Video Article

Combining Wet and Dry Lab Techniques to Guide the Crystallization of Large Coiled-coil Containing Proteins

Jenna K. Zalewski[1], Simone Heber[1,2], Joshua H. Mo[1], Keith O'Conor[1], Jeffrey D. Hildebrand[1], Andrew P. VanDemark[1]

[1]Department of Biological Sciences, University of Pittsburgh

[2]Institute of Structural Biology, German Research Center for Environmental Health

Correspondence to: Andrew P. VanDemark at andyv@pitt.edu

Abstract

Obtaining crystals for structure determination can be a difficult and time consuming proposition for any protein. Coiled-coil proteins and domains are found throughout nature, however, because of their physical properties and tendency to aggregate, they are traditionally viewed as being especially difficult to crystallize. Here, we utilize a variety of quick and simple techniques designed to identify a series of possible domain boundaries for a given coiled-coil protein, and then quickly characterize the behavior of these proteins in solution. With the addition of a strongly fluorescent tag (mRuby2), protein characterization is simple and straightforward. The target protein can be readily visualized under normal lighting and can be quantified with the use of an appropriate imager. The goal is to quickly identify candidates that can be removed from the crystallization pipeline because they are unlikely to succeed, affording more time for the best candidates and fewer funds expended on proteins that do not produce crystals. This process can be iterated to incorporate information gained from initial screening efforts, can be adapted for high-throughput expression and purification procedures, and is augmented by robotic screening for crystallization.

Video Link

The video component of this article can be found at https://www.jove.com/video/54886/

Introduction

Structure determination via X-ray crystallography has made fundamental contributions to every field of modern biology; providing an atomic view of the macromolecules that support life and how they interact with one another in a variety of contexts; allowing us to understand the mechanisms that cause disease and providing opportunities to rationally design drugs to treat disease. Crystallography has long been the dominant experimental technique for determining macromolecular structure, and currently accounts for 89.3% of the structural database (www.rcsb.org). This technique has many advantages, including the potential for very high resolution, the ability to visualize macromolecules with a broad range of sizes, relatively easy data collection, and the opportunity to visualize how the macromolecule interacts with solvent as well as ligands.

Despite numerous technological improvements in recombinant protein expression[1,2], purification[3], and molecular biology used to generate these systems[4], the single biggest obstacle in the crystallographic process remains the ability to grow diffraction quality crystals. This has been especially true for proteins which contain large coiled-coil domains. It has been estimated that as much as 5% of all amino acids are found within coiled-coils[5,6], making this a very common structural feature[7], yet these proteins are often more difficult to purify and crystallize than globular proteins[8-10]. This is further compounded by the fact that coiled-coil domains are often found within the context of a larger protein, therefore correctly predicting the boundaries of these domains is critical to avoid the inclusion of unstructured or flexible sequence that is often detrimental for crystallization.

Here we present a conceptual framework combining web-based computational analyses with experimental data from the bench, to help guide users through the initial stages of the crystallographic process including: how to select protein fragment(s) for structural studies, and how to prepare and characterize protein samples prior to crystallization attempts. We focus our analysis on two proteins containing large coiled-coil domains, Shroom (Shrm) and Rho-kinase (Rock). These proteins were chosen as they both contain coiled-coil domains and are known to form a biologically relevant complex[11-16]. Shroom and Rho-kinase (Rock) are predicted to contain ~200 and 680 residues of coiled-coil respectively, many portions of which have been characterized structurally[17-20]. The method described here provides a streamlined workflow to quickly identify fragments of coiled-coil containing protein that will be amenable for crystallization, however, the techniques described can easily be adapted for most protein or protein-complexes or modified to incorporate high-throughput approaches as available. Lastly, these methods are generally inexpensive and can be performed by users at nearly all experience levels.

Protocol

NOTE: A diagram of the conceptual framework or workflow is described in **Figure 1** for reference. The protocol can be broken down into four stages: computational or sequence based predictions, protein expression and purification, biochemical characterization, and crystallization. The examples shown analyze Shroom SD2 domains and/or Shroom-Rock complexes, but can be utilized with any protein.

1. Use Established Web-based Tools to Generate Computational Predictions of Coiled-coil Domain Boundaries

1. Collect an Evolutionarily Diverse Set of Sequence Homologs.
    1. Go to www.uniprot.org and type in Shroom in the upper search bar.
    2. Select sequences to download by checking the box next to their accession number. Sequences with a star indicate sequences reviewed by Uniprot and are generally more reliable. Take care to ensure that all sequences are complete and accurate.
    3. Save the selected sequences by clicking the Download button.
    4. Align the collection of Shroom sequences using Clustal-Omega[21] (http://www.ebi.ac.uk/Tools/msa/clustalo/). Click the "Browse" button to select the file with the Shroom sequences and then push Submit.
    5. After the alignment is complete, click the "Download Alignment File" button.
    6. Open the multiple sequence alignment generated in 1.1.5 using Jalview (File|Input Alignment|From file). Within the new alignment window, color by identity (Color|Percent Identity).

2. Calculate predictions of secondary structure, predictions of flexible or disordered sequence, and predicted coiled-coil regions within the protein(s) of interest.
    1. Go to http://gpcr.biocomp.unibo.it/cgi/predictors/cc/pred_cchmm.cgi, to calculate Coiled-Coil predictions. Copy and Paste the sequence into the Sequence Box and click Submit.
    NOTE: This analysis may take a couple of hours to complete. Since Shroom sequences are quite large, the sequences may need to be split into blocks of less than 1,000 amino acids to fit within the size constraint of the webserver. When analyzing Shrm2, it is recommended that the C-terminal 1,000 amino acids are used as this section contains the Shrm SD2 domain which is known to contain coiled-coil regions. When repeating this analysis with other proteins, it is recommended to use the full-length protein sequences when possible. If that is not an option, use the largest fragment possible or dissect the sequence based on known biochemical or functional data.
    2. Go to http://gpcr.biocomp.unibo.it/cgi/predictors/cc/pred_cchmm.cgi, to calculate Coiled-Coil predictions. Copy and Paste the sequence into the Sequence Box and click Submit.

3. Combine the results of steps 1.1 to 1.2.2 into a single comprehensive annotation of the Shroom sequence.
    NOTE: It is highly encouraged to also include any and all relevant information that can be gleaned from the literature or other sources about protein function or purification at this stage.
4. Using this comprehensive annotated sequence, predict domain boundaries (or a series of possible boundaries) for the protein, trying to maximize conservation and predicted structural features while minimizing the amount of disordered or flexible sequence. If known, the resulting protein should retain the functional properties of interest.

2. Express and Purify Proteins with the Domain Boundaries Identified in Section 1

NOTE: The goal of this section is to use a series of quick and easily quantifiable assays to screen hypothetical domain boundaries generated in Section 1.

1. Using standard molecular biology techniques, generate an expression plasmid with the desired coding sequence in frame within the His_{10}-mRuby2-XH2 plasmid (See **Figure 3** for the vector map and additional details).
2. Test expression levels for each expression plasmid using BL21(DE3) *E. coli* or other suitable expression strain in autoinduction media[1] at room temperature for 18-24 hr.
    1. Transform expression plasmids as well as an empty plasmid control into BL21(DE3) following the instructions that came with the cells or using robust transformation protocol, such as the one here (https://www.addgene.org/plasmid-protocols/bacterial-transformation/).
    2. From the freshly transformed plate, pick a single colony and grow a 5 ml starter culture at 37 °C overnight in Lysogeny Broth (LB) media with 34 µg/ml kanamycin.
    3. The following day, pellet the culture and wash the pellet with fresh LB.
    4. Add the starter culture to 50 ml of autoinduction media with 34 µg/ml kanamycin. Allow the culture to grow to saturation at either room temperature or 37 °C. Typically, grow for ~18-24 hr.
    5. Pellet cells and freeze the resulting cell pellets at -80 °C indefinitely prior to purification.
    6. Compare whole cell extracts from each expression strain and the empty vector control by SDS-PAGE to determine the strain that most efficiently produces the desired fusion protein. For His_{10}-mRuby2-Shroom SD2 fusion proteins, run 10% SDS-PAGE gels at 180 V for ~50 min or until the dye front reaches the bottom of the gel[22].
    7. Stain gel with Coomassie Blue to visualize total cellular proteins within each strain[23].

3. Perform an initial affinity chromatography step on each strain that successfully expressed mRuby2-fusion protein. Typically, perform purification steps 2.3.2-2.3.7 at room temperature, unless the protein will benefit from altered conditions such as purification at 4 °C.
    1. Thaw cell pellets from step 2.2.5.
    2. Resuspend each cell pellet in 1.5 ml of lysis buffer (10% glycerol, 500 mM NaCl, 40 mM imidazole, 20 mM Tris pH8.0, 1 mM beta-mercaptoethanol). Supplement the lysis buffer with protease inhibitors as appropriate.

3. Add 30 µl of 10 mg/ml lysozyme and incubate at room temperature for 20 min. Sonicate following the instructions for the instrument being used.
4. Transfer into a 1.5 ml tube and centrifuge in a table top centrifuge for 30 min at 14,000 x g to pellet insoluble material or unlysed cells. Save both supernatant and pellet for subsequent analysis via SDS-PAGE.
5. Perform nickel affinity purification, incubating the soluble fraction with 100 µl of Ni-NTA beads. Incubate the beads with soluble lysate for 5-10 min, inverting several times to mix beads.
6. Centrifuge at 800 x g for 30 sec in a tabletop centrifuge to pellet the beads. Follow this by washing the pellet 3-5 times with lysis buffer to clean off non-specific contaminants.
7. Elute Ruby fusion protein from the beads with lysis buffer supplemented with 1 M imidazole.
8. Use SDS-PAGE to compare the behavior of His-mRuby2-Shroom proteins in the "quick and dirty" purification above.

3. Characterize Protein Sample to Identify Those with Advantageous Properties

1. Use a spectrophotometer set at 280 nm to measure the concentration of the Ruby fusion protein, and then assess the homogeneity of the sample by loading 1-5 µg of fusion protein onto a native PAGE gel[24]. Run the 10% Native PAGE gel at 4 °C for 140 min at 175 V.
    1. Image the native PAGE using an imager equipped to visualize fluorescence, observing where the mRuby-tagged fusion migrates in this assay.
        NOTE: Be careful to observe fusion protein that is stuck in the well as this protein is likely aggregated. If a fluorescence imager is not available, one can often visualize and image concentrated samples of Ruby tagged protein under a black light or with a UV light box.

2. Perform limited proteolysis to identify stably folded domains. Incubate 95 µl of Ruby-Shrm SD2 fusion at ~1 mg/ml with 5 µl of 0.025% Subtilisin A.
    1. Sample the reaction at time points of 0, 0.5, 2, 5, 15, 60, and 120 min, removing 10 µl from the reaction for each time point and visualizing the progress of the reaction via SDS-PAGE using a 15% acrylamide gel.
    2. Stain with Coomassie Blue as in step 2.2.7 and evaluate whether a protease resistant species can be identified.

3. Integrate data from the purification efficiency, behavior in native PAGE, and limited proteolysis on the partially purified mRuby2-Shroom fusion proteins into a comprehensive assessment of the overall behavior of these proteins in solution.
    1. (Optional) If a suitable functional assay is available, check for activity at this point.

4. Producing High Quality Crystals of the Coiled-coil SD2 Domain from Shroom

NOTE: All steps within section 4.1 are performed at room temperature unless the protein would benefit from purification at a different temperature, usually 4 °C.

1. Using the 50 ml growths as a rough estimate of expression and purification potential, perform large scale expressions of mRuby2-Shroom SD2 with the goal of achieving 5-20 mg of completely purified sample. Typically 2 L of culture provides adequate starting material. After growth pellet cells by centrifugation at 8,000 x g for 10 min.
    1. Resuspend the pellet from the 2 L growth in lysis buffer as before, using ~2 ml of lysis per gram of frozen cell pellet if using lysozyme for lysis. Alternatively, resuspend at ~8 ml/g of pellet if using a homogenizer or French press. Pellet cellular debris via centrifugation at 30,000 x g for 30 min.
    2. Batch bind the soluble fraction from 4.1.1 using 10 ml of Ni-NTA resin. Pour into appropriate gravity column and allow unbound proteins to drain off.
    3. Wash resin 3-5 times with 40 ml of lysis buffer, followed by a wash with lysis buffer supplemented to 1 M NaCl.
    4. Wash resin with 40 ml of lysis buffer supplemented with 80 mM imidazole.
    5. Elute the Ruby-Shroom fusion protein with 40 ml of lysis buffer supplemented to 1 M imidazole. Manually fractionate the eluted protein into 4-10 ml fractions.
    6. Confirm fractions containing Ruby-Shroom fusion protein using 10% SDS-PAGE.
    7. Dialyze appropriate fractions (typically fractions 2-4) into 8% glycerol, 250 mM NaCl, 15 mM imidazole, 20 mM Tris pH8.0, 1 mM β-ME, using 6-8 kDa MWCO dialysis tubing. Add 1 mg of TEV protease for every 25-50 mg of fusion protein. Dialyze overnight at room temperature with slow stirring.
    8. Repeat nickel affinity purification steps 4.1.2. to 4.1.6. The Shroom SD2 domain should now remain in the unbound fraction while the His$_{10}$-mRuby2 will remain bound to the resin and will be found in the elution fractions. Confirm this using 12% SDS-PAGE staining with Coomassie Blue.
    9. Dialyze fractions containing Shroom SD2 domain into 8% glycerol, 100 mM NaCl, 20 mM Tris pH 8.0, 5 mM beta-mercaptoethanol buffer overnight.
    10. Perform anion exchange chromatography[25] using an FPLC, and eluting with a NaCl gradient from 0.1-1.0 M NaCl. Analyze fractions using 12% SDS-PAGE.
    11. Using a spin concentrator (MWCO of 10 kDa or greater depending on the protein being studied), concentrate peak fractions to ~ 0 mg/ml, and then perform size exclusion chromatography[26], analyzing peak fractions via 12% SDS-PAGE.
    12. Pool peak fractions and dialyze into 20 mM Tris pH 8.0, 50 mM NaCl, 2% glycerol, 1 mM β-ME, and concentrate to 10 mg/ml prior to crystallization.
        Optional: During this step, remove 5 µl samples at concentrations of 1, 2, 5, and 10 mg/ml during the course of concentrating the sample. Run a native PAGE loading 2-5 µl of each sample to look at the behavior of the sample during this step.

2. Screen a small array of 12 conditions to identify an optimal protein concentration for crystallization trials. The composition of these conditions is described in **Figure 7**.

1. Using 24-well sitting drop crystallization trays, perform crystallization trials using the vapor diffusion method, pipetting 500 µl of each of the 12 conditions into separate wells followed by drops that initially contain 1 µl of well solution and 1 µl of protein sample on the coverslips. Please see[27] for additional information on this technique.
2. Quickly seal the tray with clear tape and move the tray to a suitable temperature controlled environment that is vibration free. The temperature used can vary, but 4 °C, 16 °C, and 20 °C are quite common.
3. Examine the drops in the trays over the next 3 days using a microscope at up to 100X magnification. At the end of the 3-day period, score each drop as containing no precipitation (clear), light precipitation, heavy precipitation (brown), or crystals. A suitable protein concentration should contain no more than 6 heavy precipitations.
4. Using the protein concentration identified in 4.2, screen a broad range of commercially available crystallization screens. The use of a liquid handling or crystallization robot greatly speeds up this process while also minimizing error in the drops. It requires far less protein as well, allowing the user to screen more conditions with a single sample.
5. Improve initial conditions identified from broad screens by systematically varying each of the variables within the crystallization drop using 24-well screening trays as before.

Representative Results

A diagram depicting the workflow utilized in this system is shown in **Figure 1** and includes three main stages. Computational analysis of the sequence is utilized to develop hypotheses about the domain boundaries of the coiled-coil protein of interest. An example of an annotated analysis of the Shrm2 SD2 domain is shown in **Figure 2**. In this diagram, the goal was to identify possible domain boundaries for a conserved domain at the C-terminus of the cytoskeletal regulator Shroom called SD2. From this analysis was three distinct sets of hypothetical domain boundaries were generated containing coiled-coil fragments that spanning the entire conserved SD2 or minimal fragments near the C-terminus which ended up as the best candidates for crystallization. Candidates identified from the sequence analysis are then quickly tested in small scale using a protein fusion with mRuby2 (**Figure 3**) for efficient analysis. Representative small scale (50 ml of culture) purifications from two His_{10}-mRuby2-tagged proteins are shown in **Figure 4**. In this figure, the behavior of an insoluble protein is readily apparent as compared to its soluble counterpart. Poorly expressing or insoluble protein fusions are easily and quickly identified in this manner. Biochemical analyses of protein fragments by limited proteolysis are shown in **Figure 5** for a variety of fragments using wither the Ruby system described or with isolated and purified Shroom SD2 domains. In **Figure 5A**, two variants of mouse Shrm SD2 corresponding to hypothetical domain boundaries #2 and #1 in **Figure 3** were digested using a gradient of protease concentration from (0-1.0%) for 30 min at room temperature. Both of these were effectively degraded into a host of smaller products at moderate enzyme concentrations. **Figure 5B** shows the same experiment performed using hypothetical boundary #3 from *Drosophila* Shrm SD2. This protein did crystallize and its structure has been described[19]. Proteolytic analysis can also be useful when examining Ruby-fusions as generated from the protocol above. As shown in **Figure 5C**, a Ruby fusion protein with human Shrm2 SD2 (1427-1610) was digested with a high concentration (0.025%) of the non-specific protease Subtilisin at room temperature for the indicated time points. Here the linker between Ruby and Shrm was immediately cleaved as would be expected. Additionally, minor products were formed indicating this protein has two other protease sensitive regions, however, the products of degradation which were produced within 2 min remained largely intact throughout the rest of the experiment indicating the protein is actually quite stable.

A complementary approach is shown using native PAGE analysis in **Figure 6**. In **Figure 6A**, Ruby-tagged human Shrm SD2 (1427-1610) as well as protein containing the indicated point mutations within Shrm are analyzed by native PAGE. In this experiment, the wild-type protein (which forms crystals) runs as two distinct bands, while the three mutants which do not crystallize have dramatically different behavior. To demonstrate that the system can also be useful on protein complexes, a variety of Shroom-Rock complexes were analyzed by native PAGE in **Figure 6B**. In this case, the same fragment of human Shrm2 SD2 (1427-1610) was used to help clarify the analysis, while different fragments of human Rock were utilized. This approach suggested that complexes formed using Rock 700-906 and 746-906 had multiple species, were smeary, and less homogenous. Complexes utilizing 788-906 were improved, albeit not dramatically, and this species was able to crystallize, although crystals took many weeks to form and contained degraded Rock protein. Complexes generated using Rock 834-913 formed a single and more uniform species on the gel and readily crystallized overnight. **Figure 7** shows a set of common crystallization conditions that are used to inform on the behavior of the protein sample in crystallization trials, and could be used generally with any protein. Ideally, a mix of clear and precipitating conditions will be obtained. Proteins that do not form precipitates require high concentrations or more stringent buffering conditions while those that precipitate in many conditions should be used at a lower protein concentration or with buffering conditions that promote protein solubility.
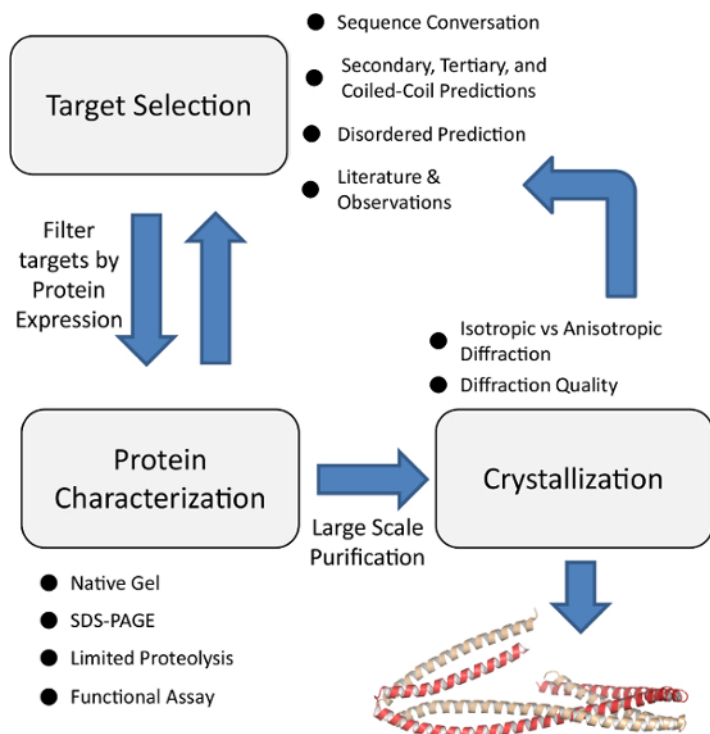
Figure 1: Workflow Diagram. A generalized diagram depicting the integration of computational sequence analysis and biochemical and other wet lab techniques into a comprehensive strategy for delineating domain boundaries and identifying protein fragments for crystallization. Please click here to view a larger version of this figure.
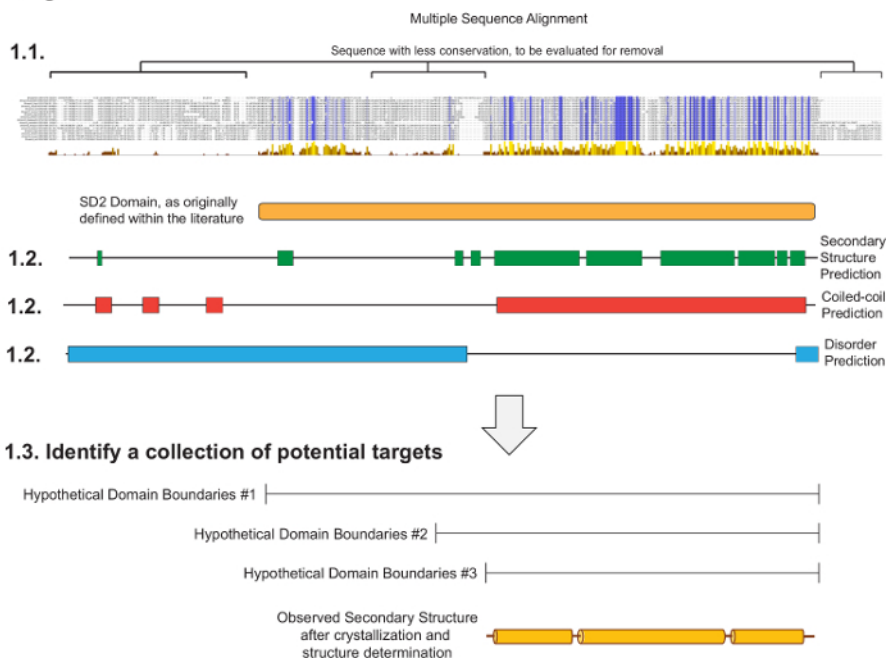

Figure 2: Annotated sequence analysis for the coiled-coil SD2 domain from Shroom. An overlay of computational analyses for the Shroom SD2 domain, including a multiple sequence alignment generated by CLUSTAL omega and colored by sequence identity within Jalview (Step 1.1). Also included are predicted secondary structure, disordered sequence predictions, and coiled-coil predictions of the Shroom SD2 domain (Step 1.2). Hypothetical domain boundaries (Step 1.3) are indicated as are the observed boundaries and secondary structure elements as revealed by subsequent crystallographic analysis[19]. Please click here to view a larger version of this figure.
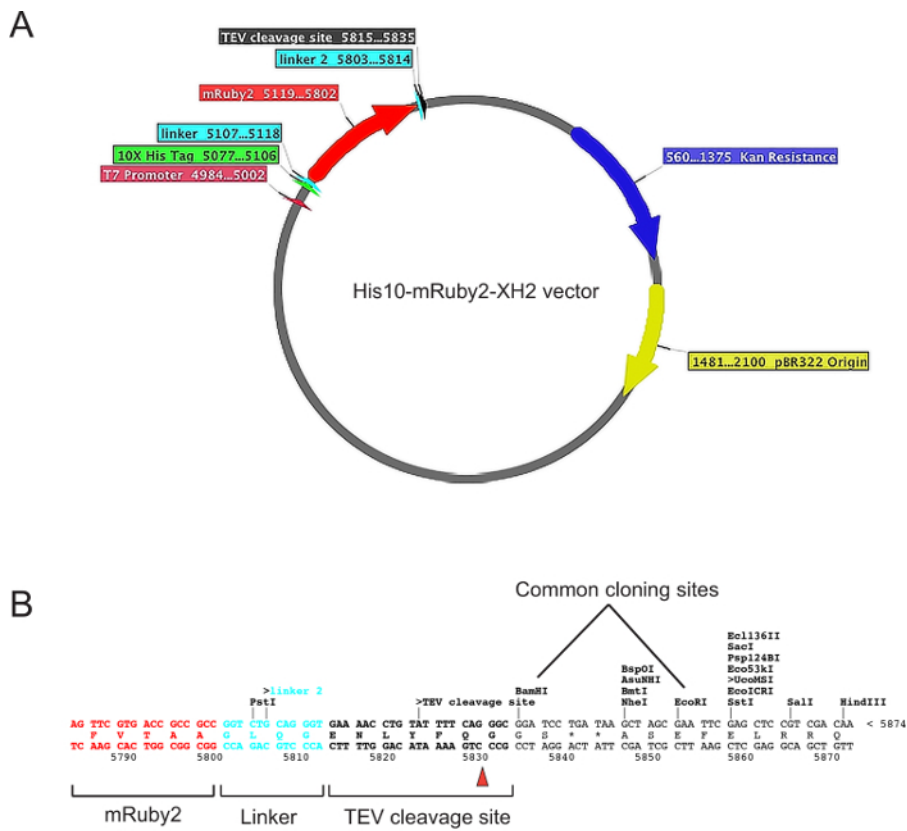
Figure 3: Diagram of the His$_{10}$-mRuby2-expression system. (**A**) Schematic of the expression vector His$_{10}$-mRuby2-XH2 vector used in this study. (**B**) Diagram of the multiple cloning site from this vector. Protein coding sequences are typically inserted into this vector via BamHI and EcoRI cloning sites. The extreme C-terminus of the mRuby2 protein is shown in red, the TEV protease cleavage site is indicated with brackets and with the site of cleavage shown as a red triangle. A linker sequence between the mRuby2 and the TEV site is shown in cyan. Please click here to view a larger version of this figure.
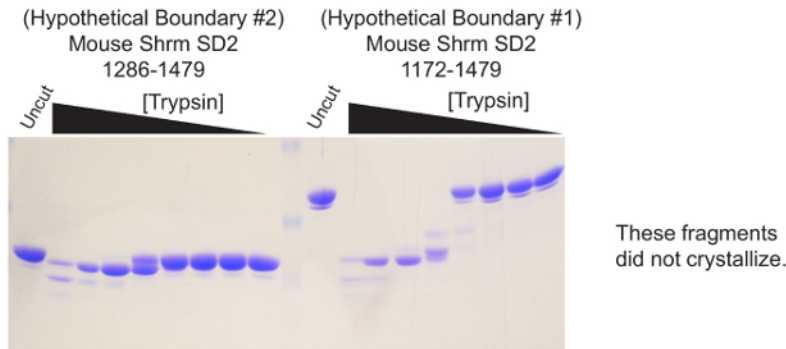
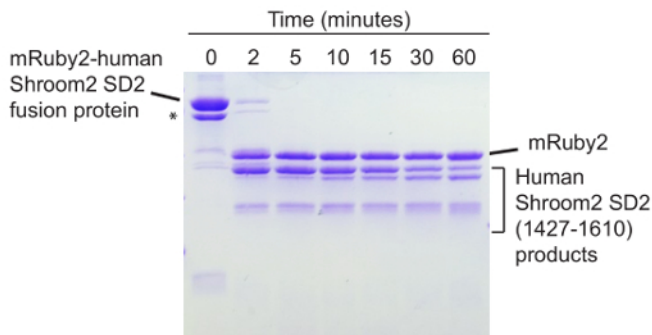## Illustration of Sample Purification of His$_{10}$-mRuby2-Shroom SD2
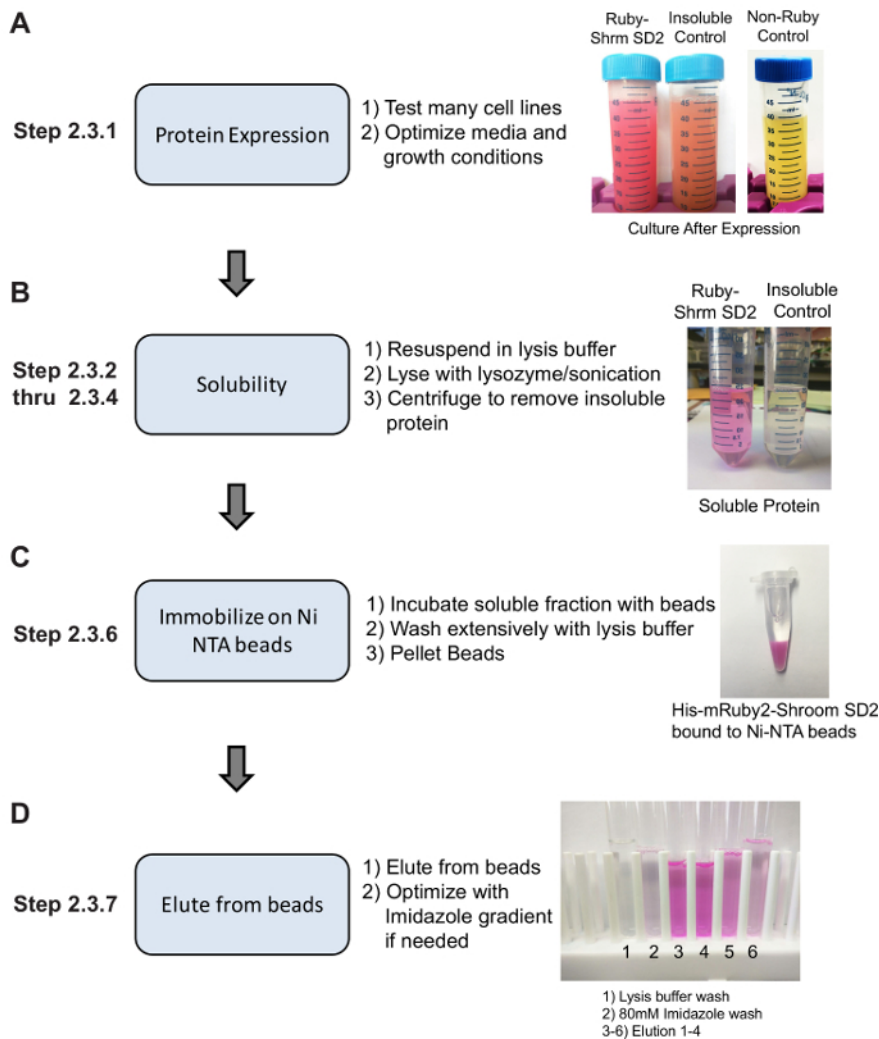
Figure 4: Representative small scale purifications of two mRuby2 tagged fusion proteins. (**A**) An image of samples of bacterial culture expression Ruby Shrm SD2 or an unrelated Ruby fusion protein known to be insoluble are pictured. A separate culture expressing a protein without a Ruby-tag is shown for comparison. (**B**) The soluble fraction from the cultures above were imaged following lysis and centrifugation at 30,000 x g for 30 min, demonstrating the visualization of soluble Ruby-Shrm SD2. (**C**) Image of the Ni-NTA beads after binding and subsequent washing steps indicating that Ruby-Shrm SD2 is being immobilized onto the beads. (**D**) Samples of washing and elution steps as described in steps 2.3.6 and 2.3.7 demonstrate that the Ruby-Shrm SD2 fusion remains bound to the resin while in the presence of up to 80 mM imidazole and is effectively eluted off the column with 1 M imidazole elution buffer. Please click here to view a larger version of this figure.
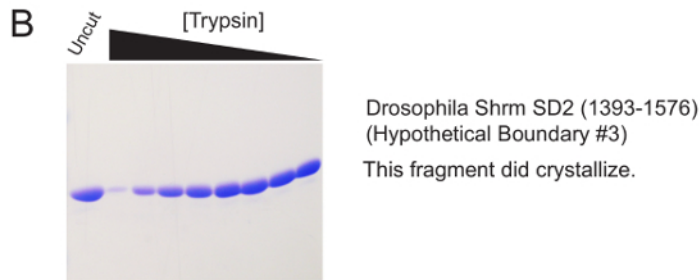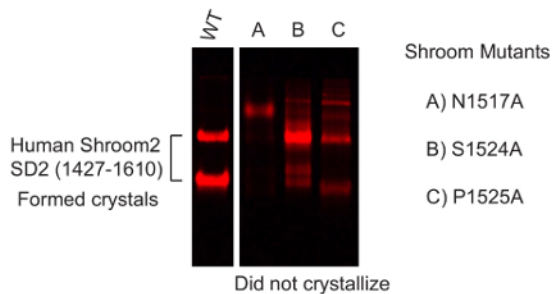
Figure 5: Limited proteolysis of candidate SD2 domains from different Shroom proteins. A comparison of four SD2 domains from various Shroom proteins. (**A** and **B**) The indicated Shrm SD2 fragments were incubated with a concentration gradient of the protease trypsin from 0 to 1.0% and the results analyzed by SDS-PAGE. The relationship of that protein fragment with the hypothetical boundaries are indicated. (**C**) SDS-PAGE analysis of limited proteolytic digestion of His_{10}-mRuby2-Shroom SD2 fusion protein using the time course method described in the protocol. The reaction occurred with 0.025% trypsin at room temperature. Digestion of the linker between Ruby and Shroom SD2 is rapid and serves as an internal control. A presumed degradation product that co-purifies with the His_{10}-mRuby2-Shroom SD2 fusion protein is indicated (*). Please click here to view a larger version of this figure.
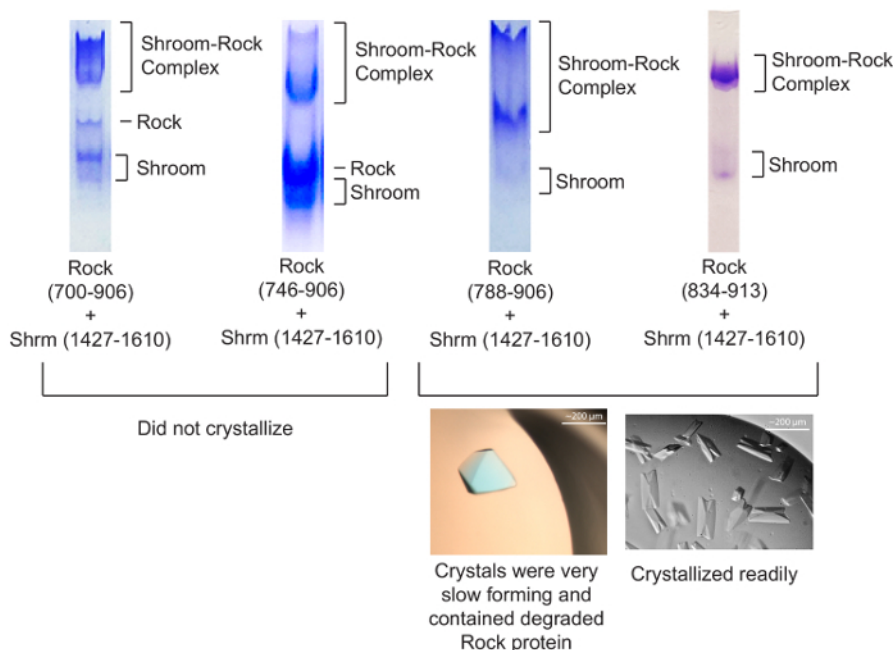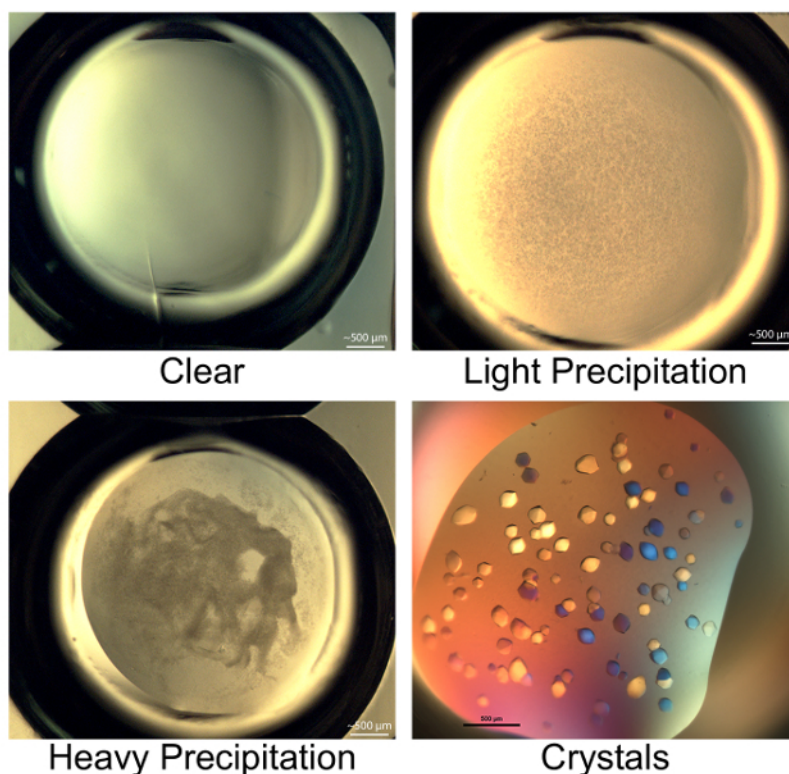
Figure 6: Observing changes in protein behavior using native gel. (**A**) 10% Native PAGE demonstrating the effect of a mutant that changes the properties of mRuby2-Shroom SD2. Under these conditions Shrm SD2 runs as two discrete bands, while indicated point mutants within the SD2 display a range of aberrant migration patterns. (**B**) A comparison of Shroom SD2-Rock complexes formed using different versions of Rock and analyzed by native PAGE. Complexes between Shroom SD2 and the indicated fragments of Rock kinase (both coiled-coil proteins) were resolved by native PAGE. Crystals were obtained for the complex containing Rock 788-906 after many weeks but complex containing Rock 834-913 crystallized rapidly. Please click here to view a larger version of this figure.

Figure 7: Quick primary screen for crystallization. (**A**) A quick screen of common crystallization conditions is used to assess the behavior of the protein sample in crystallization trials. (**B**) Examples of the simple scoring matrix to assess crystallization drops. Please click here to view a larger version of this figure.

Discussion

The protocol described here is designed to help the user identify domain boundaries within large coiled-coil proteins to facilitate their crystallization. The protocol relies on a holistic incorporation of a variety of data from computational predictions and other sources to generate a series of potential domain boundaries. These are followed by a set of biochemical experiments which are quick and inexpensive, and are used to further refine these initial hypotheses. Using this approach, the user could quickly eliminate potential protein fragments that are undesirable, and focus more attention on better candidates, thereby improving the prospects of obtaining crystals.

There are many important steps within this technique, however, none as critical for the production of crystals as the development of the initial set of hypothetical domain boundaries. This step incorporates a variety of computational approaches, as well as information obtained from the literature or functional data when available. Care should be taken to avoid using the strategy to hone immediately down to the single "best" solution as there is currently no method in place to attach a quantifiable confidence value to any of the predictions. Instead, it is best used as a method to quickly identify a small set of possible domain boundaries which need to be experimentally verified.

The properties of coiled-coil proteins which can be analyzed with this protocol are quite broad. From a computational perspective, the strength of coiled-coil predictions is limited below 20 amino acids, and as mentioned in step 1.2.1, coiled-coil regions larger than 1,000 amino acids would need to be split into multiple sections for analysis by DISOPRED. We view this later limitation as temporary as it is imposed by the webserver and may change as the method is upgraded in the future. The subsequent biochemical analysis however, can suffer in various ways. First, internal loops or regions hypersensitive to proteases may make the sample appear to be less stable than it actually is. Stable proteins that have cleaved loops or are otherwise nicked by the protease should still give a stable appearance on native PAGE, which is why it is recommended that the user explore both strategies. Native PAGE may be difficult for some proteins, however, either because they are very long and migrate slowly into gel or because their particular charge may make them run the opposite direction out of the gel entirely. In this cases, it may be helpful to explore different buffering conditions for the native PAGE gel system.

While the focus of the work presented here has been on large coiled-coil containing proteins, this protocol can be utilized for almost any protein with very few adaptations. The primary addition for globular proteins would be the addition of domain prediction software such as pDomTHREADER[28] or DomPred[29] to look for domain boundaries, and tertiary structure predictions such as Phyre2[30] to enhance the predictive power of the sequence analysis. Additionally, there are many algorithms available for performing secondary structure predictions and disordered predictions, and the inclusion of additional algorithms could be helpful. Globular proteins often possess enzymatic activity or other functional readouts that would provide important additional information during target selection. Further, moderate or high-throughput techniques can be incorporated as appropriate. For example, inexpensive systems for 1-2 ml cultures and metal affinity purifications in 96-well formats are readily available[31]. Lastly, the use of robotics for setting up large numbers of crystallization experiments using minimal sample is becoming routine, but efficient operation of robotic systems is predicated on the behavior of well characterized protein samples. Additional modifications to this protocol could include sampling a variety of affinity tags. Many different fluorescent tags are available, or His-, GST-, Thioredoxin-, Strep-, and MBP- are among dozens of available options if a fluorescent tag is not needed or helpful.

When performing this procedure, the user should be mindful of the effect that the mRuby2 tag might have on the users' protein of interest. Anecdotal evidence from using this tag on a variety of different fusions supports the fact that mRuby2 will sometimes bind quite tightly to the protein of interest, remaining bound through multiple chromatographic steps. This behavior is obviously undesirable, but unintended Ruby complexes can usually be separated and removed chromatographically. It is unclear whether this is a chaperone-like behavior as has been observed for MBP[32].

After obtaining crystals, there are still many challenges that are commonly faced with coiled-coil proteins that are not addressed in this protocol. The most common is poor diffraction quality, either due to imperfectly formed lattices or anisotropic diffraction. There are tools in place to assist with anisotropic diffraction[33], and these have been critical for solving some coiled-coil structures[18,20]. Many crystal pathologies are unfortunately difficult to overcome quickly, therefore it is often prudent to search for additional crystal forms with enhanced diffraction properties. This is facilitated by robotic screening of thousands of conditions. Alternatively there are many post-crystallization techniques to improve diffraction quality such as dehydration, or seeding[34].

Disclosures

The authors have nothing to disclose.

Acknowledgements

References

1. Studier, F. W. Stable expression clones and auto-induction for protein production in E. coli. *Methods Mol Biol.* **1091** 17-32 (2014).
2. Studier, F. W., & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol.* **189** (1), 113-130 (1986).
3. Chapman, T. Protein purification: pure but not simple. *Nature.* **434** (7034), 795-798 (2005).
4. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods.* **6** (5), 343-345 (2009).
5. Lupas, A., Van Dyke, M., & Stock, J. Predicting coiled coils from protein sequences. *Science.* **252** (5009), 1162-1164 (1991).
6. Offer, G., Hicks, M. R., & Woolfson, D. N. Generalized Crick equations for modeling noncanonical coiled coils. *J Struct Biol.* **137** (1-2), 41-53 (2002).
7. Lupas, A. N., & Gruber, M. The structure of alpha-helical coiled coils. *Adv Protein Chem.* **70** 37-78 (2005).
8. Hernandez Alvarez, B. *et al.* A new expression system for protein crystallization using trimeric coiled-coil adaptors. *Protein Eng Des Sel.* **21** (1), 11-18 (2008).
9. Slabinski, L. *et al.* The challenge of protein structure determination--lessons from structural genomics. *Protein Sci.* **16** (11), 2472-2482 (2007).
10. Slabinski, L. *et al.* XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics.* **23** (24), 3403-3405 (2007).
11. Haigo, S. L., Hildebrand, J. D., Harland, R. M., & Wallingford, J. B. Shroom induces apical constriction and is required for hingepoint formation during neural tube closure. *Curr Biol.* **13** (24), 2125-2137 (2003).
12. Hildebrand, J. D. Shroom regulates epithelial cell shape via the apical positioning of an actomyosin network. *J Cell Sci.* **118** (Pt 22), 5191-5203 (2005).
13. Dietz, M. L., Bernaciak, T. M., Vendetti, F., Kielec, J. M., & Hildebrand, J. D. Differential actin-dependent localization modulates the evolutionarily conserved activity of Shroom family proteins. *J Biol Chem.* **281** (29), 20542-20554 (2006).
14. Bolinger, C., Zasadil, L., Rizaldy, R., & Hildebrand, J. D. Specific isoforms of drosophila shroom define spatial requirements for the induction of apical constriction. *Dev Dyn.* **239** (7), 2078-2093 (2010).

15. Farber, M. J., Rizaldy, R., & Hildebrand, J. D. Shroom2 regulates contractility to control endothelial morphogenesis. *Mol Biol Cell.* **22** (6), 795-805 (2011).

16. Das, D. *et al.* The interaction between Shroom3 and Rho-kinase is required for neural tube morphogenesis in mice. *Biol Open.* **3** (9), 850-860 (2014).

17. Dvorsky, R., Blumenstein, L., Vetter, I. R., & Ahmadian, M. R. Structural insights into the interaction of ROCKI with the switch regions of RhoA. *J Biol Chem.* **279** (8), 7098-7104 (2004).

18. Mohan, S. *et al.* Structure of a highly conserved domain of Rock1 required for Shroom-mediated regulation of cell morphology. *PLoS One.* **8** (12), e81075 (2013).

19. Mohan, S. *et al.* Structure of Shroom domain 2 reveals a three-segmented coiled-coil required for dimerization, Rock binding, and apical constriction. *Mol Biol Cell.* **23** (11), 2131-2142 (2012).

20. Tu, D. *et al.* Crystal structure of a coiled-coil domain from human ROCK I. *PLoS One.* **6** (3), e18080 (2011).

21. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* **7** 539 (2011).

22. Shapiro, A. L., Vinuela, E., & Maizel, J. V., Jr. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochem Biophys Res Commun.* **28** (5), 815-820 (1967).

23. Diezel, W., Kopperschlager, G., & Hofmann, E. An improved procedure for protein staining in polyacrylamide gels with a new type of Coomassie Brilliant Blue. *Anal Biochem.* **48** (2), 617-620 (1972).

24. Wittig, I., & Schagger, H. Advantages and limitations of clear-native PAGE. *Proteomics.* **5** (17), 4338-4346 (2005).

25. Jungbauer, A., & Hahn, R. Ion-exchange chromatography. *Methods Enzymol.* **463** 349-371 (2009).

26. Yu, C. M., Mun, S., & Wang, N. H. Theoretical analysis of the effects of reversible dimerization in size exclusion chromatography. *J Chromatogr A.* **1132** (1-2), 99-108 (2006).

27. Giege, R. A historical perspective on protein crystallization from 1840 to the present day. *FEBS J.* **280** (24), 6456-6497 (2013).

28. Lobley, A., Sadowski, M. I., & Jones, D. T. pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics.* **25** (14), 1761-1767 (2009).

29. Marsden, R. L., McGuffin, L. J., & Jones, D. T. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. *Protein Sci.* **11** (12), 2814-2824 (2002).

30. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* **10** (6), 845-858 (2015).

31. Elsliger, M. A. *et al.* The JCSG high-throughput structural biology pipeline. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* **66** (Pt 10), 1137-1142 (2010).

32. Bach, H. *et al.* Escherichia coli maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. *J Mol Biol.* **312** (1), 79-93 (2001).

33. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A.* **103** (21), 8060-8065 (2006).

34. Heras, B., & Martin, J. L. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D Biol Crystallogr.* **61** (Pt 9), 1173-1180 (2005).