>
<td>IPS1 CARD domain</td>
<td>93</td>
<td>2.1</td>
<td>SAMA</td>
</tr>
<tr>
<td></td>
<td>CRFR1 ECD</td>
<td>96</td>
<td>1.96-3.4</td>
<td>AAAEFP</td>
</tr>
<tr>
<td></td>
<td>ZP3 (ZP-N domain)</td>
<td>102</td>
<td>2.3-3.1</td>
<td>AAA</td>
</tr>
<tr>
<td></td>
<td>L30</td>
<td>103</td>
<td>2.31</td>
<td>SSSVPGRGSIEGRA</td>
</tr>
<tr>
<td></td>
<td>gp2136</td>
<td>108</td>
<td>—</td>
<td>AAA</td>
</tr>
<tr>
<td></td>
<td>designed helical protein</td>
<td>108</td>
<td>1.9</td>
<td>SSSNRRRRRRRRRRR</td>
</tr>
<tr>
<td>GST (~217AA)</td>
<td>GP41 neutralizing antigen</td>
<td>6</td>
<td>2.5</td>
<td>SDLVPRGSM</td>
</tr>
<tr>
<td></td>
<td>C-terminal Fibrinogen gamma chain</td>
<td>14</td>
<td>1.8</td>
<td>SDP</td>
</tr>
<tr>
<td></td>
<td>alpha-Na/K ATPase</td>
<td>25</td>
<td>2.6</td>
<td>SDLVPRGS</td>
</tr>
<tr>
<td></td>
<td>(ankyrin binding domain)</td>
<td>31</td>
<td>2.7</td>
<td>SDLVPRGSRRASVGS</td>
</tr>
<tr>
<td></td>
<td>AML-1 NMTS</td>
<td>53</td>
<td>2.5</td>
<td>SDLVPRGS</td>
</tr>
<tr>
<td>Lysozyme (~163AA)</td>
<td>TELSAM + E80-TELSAM</td>
<td>80</td>
<td>2.4-2.6</td>
<td>AGP</td>
</tr>
<tr>
<td></td>
<td>β3AR</td>
<td>282</td>
<td>2.8</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>β2AR</td>
<td>282</td>
<td>2.4</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Adenosine A2A receptor</td>
<td>308</td>
<td>2.6</td>
<td>—</td>
</tr>
<tr>
<td>Thioredoxin (~109AA)</td>
<td>PuI60 UHM</td>
<td>100</td>
<td>2.2</td>
<td>GSAM</td>
</tr>
<tr>
<td></td>
<td>VanH</td>
<td>322</td>
<td>3.0</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table I. Successfully Solved Structures Using Carrier-Driven Crystallization

a All target proteins listed in this table are fused C-terminal to the carrier protein, unless otherwise stated.
b Information not available.
c TELSAM was fused N-terminal to the lysozyme carrier protein.
d Residues 231–262 of the β2AR have been replaced by residues 2-161 of T4 lysozyme.
e Adenosine A2A receptor residues 209–221 have been replaced by residues 2-161 of T4 lysozyme.
f VanH was fused N-terminal to the thioredoxin carrier protein. The structure of this fusion protein has not yet been reported.
method minimizes the risk of inadvertent mutagenesis of biologically important residues on the target protein—a potential hazard of the SER technique. Here, we present three case studies of unrelated target proteins (ranging in size from 198 to 324 amino acids) whose structures were determined using this approach.

Discussion
To date, in our laboratory, we used the tandem fixed-arm MBP/SER mutation system toward the crystallization of seven unrelated proteins whose structures could not be solved by conventional means. We have successfully solved the structures of three of these proteins—2OST (Table I), 42 RACK1A, and Derp7 (Mueller et al (doi:10.1016)). We have also obtained crystals for two (TargetA and TargetB) of the remaining four proteins (Table III). Co-crystals of TargetA (431 amino acids) in complex with DNA diffract to 2.3 Å resolution. However, these crystals are badly twinned, and simultaneously, have problems with pseudo-symmetry. Efforts are currently underway to optimize the crystal growth conditions for TargetB (292 amino acids). Thus far, the other two proteins (Targets C and D) have not yielded diffraction-quality crystals.

Taken together, in a field where less than 10% of expressed proteins are solved and deposited (http://www.jcsg.org), we have increased our success rate to over 40%, with proteins that previously have been difficult to solve. Because of the small sample size (n = 7), these values may not be statistically significant, but are definitely encouraging.

Though the MBP/SER system can increase the likelihood of successfully crystallizing difficult proteins, the approach is not infallible. There are specific issues to consider. While N-terminal fusion of the MBP to the target protein can be beneficial for protein expression and purification, the significant size of the MBP may serve to limit the size of the target protein that can then be expressed in E. coli. Historically, MBP fusion proteins with smaller targets, relative to the size of the MBP, have yielded more structures (Table I), though a few targets larger than 150 residues have been determined. The synergistic nature of the additional SER mutations may increase the likelihood that a larger target protein could be crystallized. Placement of the MBP at the N-terminus of the fusion protein is generally considered to be advantageous due to its chaperone-like behavior. However, in cases where the N-terminus of the target protein is crucial for functionality or for protein-protein interaction, N-terminal MBP could be disruptive. On the other hand, it should be noted that MBP fusion proteins have been used to solve structures of large multi-protein complexes.44

In short, the tandem fixed-arm MBP/SER system represents a simple, cost-effective method for crystallization of problematic proteins that encompasses the advantageous properties of carrier-driven crystallization and surface entropy reduction. This technique represents another tool that can be used as a viable rescue strategy for problematic protein crystallography.

Experimental Methods

Vector design
The pMAL-c2x (New England Biolabs) vector was used as the backbone to create the pMALX vector, based on the reported vector,30,36 with a few alterations. The MBP amino acid sequence was originally modified to include mutations E359A/K362A/D363A and was truncated after Asn367, removing the Factor Xa protease cleavage site. The linker region encodes three alanine residues, as in the original vector,36 but was created by insertion of NotI and NheI restriction sites upstream of the EcoRI site in the original pMAL-c2x multicloning region (Fig. 1(A)).

Table II. SER Mutations Present in the MBP-SER Cassettes

<table>
<thead>
<tr>
<th>Vector</th>
<th>SER mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMALX(A)</td>
<td>D82A/K83A</td>
</tr>
<tr>
<td>pMALX (B)</td>
<td>E172A/N173A</td>
</tr>
<tr>
<td>pMALX(C)</td>
<td>D82A/K83A/K239A</td>
</tr>
<tr>
<td>pMALX(D)</td>
<td>E172A/N173A/K239A</td>
</tr>
<tr>
<td>pMALX(E)</td>
<td>D82A/K83A/E172A/N173A/K239A</td>
</tr>
</tbody>
</table>

Table III. Proteins for Which the MBP(SER) System has been Thus Far Unsuccessful

<table>
<thead>
<tr>
<th>Target protein</th>
<th>Biological role</th>
<th>Size</th>
<th>Crystals?</th>
<th>Solved?</th>
</tr>
</thead>
<tbody>
<tr>
<td>TargetA + DNA</td>
<td>Y-family polymerase</td>
<td>431 amino acids</td>
<td>Yes</td>
<td>No (perfect hemihedral twinning)</td>
</tr>
<tr>
<td>TargetB</td>
<td>Protein sulfotransferase</td>
<td>292 amino acids</td>
<td>Yes</td>
<td>No (insufficient diffraction quality)</td>
</tr>
<tr>
<td>TargetC + DNA</td>
<td>X-family polymerase</td>
<td>357 amino acids</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>TargetD</td>
<td>Transcriptional regulator</td>
<td>242 amino acids</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

* Before transferring these proteins to the MBP(SER) system, we first attempted to crystallize them alone or in the presence of ligands. Where appropriate, different species of proteins were used, utilizing multiple construct lengths in each case. These particular target proteins displayed low solubility in the absence of a fusion partner, so SER mutations were not used. In multiple cases, chemical modification was also used. However, these proteins generated no usable crystals.
Creation of SER mutations in MBP

The pMALX vector was used as a template to create the SER mutations, using the QuikChange Mutagenesis protocol (Stratagene). By examining the published structure of MBP in the presence of maltose (PDB ID code 1HSJ38), we generated several mutations on solvent-exposed loop regions in different areas of the protein [Fig. 1(B) and Table II]. These mutations were generated either singly or in pairs on a particular loop, and were used in a combinatorial fashion to create five interchangeable cassettes: pMALX(A) with mutations D82A/K83A, pMALX(B) with mutations E172A/N173A, pMALX(C) with mutations D82A/K83A/K239A, pMALX(D) with mutations E172A/N173A/K239A, and pMALX(E) with mutations D82A/K83A/E172A/N173A/K239A [Fig. 1(B) and Table II].

Rationale of construct design

To maximize the chances of successful crystallization using this technique, initial construct design is crucial. The length of the linker region between the MBP and the protein of interest has been shortened to the extent that it should function more like a “fixed arm,” merging the two proteins into an inflexible crystallization unit (henceforth referred to as the “fusion protein”). Because the exact structure of the N-terminal end of the target protein is unknown, several N-terminal truncation constructs should be screened. If the linker is too long, the fusion protein may have undesired conformational flexibility. Yet, if the linker is not long enough, the connection to the MBP C-terminal α-helix may disrupt the structure of the downstream target protein, leading to misfolding or insolubility.

There are many tools to aid in construct design. We combine available biochemical data for the target protein and structural information from related or orthologous proteins with secondary structure prediction. Given known structural information on plausible domain organization, we design several constructs, utilizing different N-terminal truncations. This method maximizes the likelihood of successfully melding the N-terminal secondary structure of the target protein with the C-terminal α-helix of MBP. If structures of homologous proteins exist, manual docking of the N-terminus of the target protein to the MBP C-terminal α-helix can provide a visual aid to construct design.

Cloning and expression

Multiple constructs of the target protein are fused to the MBP carrier protein using traditional cloning methods. Restriction sites are added to the coding sequence of the target protein by PCR. The PCR products are digested with the appropriate restriction enzyme and ligated into the pMALX(wt) vector. All constructs are sequenced to ensure the absence of undesirable mutations, then transformed into an appropriate bacterial expression cell line. Small-scale test expressions (1L cultures) are performed to determine which constructs express soluble protein. In short, the cells are grown to density (OD600 = 0.5–1.0) at 37°C, at which time the temperature is decreased to 18°C for 30 min. Protein expression is induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.2–0.5 mM and incubation proceeds for 14–16 h. The cells are pelleted by centrifugation and lysed by sonication in an appropriate buffer. The soluble fraction is obtained by centrifugation of the lysate, then bound in-batch to amylose resin (New England Biolabs). Unbound protein is washed from the resin and the bound protein eluted by addition of 40 mM D(-)-maltose dissolved in sonication buffer. The eluted protein is concentrated and loaded onto a Superdex200 16/60 size exclusion column (GE Healthcare) equilibrated in elution buffer containing 40 mM D(-)-maltose. The ligand is added to the
chromatography buffer for optimal protein separation because the MBP fusion protein can exhibit nonspecific interaction with the column matrix in the absence of maltose (data not shown). Size exclusion chromatography is likely crucial for success using the MBP/SER technique. Because the presence of MBP dramatically increases the expression and solubility of the resulting fusion protein, a substantial population of the protein can be incorrectly folded or aggregated. Such protein appears in the void volume peak from a size exclusion column and is easily separated from the more well-behaved and crystallizable target population.

As a general rule, the smallest target protein constructs with the shortest linker, and exhibiting the highest level of desired activity or function are used for crystallization. Active, soluble, and well-behaved constructs are transferred to the pMALX vectors harboring MBP/SER mutations by simple restriction and ligation. All constructs, including the wildtype, are then prepared in a large-scale purification (12 L cultures), using the same protocol as described above.

In many cases, the fusion protein is sufficiently pure for crystallization after the two-step purification protocol of affinity and size exclusion chromatography. However, further chromatography steps can be carried out if a higher extent of purification is required. After purification, the fusion protein is dialyzed into a buffer appropriate for crystallization— a low salt, non-phosphate buffer, with the addition of 5 mM D-(+)-maltose. Maltose is included in the crystallization buffer because the MBP exhibits conformational heterogeneity dependent on the presence or absence of ligand. Crystallization trials are then carried out using commercially available screens. Crystals obtained in the initial screens are subsequently optimized for size and quality and then used for data collection.

Solving the phasing problem with molecular replacement

A key advantage of the MBP fusion system is that the MBP protein structure can be used to solve the crystallographic phase problem by molecular replacement. Once molecular replacement solutions for the MBP have been found, these molecules can be fixed, and a second molecular replacement search for the target protein can be performed if quality search models exist. However, if no such search model exists, the phase information gained from the MBP may be sufficient for the target molecule to be built manually, utilizing an iterative process of density modification and solvent flattening to improve the maps, followed by model building and refinement. Alternatively, automated model building programs such as AutoBuild in Phenix have been tested and shown in some cases to provide helpful starting models for chain tracing the target protein if no acceptable molecular replacement models exist.

Case study #1: 2OST from Gallus gallus

Heparan sulfate (HS) is a highly sulfated linear polysaccharide found ubiquitously on the surface of cells, and in the surrounding extracellular matrix. Specifically sulfated saccharide sequences act as binding sites for proteins involved in a myriad of biological processes, ranging from embryonic development, inflammatory response, and blood coagulation (reviewed in 80). 2-O-sulfotransferase (2OST) is involved in HS biosynthesis, transferring a sulfo group to the 2-OH position of iduronic or glucuronic acid within an HS chain. This sulfation is critical for substrate recognition in blood coagulation and for signal transduction through fibroblast growth factor-mediated pathways.82

We recently determined the crystal structure of chicken 2OST using the MBP/SER system.42 Initial attempts to structurally characterize 2OST involved crystallization of a His6x-tagged construct of the catalytic domain (amino acids Arg63-Asn356) of hamster 2OST (94% similarity to chicken 2OST). This construct formed somewhat disordered crystalline rods which diffracted poorly to 3.3 Å. The structure could not be solved from these data. To generate a maximal change in the crystallization surface, while preserving physiological relevance, the catalytic domain (Asp69-Asn356) of chicken 2OST (92% similarity to human 2OST) was fused downstream of the MBP carrier protein. Asp69 was selected as the starting residue based on conserved secondary structural elements found in both cytosolic and membrane-bound sulfotransferases. The sequence encoding these amino acids was inserted between the NotI and BamHI restriction sites in the pMALX(wt) vector.42

Expression of the MBP-2OST fusion protein gave a greatly increased yield of soluble protein, compared to the His6x-tagged protein (unpublished data). The fusion protein also exhibited markedly improved solubility, both in solution and during crystallization trials, and proved active in sulfotransferase assays.42 The MBP-2OST protein crystals were obtained at 4°C using the sitting-drop vapor diffusion method.83

The published structure of MBP (PDB ID code 1HSJ) was used to solve the phase problem via molecular replacement with the program MOLREP from the CCP4 Suite.4 A single molecule of MBP was found within the asymmetric unit in this crystal form. A generic search model comprised of the conserved secondary structural elements distinctive to known sulfotransferases was generated using amino acids Asp39-Tyr62, Arg78-Glu83, Ile104-Ile146, Leu189-Phe210, and Pro269-Lys285 from estrogen sulfotransferase (EST, PDB ID code 1AQU) and...
was used to determine the relative position of 2OST to MBP with a second round of molecular replacement in MOLREP. Density modification in CNS was utilized to improve the phases so that model building could be carried out. The final model includes all residues in MBP and residues Asp69-Lys354 of chicken 2OST [Fig. 2(A)].

Though only a single molecule of MBP-2OST is present in the asymmetric unit in this space group, examination of the crystal packing reveals a three-fold crystallographic axis running through the center of a trimer [Fig. 2(B)]. Size exclusion chromatography of MBP-2OST suggests that the protein behaves as a trimer in solution. The trimeric interface is mediated by C-terminal residues Asn345-Tyr352, which form an additional β-strand antiparallel to strand β5 of another monomer. 2OST truncation mutants lacking the C-terminal residues exhibited greatly decreased trimer formation, and concomitantly reduced sulfotransferase activity. These observations indicate that the trimeric form may be the biologically active unit, and most notably, that addition of the MBP carrier protein did not disrupt this interaction [Fig. 2(B)]. The position of all three MBP molecules, grouped proximally at the N-termini of the 2OST molecules, could be viewed as a possible arrangement of the N-terminal transmembrane domains present in the native proteins. Such an arrangement is consistent with current knowledge of the physiological behavior of 2OST. In vivo, this sulfotransferase is associated with the luminal side of the Golgi membrane, where it modifies oligosaccharide groups on glycoproteins later bound for the cell membrane and the extracellular matrix.

In this case study, SER mutations were not required to obtain diffraction quality crystals. The E359A mutation within the C-terminal α-helix of MBP was fortuitous in this instance, as deletion of the wildtype glutamate residue allows a tighter interaction between the MBP and 2OST molecules [Fig. 2(C)]. A glutamate at position 359 would have likely created an electrostatic clash with the adjacent Glu265 residue on the 2OST molecule, preventing domain arrangement and crystallization.

Case study #2: RACK1A from Arabidopsis thaliana

The receptor for activated C-kinase 1 (RACK1) belongs to the WD40 repeat family of β-propeller proteins and is highly conserved in eukaryotes. RACK1 is believed to function as a scaffold protein, mediating multiple simultaneous protein-protein interactions involved in diverse cellular processes including regulation of transcription, signal transduction pathways, and ribosome assembly (reviewed in Refs. 89 and 90). RACK1 is highly conserved in eukaryotes—mouse and chicken orthologs are identical to the human protein. RACK1A, one of three RACK1 isoforms in A. thaliana, exhibits 66% sequence conservation to human RACK1. The high degree of sequence conservation across these diverse species suggests a conserved function in eukaryotes.

Initial crystallization attempts with untagged RACK1A failed to produce crystals with commercially available crystallization matrix screens. Therefore, the sequence encoding RACK1A amino acids Gly4-Tyr327 was cloned into the pMALX expression
system, using the NotI and BamHI restriction sites, leaving the shortest possible linker between the two proteins. Gly4 was chosen as the starting residue based on secondary structure prediction, and by manually docking the N-terminus of related β-propeller proteins to the C-terminus of the MBP. 43

The structure of the MBP(C)-RACK1A is shown in Figure 3(A). A detailed analysis of the structure within the crystal lattice identifies several key features. Shortening the linker between the MBP and the RACK1A was effective in this fusion protein—only one residue at the junction (Gly4) was not involved in a secondary structural element. Given the brevity of the linker connecting the two proteins, there are surprisingly few inter-domain interactions between MBP and RACK1A [Fig. 3(A)]. Residues along the interface are a minimum of 5 Å apart. In addition, all interactions forming the crystal lattice are either MBP-MBP or MBP-RACK1A. There are no RACK1A-RACK1A interactions, which may explain the failure of crystallization attempts with RACK1A alone.

The position of the D82A/K83A/K239A mutations within the lattice suggests a preference for crystal formation with these specific mutations over other surface mutations [Fig. 3(B,C)]. D82A and K83A are located on the surface of the MBP, near two different symmetry-related molecules [Fig. 3(B)]. The presence of large aspartate and lysine residues could have negatively influenced crystal packing. Similarly, the loop containing K239A forms a tight crystal lattice contact, and a lysine side chain could have disrupted crystal formation [Fig. 3(C)]. Conversely, Asn173 participates in a hydrogen bonding lattice contact with the backbone of Cys138 of a symmetry-related RACK1A molecule [Fig. 3(D)]. An alanine substitution at this position could have been detrimental to formation of this lattice contact. Because of the impossibility of determining beforehand which residues may be either advantageous or detrimental to lattice formation, the availability of multiple cassettes with different combinations of SER mutations improves the odds of crystallization.

Case study #3: Der p 7 from Dermatophagoides pteronyssinus

Conditions stemming from allergic response—that is, rhinitis, asthma, anaphylaxis—are currently a major health concern. One of the most common sources of indoor allergens is the house dust mite. To date, multiple groups of dust mite allergens have been identified, many of which have been found to

Figure 3. Crystal structure of the MBP(C)-RACK1A fusion protein (PDB ID code 3DM0). A: Ribbon diagram of the MBP(C)-RACK1A fusion protein. The bound maltose is drawn in stick (cyan). The structure of the MBP is shown in gray, relative to that of the RACK1A (multicolored). The β-strands of the RACK1A β-propeller domain are shown as ribbons, with each WD motif colored and labeled. The residues of the disordered loop (Lys277-Lys294) in WD6 (purple) are labeled. B, C, and D: Secondary structural elements of the MBP(C) protein are drawn as gray ribbons, with the SER mutations drawn in stick. The positions of symmetry-related MBP molecules (gray) and RACK1A (green, blue, yellow) are drawn in stick. B: Position of the D82A/K83A SER mutations in crystal lattice formation. C: Position of the K239A mutation in crystal lattice formation. D: Role of Asn173 in crystal lattice formation.
be enzymes.\textsuperscript{91} Allergens belonging to Groups 5 and 7 are proteins of unknown function.\textsuperscript{92} Der p 7-reactive IgE serum antibodies are found in only 53\% of allergic patients,\textsuperscript{93} however, the scale of the response can be as great as for the more major allergens.\textsuperscript{94} To gain some insight into the abnormal patient response, and the natural protein function in dust mites, we solved the structure of the Group 7 allergen from \textit{Dermatophagoides pteronyssinus}, Der p 7 (Mueller \textit{et al}, doi:10.1016).

Based on sequence alignment, Der p 7 is not related to any known protein. However, analyzing the protein sequence with the GenTHREADER alignment server\textsuperscript{95} suggested that Der p 7 may have a fold similar to two lipid-binding proteins—human Bactericidal/permeability-increasing protein (BPI, PDB ID code 1EWF\textsuperscript{96}) and the Takeout 1 protein from \textit{Epiphyas postvittana} (PDB ID code 3E8T\textsuperscript{97})—though with expectedly low sequence identities (14\% and 12\%, respectively).

His6x-tagged Der p 7 was readily expressed in soluble form in \textit{E. coli}, and quickly generated a variety of crystals. However, these crystals were poorly ordered, and did not yield quality diffraction. Der p 7 residues Asp18-Gln215 were cloned into the pMALX cassettes using the NotI and BamHI restriction sites. The N-terminal region of Der p 7 carries a signal peptide, which was removed in this construct. Residue 18 was used as the starting residue for this construct based on secondary structure prediction. The fusion protein behaved similarly to the natural allergen in IgE binding assays (Mueller \textit{et al}, doi:10.1016). Crystals of MBP-Der p 7 were obtained using the MBP cassette bearing the D82A/K83A/E172A/N173A/K239A SER mutations. These crystals diffracted at 2.35 Å (Mueller \textit{et al}, doi:10.1016).

In this crystal form, there are three MBP(E)-Der p 7 molecules within the asymmetric unit. Following positioning of the MBP molecules using molecular replacement with an MBP model (PDB ID code 3DM0\textsuperscript{98}), discontinuous electron density was visible for the Der p 7 molecule. Although molecular replacement models using the structurally related proteins BPI (PDB ID code 1EWF\textsuperscript{96}) and Takeout1 (PDB ID code 3E8T\textsuperscript{97}) were unsuccessful, these structures provided a guide for chain tracing and manual model building. Molecule A from the final MBP(E)-Der p 7 model is shown in Figure 4(A). All three of the Der p 7 molecules superimpose well (RMSD of less than 0.51Å over 197 C\textsubscript{a} atoms). Though Der p 7 currently has no known enzymatic activity, the structural similarities between Der p 7 and BPI or Takeout1 suggested a possible role in binding of hydrophobic ligands. Subsequent HSQC and STD-NMR experiments have demonstrated interaction of the bacterial lipopeptide polymixin B with Der p 7 (Mueller \textit{et al}, doi:10.1016). Such interactions with hydrophobic ligands may provide insight into the allergenicity of Der p 7. The binding of hydrophobic ligands is thought to inappropriately stimulate an allergic immune response.\textsuperscript{99}

A detailed analysis of the MBP(E)-Der p 7 crystal packing yielded some interesting observations. The most extensive interactions appear to involve MBP-MBP interactions (Fig. 4(B)), which, as for 2OST and for RACK1A, appear to be the driving force for crystal lattice formation. All permutations of MBP-MBP,
MBP-Der p 7, and Der p 7-Der p 7 interactions exist between symmetry-related molecules.

Determining the role of the SER mutations is more complex for the Der p 7 crystal lattice than for either 2OST or RACK1A. For both the Molecules A and B, K239A lies along the interaction surface between the two MBP molecules. The presence of a lysine side chain at this junction could have disrupted lattice packing if the original lysine residue were still present [Fig. 3(C)]. For the D82A/K83A and E172A/N173A mutations, there would have been ample space within the crystal to accommodate the large side chains. The general reduction of side chain entropy could have been a contributing factor. In the case of K239A, however, there is clear evidence that this particular SER mutation may have played a key role in forming this crystal lattice.

Acknowledgments
The authors thank Traci Hall and Michael Murray for critical reading and thoughtful comments on the manuscript. The expression vectors are freely available upon request.

References

27. Ke A, Wolberger C (2003) Insights into binding cooperativity of MATa1/MATalpha2 from the crystal structure


65. Chen GQ, Gouaux JE (1996) Overexpression of bac-
64. Grisham R, Little J, Aharony D (1994) Expression of
62. Edwards AM, Arrowsmith CH, Christendat D, Dhar-
61. Sachdev D, Chirgwin JM (2000) Fusions to maltose-
60. Kapust RB, Waugh DS (1999) Escherichia coli maltose-
59. Shih YP, Kung WM, Chen JC, Yeh CH, Wang AH, 
58. Pinhal MA, Smith B, Olson S, Aikawa J, Kimata K, 
57. Rong J, Habuchi H, Kimata K, Lindahl U, Kusche-
56. Shen HD, Chua KY, Lin KL, Hsieh KH, Thomas WR 
54. Vaing AT (1997) MOLEP: An automated program for 
53. Pinhal MA, Smith B, Olson S, Aikawa J, Kimata K, 
51. Rong J, Habuchi H, Kimata K, Lindahl U, Kusche-
50. Bishop JR, Schuksz M, Esko JD (2007) Heparan sul-
48. Pinhal MA, Smith B, Olson S, Aikawa J, Kimata K, 
45. Kakuta Y, Pedersen LG, Carter CW, Negishi M, Peders-
44. Shen HD, Chua KY, Lin KL, Hsieh KH, Thomas WR 
43. Chapman MD, Pomes A, Breiteneder H, Ferreira F, 
42. Thomas WR (1993) Mite allergens groups I-VII. A cata-
41. Thomas WR (1993) Mite allergens groups I-VII. A cata-
40. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Mor-
39. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Mor-
38. Thomas WR (1993) Mite allergens groups I-VII. A cata-
36. Spurlinko JC, Lin GY, Quiocco FA (1991) The 2.3-A reso-
35. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Mor-
34. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
33. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Mor-
32. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
31. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Mor-
30. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
29. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
28. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
27. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
26. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
25. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
24. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
23. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
22. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
21. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
20. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
19. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
18. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
17. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
16. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
15. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
14. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
13. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
12. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
11. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
10. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
9. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
8. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
7. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
6. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
5. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
4. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
3. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
2. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ, 
1. Cooper DR, Boczek T, Grelewska K, Pannu NS, Read RJ,