Application of Protein Engineering Strategies in Structural Biology for Enhancing Protein Crystallization

Xiaoli Gao[1a#], Yanfeng Zhang[1a#]*, Yixiao Shen[2], Xiangqun Shen[3]

[1]Department of Genetics, Michigan State University, East Lansing, MI, USA
[2]College of Food Science, and [3]College of Horticulture, Shenyang Agricultural University, Shenyang, Liaoning Province, China
[a]Present address: Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA
[#]These authors contributed equally to this work
*Corresponding author: zhangy11@gmail.com

Abstract: Structural biology is one of the most important areas in biological sciences since detailed 3-D atomic protein structure not only gives direct information on protein function, but also provides useful knowledge on protein engineering and drug design. X-ray crystallography is one of the most powerful tools for high-resolution protein structure determination. It requires growth of protein crystals, which is extremely challenging for some proteins and usually pose it the most rate-limiting step for protein structure determination. However, protein engineering methods improving the entropy of crystallization sometimes enhance protein crystallization. In this review, we summarized the most commonly used protein engineering strategies for improve the chance of protein crystallization.

[Xiaoli Gao, Yanfeng Zhang, Yixiao Shen, Xiangqun Shen. Application of Protein Engineering Strategies in Structural Biology for Enhancing Protein Crystallization. Journal of American Science 2011;7(7):696-699] (ISSN: 1545-1003). http://www.americanscience.org.

Keywords: Protein engineering, crystallization, X-ray crystallography, limited proteolysis, MBP, surface entropy reduction

Introduction

X-ray crystallography and NMR are two most powerful approaches for studying high resolution protein structure. NMR does not require protein crystals, but sometimes is difficult to solve structure of protein with large size. X-ray crystallography does not have a size limitation, but needs diffractable protein crystals. For some protein, growing crystals are extremely challenging. However, modification of protein or crystallization conditions sometimes may improve the chance of protein crystallization. One common way of crystallizing proteins is to co-crystallize proteins in complex with cofactors, inhibitors or antibody fragments. The conformational change induced by such ligand bindings may be favorable to the crystallization by exposing new crystal contacts or by ordering the protein structure. For instance, for a GTPase, crystallizing the apo form protein and co-crystallizing the GTPase with GDP should be attempted. Co-crystallizing the GTPase with GTP analogs such as GppNHp and GTP-γS is also helpful. If a protein is membrane related or interacts with membrane components, crystallizing the protein with lipid analogs such as detergents and the head group of lipids will be necessary.

Recent studies have shown that the protein itself should be modified during protein crystallization [1]. Besides altering precipitating agents, the protein should be considered as an important variable in the crystallization screen. Strategies of modifications of target proteins via protein engineering tools in order to improve the chance of protein crystallization are summarized below.

Limited proteolysis to determine protein domain boundaries

Proteins, especially from eukaryotes, are usually complicate and contain several domains. The flexibility caused by connections between domains generates high conformational heterogeneity and usually is one of the most important factors to be considered for crystallization [2]. To solve this problem, it is useful to identify crystallizable functional domains of such proteins for X-ray crystallographic study. One challenging aspect of this study is to identify the domain boundaries. Multiple protein sequence alignments combined with secondary structure prediction is one of the methods to identify the domain boundaries. However, a more promising method is limited proteolysis followed by mass spectrometry protein sequence determination. The rationale for limited proteolysis is that the connections between domains are structural flexible and therefore sensitive to protease digestion. In contrast, bulk protein domains are compact and are less accessible by proteases [2]. Therefore, limited proteolysis is applicable to multi-domain proteins if full-length proteins fail to crystallize.

The fastest way to perform limited proteolysis is in-drop proteolysis [3,4]. Different kinds of

proteases such as trypsin, chymotrypsin, papain, and proteinase K will be added to the solution containing the full-length protein just before the crystallization screening is carried out. Crystals grown in the drops will be resolubilized and subjected to mass spectrometry for mass determination and sequence analysis [3]. The in-drop method is easy to manipulate, however, its disadvantage is the heterogeneity resulted from other digested fragments that may cause failure of the compact digested domains to crystallize. If there is no crystal grown using the in-drop method, the regular limited proteolysis will be performed.

To perform regular limited proteolysis, the partial digestion patterns of full-length target proteins using the proteases listed above will be compared. The digestion will be applied to SDS-PAGE and the products will be characterized either by electroblotting to PVDF followed by N-terminal sequencing or by mass spectrometry. Based on the obtained sequences, the protease-resistant regions of the target proteins will be cloned into expression vectors. The expressed recombination proteins will be used for crystallization trials.

Surface-entropy reduction approach to enhance protein crystallizability

The basis of the surface-entropy reduction is to reduce the entropic cost of protein crystallization by modifying the target protein using protein engineering methods. Typically, residues with large flexible side chains in solvent-exposed loops cost more entropy to crystallize and should be mutated to small amino acids [5]. Sometimes, instead of single amino-acids, entire flexible regions of a protein generating conformational heterogeneity may also cause difficulty in crystallization. Thus, deleting or replacing the flexible regions may be helpful in the crystallization by reducing the interfering effects from the heterogeneity [1,6,7].

For example, surface-energy prediction of a GTPase protein revealed that the high ratio of charged and flexible residues of a domain in the middle of the protein causes high surface entropy (Figure 1). Secondary structure predictions and protein disorder prediction showed that this domain does not have stable secondary structure and probably forms a large solvent-exposed loop (Figure 2). Comparison of domain topology of the target protein with crystallized homologues from other organisms indicated that the domain is not present in the crystallized homologues. These analyses indicate that the domain may be highly flexible causing an energy barrier for crystal formation. Therefore, replacement of this domain with a shorter but less mobile linker by protein engineering may be a way to enhance the likelihood of protein crystallization.

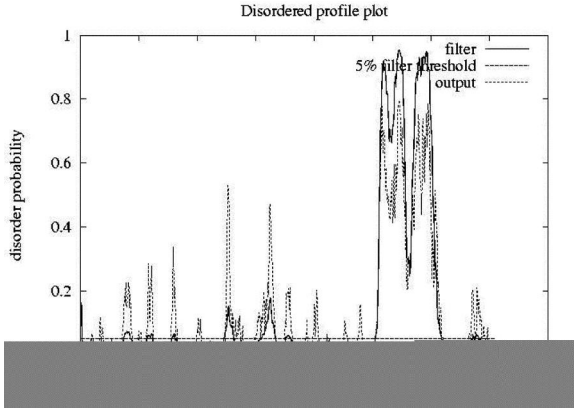

Figure 1: Prediction of disordered region of a GTPase by the server DISOPRED. Residues from 500 to 600 were indicated as the disordered region.

Some candidates for the linkers are FLAG or *Strep* II and the protein expression level in *E. coli*, protein solubility and the GTPase activity can be used as indicators of protein folding of the target protein.

Large fusion tags as crystallization partner

Large fusion tags such as maltose binding protein (MBP) and glutathione-S-transferase (GST) have been commonly used to enhance the expression, improve the yield and stability, and facilitate purification of the protein to which they are fused [8-10]. Recently, several protein crystal structures have been reported by fusion with *E. coli* MBP [11-14]. With a modified linker between the MBP and the protein of interest, the presence of the MBP does not seem to interfere with the native structure of the target protein, as indicated by the crystal structures. Several advantages of co-crystallizing a protein with fusion partner have been recognized. First, large fusion tags may enhance solubility and stability of target proteins by avoiding formation of inclusion bodies in *E. coli* during expression. Second, since the crystal contacts are dominated by MBP/MBP or MBP/protein interactions, the fusion partner could facilitate crystallization by increasing those contacts. Third, the conditions used to crystallize the tag and the crystal contacts found in the tag crystal may be used to guide the crystallization of the fusion protein. Last but not least, the three-dimensional structure of tags can be used as a search model to solve the crystallographic phase problem by molecular replacement.

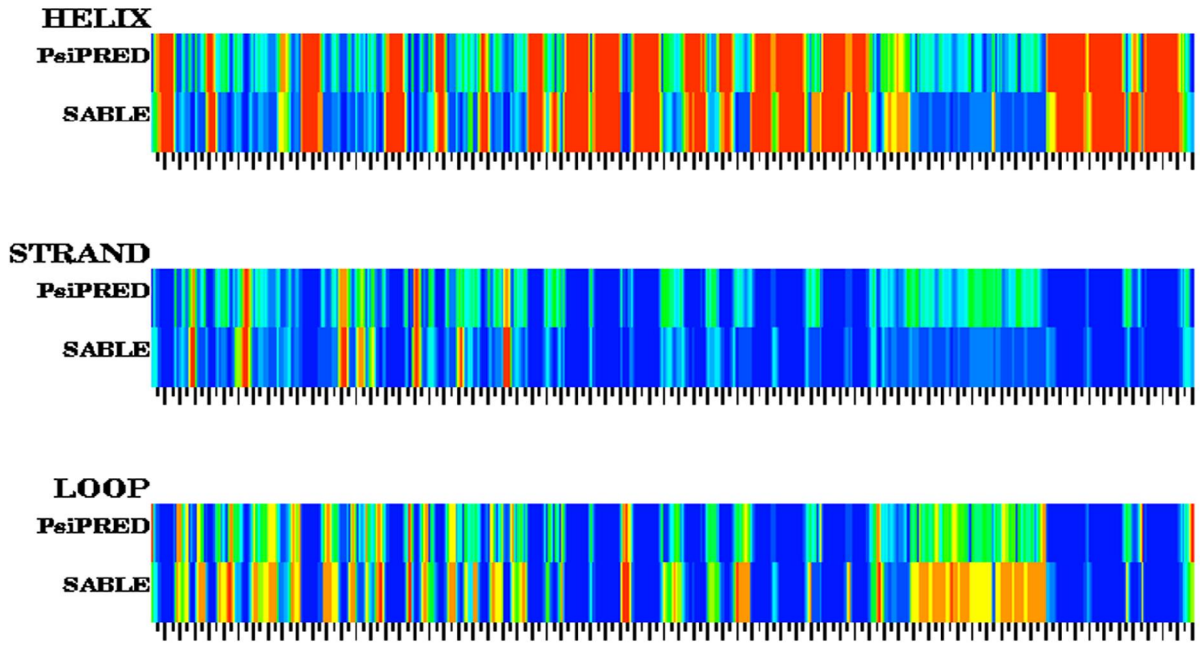

Figure 2: **An example of comparison of secondary structure predictions of a GTPase from different servers (PsiPRED and SABLE) by visualizing the predictions.** Red, high possibility of being an indicated secondary structure; blue, low possibility of being an indicated secondary structure; yellow and green, possibility between the red and blue.

One challenging aspect of co-crystallizing fusion proteins is the modification of the linker between the tag and the target protein. The characteristic shared by most successfully crystallized fusion proteins is a short rigid connection such as three alanines instead of a long flexible linker. Shorter linkers may help avoid the conformational heterogeneity introduced by flexible linkers. Therefore, the linker region should be short but allows flexibilities for the target protein. Meanwhile, besides *E. coli* MBP, the *Pyrococcus furiosus* (*Pfu*) MBP is also a good choice as a fusion partner since the *Pfu* MBP has been proven to be a more potent solubilizing agent than *E.coli* MBP [15].

An alternative large fusion tag is the DsRed Monomer (Invitrogen), which is an engineered mutant of a red fluorescent protein from *Discosoma* sp. reef coral. The tetrameric form of native DsRed is not a suitable crystallization carrier because fusion with an oligomeric tag may result in a chimera protein with a non-native quaternary structure of the target protein. However, monomeric form of engineered DsRed Monomer is unlikely to affect the native structure of the target protein. A significant advantage of DsRed Monomer is that the red color of the fusion protein can be visualized directly by the naked eye, so that the protein expression and purification process can easily be monitored. Moreover, the red color can also be used as an indicator to differentiate the protein crystals from those of salts, which is a common problem associated with crystallization using microbatch evaporation methods.

Lipidic cubic phase

For membrane proteins, an alternative crystallization method is to use a lipidic cubic phase, also called the *in meso* method. Several membrane proteins have been crystallized for high-resolution structure determination using this method [16-19]. In the cubic phase, the lipidic compartments are interpenetrated by a freely communicating system of aqueous channels [20]. Although the exact mechanism of the *in meso* crystallization remains unclear, the cubic phase may provide a lipid bilayer that is an environment similar to the biological membranes. The membrane protein may reconstitute into the bilayer and crystals nucleate and grow upon addition of precipitants [21].

Acknowledgements

Y. F. Z would like to thank Dr. R. Michael Garavito for educating protein chemistry and X-ray crystallography.

References

[1] G.E. Dale, C. Oefner, A. D'Arcy, The protein as a variable in protein crystallization, J Struct Biol 142 (2003) 88-97.

[2] C.M. Koth, S.M. Orlicky, S.M. Larson, A.M. Edwards, Use of limited proteolysis to identify protein domains suitable for structural analysis, Methods Enzymol 368 (2003) 77-84.

[3] R.K. Gaur, M.B. Kupper, R. Fischer, K.M. Hoffmann, Preliminary X-ray analysis of a human V(H) fragment at 1.8 A resolution, Acta Crystallogr D Biol Crystallogr 60 (2004) 965-967.

[4] S. Johnson, P. Roversi, M. Espina, J.E. Deane, S. Birket, W.D. Picking, A. Blocker, W.L. Picking, S.M. Lea, Expression, limited proteolysis and preliminary crystallographic analysis of IpaD, a component of the Shigella flexneri type III secretion system, Acta Crystallogr Sect F Struct Biol Cryst Commun 62 (2006) 865-868.

[5] Z.S. Derewenda, P.G. Vekilov, Entropy and surface engineering in protein crystallization, Acta Crystallogr D Biol Crystallogr 62 (2006) 116-124.

[6] T.U. Schwartz, R. Walczak, G. Blobel, Circular permutation as a tool to reduce surface entropy triggers crystallization of the signal recognition particle receptor beta subunit, Protein Sci 13 (2004) 2814-2818.

[7] A.R. Kim, T. Dobransky, R.J. Rylett, B.H. Shilton, Surface-entropy reduction used in the crystallization of human choline acetyltransferase, Acta Crystallogr D Biol Crystallogr 61 (2005) 1306-1310.

[8] D. Sachdev, J.M. Chirgwin, Fusions to maltose-binding protein: control of folding and solubility in protein purification, Methods Enzymol 326 (2000) 312-321.

[9] D.B. Smith, Generating fusions to glutathione S-transferase for protein studies, Methods Enzymol 326 (2000) 254-270.

[10] A. Skerra, T.G. Schmidt, Use of the Strep-Tag and streptavidin for detection and purification of recombinant proteins, Methods Enzymol 326 (2000) 271-304.

[11] B. Kobe, R.J. Center, B.E. Kemp, P. Poumbourios, Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins, Proc Natl Acad Sci U S A 96 (1999) 4319-4324.

[12] Y. Liu, A. Manna, R. Li, W.E. Martin, R.C. Murphy, A.L. Cheung, G. Zhang, Crystal structure of the SarR protein from Staphylococcus aureus, Proc Natl Acad Sci U S A 98 (2001) 6877-6882.

[13] A. Ke, C. Wolberger, Insights into binding cooperativity of MATa1/MATalpha2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera, Protein Sci 12 (2003) 306-312.

[14] Y.F. Zhang, X.L. Gao, R.M. Garavito, Structural analysis of the intracellular domain of (pro)renin receptor fused to maltose-binding protein, Biochemical and Biophysical Research Communications 407 (2011) 674-679.

[15] J.D. Fox, K.M. Routzahn, M.H. Bucher, D.S. Waugh, Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers, FEBS Lett 537 (2003) 53-57.

[16] M. Kolbe, H. Besir, L.O. Essen, D. Oesterhelt, Structure of the light-driven chloride pump halorhodopsin at 1.8 A resolution, Science 288 (2000) 1390-1396.

[17] H. Luecke, B. Schobert, J.K. Lanyi, E.N. Spudich, J.L. Spudich, Crystal structure of sensory rhodopsin II at 2.4 angstroms: insights into color tuning and transducer interaction, Science 293 (2001) 1499-1503.

[18] V.I. Gordeliy, J. Labahn, R. Moukhametzianov, R. Efremov, J. Granzin, R. Schlesinger, G. Buldt, T. Savopol, A.J. Scheidig, J.P. Klare, M. Engelhard, Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex, Nature 419 (2002) 484-487.

[19] G. Katona, U. Andreasson, E.M. Landau, L.E. Andreasson, R. Neutze, Lipidic cubic phase crystal structure of the photosynthetic reaction centre from Rhodobacter sphaeroides at 2.35A resolution, J Mol Biol 331 (2003) 681-692.

[20] E.M. Landau, J.P. Rosenbusch, Lipidic cubic phases: a novel concept for the crystallization of membrane proteins, Proc Natl Acad Sci U S A 93 (1996) 14532-14535.

[21] M. Caffrey, Membrane protein crystallization, J Struct Biol 142 (2003) 108-132.

7/1/2011