Salvage or Recovery of Failed Targets by Mutagenesis to Reduce Surface Entropy

ARTICLE *in* METHODS IN MOLECULAR BIOLOGY (CLIFTON, N.J.) · MARCH 2014

Impact Factor: 1.29 · DOI: 10.1007/978-1-4939-0354-2_16 · Source: PubMed

3 AUTHORS, INCLUDING:

Zygmunt S Derewenda
University of Virginia
170 PUBLICATIONS **10,379** CITATIONS

Chapter 16

Salvage or Recovery of Failed Targets by Mutagenesis to Reduce Surface Entropy

Lukasz Goldschmidt, David Eisenberg, and Zygmunt S. Derewenda

Abstract

The success of macromolecular crystallization depends on the protein's ability to form specific, cohesive intermolecular interactions that serve as crystal contacts. In the cases where the protein lacks surface patches conducive to such interactions, crystallization may not occur. However, it is possible to enhance the likelihood of crystallization by engineering such patches through site-directed mutagenesis, targeting specifically residues with high side chain entropy and replacing them with small amino acids (i.e., surface entropy reduction, SER). This method has proven successful in hundreds of crystallographic analyses of proteins otherwise recalcitrant to crystallization. Three representative cases of the application of the SER strategy, assisted by the automated prediction of the mutation sites using the SER prediction (SERp) server are described.

Key words Protein crystallization, Surface entropy reduction, Site-directed mutagenesis, Protein engineering, Crystal contacts

1 Introduction

That even a single amino acid replacement at the protein's surface can critically affect solubility and propensity for crystallization has long been established long [1, 2]. However, the idea that it is possible to rationally engineer patches that would systematically enhance protein crystallizability is relatively new [3, 4]. The Surface Entropy Reduction (SER) concept was initially suggested on the basis of a thermodynamic argument that the presence on the protein's surface of residues with large, flexible side chains, characterized by high conformational entropy (e.g., Lys and Glu), interferes with the ability of the protein to form stable crystal contacts. This is because the immobilization of such side chains at the site of a contact would increase the entropic cost of crystallization [3, 5, 6]. Thus, mutational modification of surface patches rich in large amino acids and their replacement with residues with small side chains was suggested as a way of engineering sites with enhanced

Wayne F. Anderson (ed.), *Structural Genomics and Drug Discovery: Methods and Protocols*, Methods in Molecular Biology, vol. 1140, DOI 10.1007/978-1-4939-0354-2_16,

ability to mediate crystal contacts [7]. Subsequent experimental [3, 5], theoretical [8], and data mining [9, 10] studies strongly supported this hypothesis. A number of protein structures solved using crystals generated by the SER strategy revealed that the mutated sites almost invariably were involved in crystal contacts, providing a firm validation of the underlying hypothesis [4, 11–14]. The three types of amino acids that are of particular significance are Lys, Glu, and Gln. Crystallizable variants can be obtained by mutating these amino acids to such residues as Ala, Ser, Thr, or even Tyr, the side chain of which has only two degrees of rotational freedom [7]. It has been shown that often two or three simultaneous mutations of residues located in close proximity on the surface are necessary for optimal effect [7]. The key question is what specific sites on a given protein's surface should be targeted for mutagenesis. Therefore, we implemented an automated algorithm that allows the investigator to identify such sites from amino acid (or cDNA) sequence alone was implemented [15].

Here we describe three representative cases of the application of the surface entropy reduction strategy using automated identification of target surface sites, assisted by the SERp server. The information obtained from the server can be used in several ways. In one example, three different variants were generated, each with a different mutated patch. In the second example, only one surface site was chosen, but three distinct variants were generated by replacing the target residues with alanines, threonines, or tyrosines. In the final example, five patches identified by SERp were all mutated into single variants using alanines for replacement. In each case, one or more of the variants yielded crystals suitable for diffraction studies and the corresponding structures were published and deposited in the Protein Data Bank.

2 Materials

The SER strategy is typically used when the target protein, or a complex, fails to crystallize despite extensive screening. Thus, all relevant expression constructs and purification protocols are normally well established and followed.

The Surface Entropy Reduction prediction (SERp) server is accessible to academic and commercial users free of charge online at http://services.mbi.ucla.edu/SER. A job submission requires only the amino acid or cDNA sequence of the target protein and a valid e-mail address. The default parameters are recommended for new users, though the ability to adjust most processing parameters is also provided.

Desired SER mutations are typically introduced into the target protein using QuikChange® site-directed mutagenesis kit (Agilent Technologies) and should be confirmed by direct DNA

sequencing. Multiple mutations in close proximity can be introduced in a single reaction if they can be placed within a single primer, or alternatively the Multi-Site version of the QuikChange protocol, or QuikChange® Multi, can be used. The latest generation of these kits (i.e., QuikChange® Lightning) provide accelerated protocols capable of handling short and long templates. Protein expression and purification are then carried out as established for the wild-type protein.

Commercial crystallization kits can be purchased from various companies and methods for high-throughput crystallization screening are described elsewhere in this volume (*see* **Note 1**).

3 Methods

To illustrate the utility of the SER approach, we describe the application of the method to three distinct examples of proteins that failed to produce diffraction-quality crystals in their wild type form. These proteins are: (1) A 144-residue, monomeric prokaryotic disulfide isomerase YphP from *B. subtilis*; (2) Met6p, a 789-residue fungal methionine synthase; and (3) the 364-residue catalytic domain (amino acids 364) of the human vaccinia-related protein kinase 1 (VRK1).

3.1 Using the SERp Server: General Remarks

After opening the SERp server's URL in the browser, the user simply pastes the protein (or cDNA) sequence into the proper text box and enters the e-mail address, to which the results of the analysis are to be sent. Although providing a short name or identifier for the protein is optional, use of this field is encouraged to help with the identification of results. All results are stored on the server and can be retrieved for up to 30 days after the submission by using the submitter's e-mail address and a personal identification number (PIN). The server can also be used to screen a larger set of sequences by using the Batch Submission mode. This mode accepts the common FASTA file format as input. Results can be conveniently viewed and sorted on the server's Web site; selected results can also be e-mailed to the submitter.

The default processing parameter values are suitable for most users and do not need to be adjusted. At the core of the SER algorithm are three primary analyses: the entropy profile, secondary structure prediction, and the evolutionary conservation prediction [15]. Those analyses should remain enabled unless the user has a strong reason to omit one. For example, to exclude evolutionary information from multiple sequence alignments, the evolutionary conservation prediction can be turned off by unchecking the corresponding check box, or by setting the weight for this analysis to zero. The server can operate without secondary structure prediction if the user deems these predictions to be unreliable, although we

found the secondary structure predictions to be beneficial in virtually all cases, and thus do not recommend disabling this analysis.

Once the submitted sequence has been analyzed (typical processing times are less than a minute), a list of suggested mutations is presented on the results "Summary" tab. Mutations are organized in clusters and are reported in order of predicted success. Clusters with a SERp score of 3.0 (arbitrary scale) and above are considered as optimal targets. Details about the calculation of each cluster's SERp score can be found on the "Score Details" tab, where the contribution of each primary analysis is shown on a per-residue basis. Important data are also presented in graphical form on the "Graphs" tab. The first graph titled "Overall Score" explains why a particular set of mutations is predicted to be effective in conferring enhanced crystallizability.

Briefly, ideal candidate sites are non-conserved, and contain clusters of high side chain entropy residues (i.e., Lys, Glu and Gln) that lie in solvent-exposed loops of the protein. The stacked graph shows the average side chain entropy in red, and the confidence that a particular region falls between secondary structure elements in blue. Such regions are presumed to have a higher likelihood to lie on the surface. Residues that are found to be conserved in multiple sequence alignments are panelized relative to their conservation level, while the occurrence of proposed replacement residues is scored favorably and is also shown in green on the stacked graph.

Each proposed cluster is highlighted on the graph, and its rank and score are shown near the top. The residue patterns pre- and post-mutation within each cluster are also important and are thus included near the bottom of the graph. The overall aim is to generate new consecutive patches of residues with low side chain entropy, which have the potential to form crystal contacts. Such new contacts can then contribute to the stability of the crystal or produce new crystal forms with superior diffraction quality. Regions that could form a low-entropy patch are shaded in gray, and the proposed replacement residues within are highlighted in green.

Results from the Meta Search are also summarized on this graph, but are not utilized in the selection algorithm. It is up to user to evaluate those additional pieces of information. For example, detected sequence signatures for highly conserved regions of proteins are shown at the bottom of the graph in pink. If possible, mutations in those regions should be avoided. Homologous structures in the PDB with high sequence similarity are used to estimate the surface accessibility of residue (*see* **Note 2**). Results from this analysis are shown at the bottom of the graph for each residue, with darker gray colors indicating higher solvent accessibility.

Analysis of larger or more difficult targets may benefit from minor adjustments of the processing parameters. For example, the cluster limit cutoff can be increased to suggest more cluster candidates. If other replacement residues besides alanine should be considered,

those residues can be listed as potential "Target amino acids" in the respective field (*see* **Note 3**). To avoid replacement of a certain residue type or to consider mutations of other residues with high side chain entropy, the "Mutable amino acid" list can be adjusted (*see* **Note 4**). By default the server will suggest a maximum of three mutations per cluster, where a cluster is defined as a patch of consecutive "high entropy" or "target" residues, with gaps no larger than two amino acids in length. Larger clusters with more replacements can be also considered if the target's solubility allows it.

3.1.1 *B. subtilis* Prokaryotic Disulfide Isomerase YphP

The SER analysis for the amino acid sequence of YphP from *B. subtilis*, with default parameters, identifies three candidate clusters containing high side chain entropy residues (underlined) in putative solvent exposed locations: ^{39}GluLysAlaGlu^{42}; ^{100}GlnGlu^{101}; and ^{113}Lys,Glu^{114}. The respective scores for these sites are: 5.2, 3.8, and 4.4 (Fig. 1). Six variants were generated, each containing mutations within a single cluster, either to alanines or tyrosines (*see* **Note 2**). Although the first cluster has the highest SERp score, the authors found that the variants targeting this site yielded only poor-quality crystals under a variety of conditions. However, the second most highly scored site, with alanines replacing the two high entropy residues, gave reproducible crystals in two different conditions and ultimately yielded a structure at 2.3 Å resolution [16].

3.1.2 *C. albicans* Methionine Synthase Met6p

The SERp analysis of the amino acid sequence of the methionine synthase Met6p identifies only three potential target clusters; of these, a three-residue cluster (^{103}LysLysAlaThrGlu^{107}) in a predicted surface exposed loop has by far the highest score, 6.1 (Fig. 2). A comparison with known homologues of bacterial and plant origin confirmed the surface location of the loop [17]. The investigators chose to focus on this site alone, and generated three variants, so that each had the high-entropy residues replaced with alanines, threonines, or tyrosines (*see* **Note 2**). The catalytic activity of the protein was not affected. Each of the three variants yielded a distinct crystal form, suitable for diffraction studies. The resolution of the diffraction data ranged between 2.0 to 2.8 Å.

3.1.3 The Catalytic Domain of the Human Vaccinia-Related Kinase 1 (VRK1)

This potentially important drug target was under study for several years before the SER strategy finally brought the project to fruition (Dr. C. K. Allerston, personal communication). The SERp server identified a total of four clusters rich in high-entropy residues, i.e., ^{34}LysLysGlu^{36}; ^{212}GluTyrLysGlu^{215}; ^{292}GluLysAsnLys^{295}; and ^{359}LysLys^{360} (the current version actually identifies five clusters, including ^{312}GluLys^{313}—*see* **Note 2,** Fig. 3). In this case, the investigators decided to mutate all these clusters simultaneously, thus introducing 11 mutations into protein (*see* **Notes 5** and **6**). The modified variant yielded a crystal form that diffracted to 2.4 Å (unpublished; PDB code 3OP5).

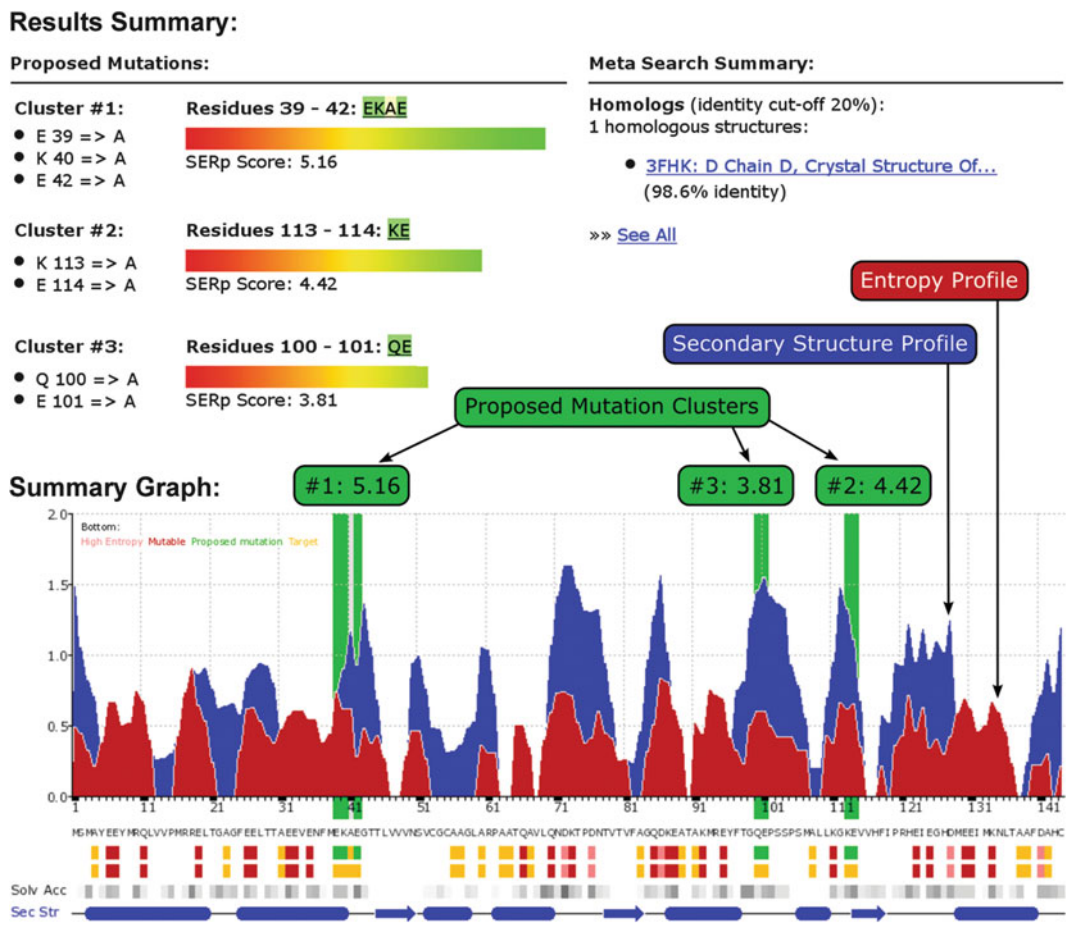

Fig. 1 Summary of the key results from the SER analysis of the *B. subtilis* prokaryotic disulfide isomerase YphP. Proposed mutations are organized in clusters and are reported in order of predicted success (*top left*). Clusters with a SERp score of 3.0 and above are optimal targets, as indicated by the *color gradient* of the *horizontal bar*. The residue composition of each cluster is shown. Residues suggested for mutation are shaded *green*, existing low entropy residues are shaded *yellow*. Results from the Meta Search for homologous structured in the Protein Data Bank (PDB) for conserved sequence signatures are summarized on the *top right*. The stacked summary graph shows the sidechain entropy and secondary structure profiles in *red* and *blue*, respectively. All clusters are *highlighted* on the graph, with the cluster rank and score annotations near the top (*green boxes*). The residue sequence and the residue patterns pre and post mutation are displayed below the profile (high entropy residues in *pink*, mutable resides in *red*, suggested mutations in *green* and replacement residues in *yellow*). The estimated solvent accessibility of each residue, computed from homologous structures in the PDB, is shown next with darker *grays* indicating higher accessibility. Finally, the predicted secondary structure is shown using the common cartoon diagram

4 Notes

1. We do not observe any specific correlation between the nature of the engineered variants and conditions under which they crystallize, although no systematic analysis has been carried out.

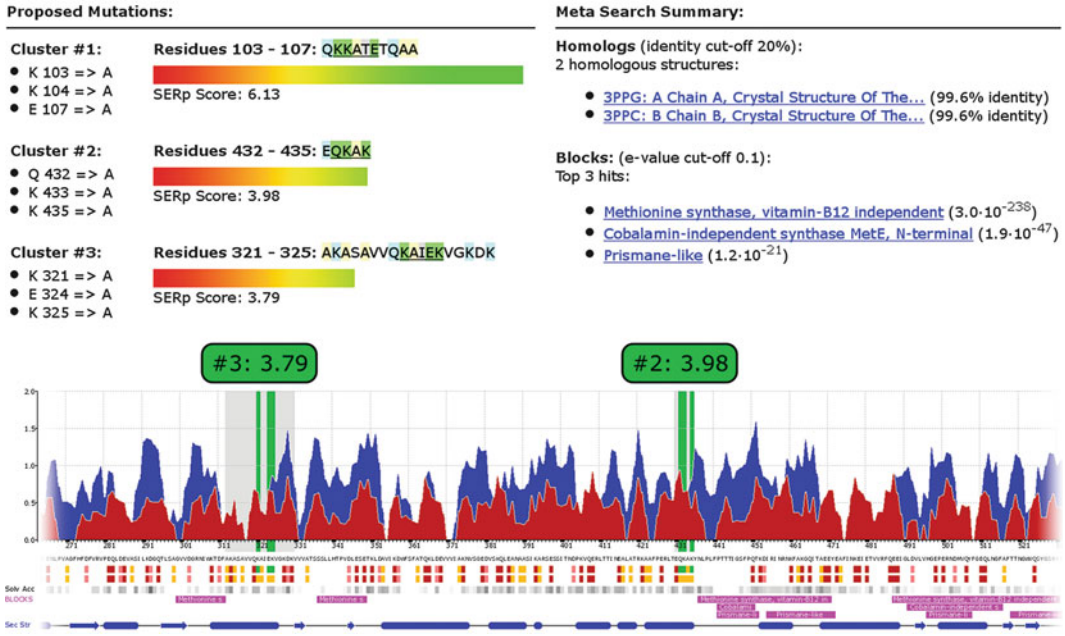

Fig. 2 SER analysis results for *C. albicans* methionine synthase Met6p. The annotations in this figure are analogous to those in Fig. 1. For this target, additionally several sequence signatures were found. The top three hits are named in the Blocks section in the Meta Search Summary with the respective expectation values. Regions that match each signature are marked in *pink* below the stacked profile graph. Due to space reasons, only the residue region 270–532 is shown

In those cases where the wild-type protein is known to crystallize, albeit yielding poor-quality crystals, the variants typically crystallize under different conditions. It is therefore recommended that all SER variants are screened de novo. In our laboratories we use various commercial screens for this purpose, but we observe that chances of success or vastly increased with the alternative reservoir strategy [7, 18].

2. The databases utilized by the SERp server are updated regularly (at least monthly), and thus cluster scores and rankings may change over time. At present, homology modeling is not explicitly used by the SERp algorithm. This information, and utilization of information about crystal contacts in homologous structures, will be used in version 2.0.
3. The current algorithm has no capability of predicting what type of an amino acid is best suited for a particular cluster, and by default suggests Ala. Most of the successful examples of the SER application use this strategy, but as shown in our examples, other types of amino acids (e.g., Thr or Tyr) may also yield good results. At the moment this is purely the investigator's call.
4. The current SERp algorithm targets Lys, Glu and Gln residues for mutagenesis. Obviously, removal of a charged, polar residue

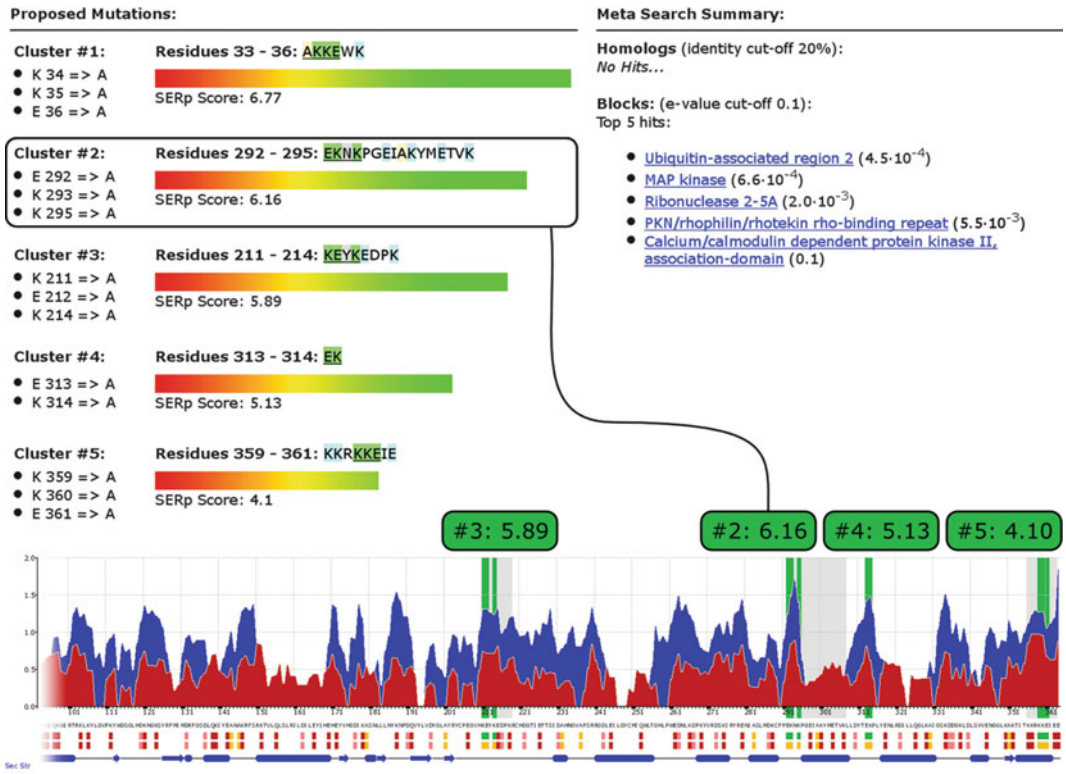

Fig. 3 SER analysis results for the catalytic domain of the human vaccinia-related kinase 1, VRK1. Similar to Figs. 1 and 2, the highest-scoring five clusters are shown, annotated equivalently. A large cluster comprised of residues 292–307 is highlighted (*box*) because of regularly spaced residues with high side chain entropy. Within the cluster only three residues near the beginning on the cluster are suggested for mutation because they lie between two helices, as evidenced by the peak of the stacked graph (*bottom*). Due to space reasons, only the C-terminal region with residues 100–363 is shown

from the surface will have an effect on the protein solubility. While this is not built into our prediction, the next version of the server will attempt to predict mutations that are least deleterious to solubility. At the moment, we recommend that the users target Lys-rich clusters in preference to Glu.

5. The "multi-patch" strategy is rarely used and requires further investigation. Nevertheless, several crystal structures have been successfully obtained using crystals obtained by this approach. One other such example is the *Arabidopsis* prenyl pyrophosphate synthase [19].
6. In all the examples listed, and in the vast majority of the crystal structures obtained through the SER strategy, the engineered surface patches are directly involved in crystal contacts. It is always instructive to examine not only the details of the target protein or complex, but also those of the packing features to better understand the crystallization process.

Acknowledgment

This work was supported by the National Institutes of Health, grant GM095847.

References

1. McElroy HH, Sisson GW, Schottlin WE et al (1992) Studies on engineering crystallizability by mutation of surface residues of human thymidylate synthase. J Cryst Growth 122:265–272
2. Dale GE, Broger C, Langen H, D'Arcy A, Stuber D (1994) Improving protein solubility through rationally designed amino acid replacements: solubilization of the trimethoprim-resistant type S1 dihydrofolate reductase. Protein Eng 7:933–939
3. Longenecker KL, Garrard SM, Sheffield PJ, Derewenda ZS (2001) Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. Acta Crystallogr D 57:679–688
4. Derewenda ZS (2004) Rational protein crystallization by mutational surface engineering. Structure 12:529–535
5. Mateja A, Devedjiev Y, Krowarsch D, Longenecker K, Dauter Z, Otlewski J, Derewenda ZS (2002) The impact of Glu → Ala and Glu → Asp mutations on the crystallization properties of RhoGDI: the structure of RhoGDI at 1.3 A resolution. Acta Crystallogr D 58:1983–1991
6. Derewenda ZS, Vekilov PG (2006) Entropy and surface engineering in protein crystallization. Acta Crystallogr D 62:116–124
7. Cooper DR, Boczek T, Grelewska K, Pinkowska M, Sikorska M, Zawadzki M, Derewenda Z (2007) Protein crystallization by surface entropy reduction: optimization of the SER strategy. Acta Crystallogr D 63:636–645
8. Pellicane G, Smith G, Sarkisov L (2008) Molecular dynamics characterization of protein crystal contacts in aqueous solutions. Phys Rev Lett 101:248102
9. Cieslik M, Derewenda ZS (2009) The role of entropy and polarity in intermolecular contacts in protein crystals. Acta Crystallogr D 65:500–509
10. Price WN 2nd, Chen Y, Handelman SK, Neely H, Manor P, Karlin R, Nair R, Liu J, Baran M, Everett J, Tong SN, Forouhar F, Swaminathan SS, Acton T, Xiao R, Luft JR, Lauricella A, DeTitta GT, Rost B, Montelione GT, Hunt JF (2009) Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. Nat Biotechnol 27:51–57
11. Longenecker KL, Lewis ME, Chikumi H, Gutkind JS, Derewenda ZS (2001) Structure of the RGS-like domain from PDZ-RhoGEF: linking heterotrimeric g protein-coupled signaling to Rho GTPases. Structure 9:559–569
12. Derewenda U, Mateja A, Devedjiev Y, Routzahn KM, Evdokimov AG, Derewenda ZS, Waugh DS (2004) The structure of *Yersinia pestis* V-antigen, an essential virulence factor and mediator of immunity against plague. Structure 12:301–306
13. Janda I, Devedjiev Y, Derewenda U, Dauter Z, Bielnicki J, Cooper DR, Graf PC, Joachimiak A, Jakob U, Derewenda ZS (2004) The crystal structure of the reduced, Zn^{2+}-bound form of the *B. subtilis* Hsp33 chaperone and its implications for the activation mechanism. Structure 12:1901–1907
14. Bielnicki J, Devedjiev Y, Derewenda U, Dauter Z, Joachimiak A, Derewenda ZS (2006) *B. subtilis* ykuD protein at 2.0 A resolution: insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. Proteins 62:144–151
15. Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D (2007) Toward rational protein crystallization: a web server for the design of crystallizable protein variants. Protein Sci 16:1569–1576
16. Derewenda U, Boczek T, Gorres KL, Yu M, Hung LW, Cooper D, Joachimiak A, Raines RT, Derewenda ZS (2009) Structure and function of *Bacillus subtilis* YphP, a prokaryotic disulfide isomerase with a CXC catalytic motif. Biochemistry 48:8664–8671
17. Ubhi D, Kavanagh KL, Monzingo AF, Robertus JD (2011) Structure of *Candida albicans* methionine synthase determined by employing surface residue mutagenesis. Arch Biochem Biophys 513:19–26
18. Newman J (2005) Expanding screening space through the use of alternative reservoirs in vapor-diffusion experiments. Acta Crystallogr D 61:490–493
19. Hsieh FL, Chang TH, Ko TP, Wang AH (2011) Structure and mechanism of an *Arabidopsis* medium/long-chain-length prenyl pyrophosphate synthase. Plant Physiol 155: 1079–1090